Entropy and Information within Intrinsically Disordered Protein Regions

Iva Pritišanac [1,2], Robert M. Vernon [1], Alan M. Moses [2,3,4,*] and Julie D. Forman Kay [1,5,*]

1   Program in Molecular Medicine, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada
2   Department of Cell & Systems Biology, University of Toronto, Toronto, ON M5S 3B2, Canada
3   Department of Computer Science, University of Toronto, Toronto, ON M5T 3A1, Canada
4   Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON M5S 3B2, Canada
5   Department of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada
*   Correspondence: alan.moses@utoronto.ca (A.M.M.); forman@sickkids.ca (J.D.F.-K.)

Abstract: Bioinformatics and biophysical studies of intrinsically disordered proteins and regions (IDRs) note the high entropy at individual sequence positions and in conformations sampled in solution. This prevents application of the canonical sequence-structure-function paradigm to IDRs and motivates the development of new methods to extract information from IDR sequences. We argue that the information in IDR sequences cannot be fully revealed through positional conservation, which largely measures stable structural contacts and interaction motifs. Instead, considerations of evolutionary conservation of molecular features can reveal the full extent of information in IDRs. Experimental quantification of the large conformational entropy of IDRs is challenging but can be approximated through the extent of conformational sampling measured by a combination of NMR spectroscopy and lower-resolution structural biology techniques, which can be further interpreted with simulations. Conformational entropy and other biophysical features can be modulated by post-translational modifications that provide functional advantages to IDRs by tuning their energy landscapes and enabling a variety of functional interactions and modes of regulation. The diverse mosaic of functional states of IDRs and their conformational features within complexes demands novel metrics of information, which will reflect the complicated sequence-conformational ensemble-function relationship of IDRs.

Keywords: intrinsically disordered proteins; Shannon entropy; information theory; evolutionary conservation; conformational entropy; low-complexity sequences; liquid-liquid phase separation; post-translational modifications; conformational ensembles; biophysics

1. Information—Central to the Central Dogma

The concept of information is deeply rooted in the central dogma of molecular biology. According to the simplest version of the dogma, a coding stretch of DNA encodes the information for RNA, which in turn encodes a protein that encodes a function [1]. When defining the dogma, Crick specified that information refers to "the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein." [1]. Given that the essential molecules of life act as encoders and transmitters of information, life itself can be considered a manifestation of the flow of information (Box 1). Indeed, the ability to process information has been proposed as one of the criteria to define life [2]. Since the discovery that the 3D structures of folded proteins are encoded in their amino acid sequence it has been understood that this genetic information translates to structure and function through a thermodynamic process that minimizes free energy, which for structured proteins

involves forming functional ordered conformations [3]. This structure-function relationship makes the determination of 3D structures of proteins a major focus of biological research [4] with structural information used to rationalize molecular mechanisms and therapeutic design [5].

1.1. Information in IDRs—Problems for The Paradigm

Proteins that lack stable tertiary structure in their native form, known as intrinsically disordered proteins and regions (IDRs), comprise a large fraction of the eukaryotic proteome [6,7]. Many studies indicate that IDRs often exhibit large evolutionary sequence variation [8–10] and sample a vast 3D conformational space [11]. However, the apparent randomness of IDR evolution and 3D conformations is at odds with the diversity and significance of IDRs' biological roles in regulatory [8,12,13] and signaling processes [14,15], and their frequent implications in disease [14,16–18]. Recent methodological developments are increasingly unveiling "hidden" information in IDRs and in so doing are forcing us to reconsider how thermodynamic protein behavior can translate genetic information to function. Here, we discuss entropy in the sequences and conformational ensembles of IDRs and the requirement for new ways of converting entropy to information in the context of their biological functions.

We briefly outline the concepts that are essential for understanding the terminology (Section 1). In Section 2, we argue that methods based on positional information in multiple sequence alignments need to be abandoned in order to quantify the information in the sequences of IDRs. In Section 3, we discuss experimental techniques that characterize various aspects of the high conformational entropy of IDRs. We explore how conformational plasticity is utilized for the interactions of IDRs with their targets, and how it is impacted by post-translational modification and liquid-liquid phase separation. We conclude by discussing future perspectives for defining and exploiting the information that is available in the sequences and conformational ensembles of IDRs (Section 4).

1.2. Uniting Different Entropies and Extracting Information

Entropy is a concept that transcends and unites different areas of science [19,20]. The term itself was born with thermodynamics [21], with subsequent atomic interpretation of entropy giving foundations to statistical mechanics [22,23]. In the general terms of information theory, entropy is a measure of the uncertainty in the identity of a random variable or the state of a system, given as:

$$H(P) = -K \sum_i p_i log_2 p_i \tag{1}$$

where K is a positive constant and P stands for the probability mass function (PMF), $P = \{p_1 \ldots p_n\}$, or the probability of a random discrete variable or a system to be found in states $i = \{1, 2 \ldots n\}$. If the logarithm in Equation (1) is base 2, entropy is measured in bits. In information theory, K is set to unity, whereas in statistical mechanics, K becomes the Boltzmann constant ($K = k_B = 1.3808 \times 10^{-23}$ J/K). In his pivotal work, Shannon [24] demonstrated that Equation (1) captures the intuitive notions of entropy: it is positive, increases with increasing uncertainty, and is additive with respect to the independent sources of uncertainty.

In practice, to measure information content, relative entropy or Kullback-Leibler (KL) divergence is often employed, which measures entropy of a probability distribution with respect to another probability distribution, given as:

$$D(P \parallel Q) = \sum_{i \in A} p_i log_2 \frac{p_i}{q_i} \tag{2}$$

where P and Q are defined over a finite set A, $i \in A$. KL divergence can be computed between a reference probability distribution ($Q = P^*$), and the measured one (P) to evaluate the dissimilarity between the two. Here we refer to the KL divergence between the distribution of symbols in biological sequences under "total uncertainty" (Q) and the observed distribution (P) as the "Sequence Information", which

is distinct from the "Mutual Information," (Box 2) a more formal measure of information that is also widely used in biological sequence analysis.

2. Sequence Entropy Metrics Fail to Extract Information from IDRs

2.1. Sequence Entropy Can be Computed Horizontally and Vertically

Complexity, information, and entropy have been difficult to apply for practical gain in many biological settings. However, Shannon entropy and related information theoretic measures, such as relative entropy and mutual information (Box 2), have been widely applied to study biological sequences [25]. To apply these concepts, biological sequences are thought of as strings of symbols, the four "letters" of DNA and RNA and the 20 of proteins. If these letters are interpreted as messages from an information source in the Shannon sense, the entropy in a biological sequence can be computed using the standard formula (Equation (1)).

Box 1. What is information?

Information is the reduction in entropy (here thought of as uncertainty) in the receiver [26]. In the biological context of the central dogma, we can imagine an mRNA carrying information to the ribosome. Before the mRNA arrived, the ribosome might be expected to assemble polypeptides with amino acids in a random order. However, once the mRNA arrives, out of the astronomically many sequences that are possible, the ribosome assembles one specific polypeptide sequence (note that this is more than one in reality because of noise, i.e., errors in the translation process). Hence, the entropy of the produced polypeptide is dramatically reduced: information has been transmitted. This example is instructive because it highlights the dependence on the receiver, in this case the ribosome: if the ribosome did not recognize the mRNA as a signal, no information would be transmitted.

Thus, information is measured as the difference in entropy before and after a probabilistic event, a measurement, or receiving of a message:

$$I(X) = H_{before} - H_{after} \qquad (3)$$

The maximum information is gained when H_{after} equals 0 and thus all H_{before} is converted to information, so the entropy sets the maximum for the possible information transmitted. This, however, need not be the case in practice, as communication channels can contain noise or some degree of uncertainty can remain after receiving a message or performing a measurement ($H_{after} > 0$). Note that relative entropy and information are often used interchangeably in the bioinformatics literature.

Shannon entropy can be measured horizontally, i.e., in a window across a protein sequence to detect a statistical bias in the use of the 20 amino acid alphabet in protein sequences [27–29]. In practice, the bias is evaluated relative to a uniform or an empirical background distribution (Equation (2)). The sequences with reduced amino acid alphabets will show reduced (relative) entropy values, which can be used to classify protein sequences as 'high' or 'low' complexity, with some IDRs sequences displaying low complexity [27–29] (see Section 2.5). There are statistical limitations of such horizontal entropy evaluations due to limited samples, and it is unclear whether and how these metrics on their own could be used to extract information specific to a protein function. An alternative way to measure entropy as a way to gain functionally relevant information from biological sequences is to focus on a functionally equivalent set of sequences and evaluate entropy vertically across their alignment [30,31].

Box 2. Mutual information.

Mutual information is another entropy-based metric often employed to measure the (in)dependence of two random variables or probability distributions. If two random variables *X* and *Y* are independent, then their joint probability distribution, *P(X,Y)*, will be equal to the product of their individual probability distributions:

$$P(X, Y) = P(X)P(Y) \tag{4}$$

The degree of dependence can be revealed by computing the relative entropy (KL divergence, Equation (2)) between the joint probability distribution and the product of the individual distributions:

$$MI(X;Y) = D(P(X,Y) \parallel P(X)P(Y)) = \sum_{i,j} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \tag{5}$$

For independent variables for which Equation (4) holds, the mutual information is equal to zero. Otherwise, the mutual information corresponds to a reduction in entropy, i.e., a gain in information, about the outcome of *X* given the knowledge of *Y*. The latter is best expressed in an equivalent, theoretical expression of mutual information that uses the concept of conditional entropy, which corresponds to the entropy of *X* given the knowledge of *Y*, *H(X|Y)*, and vice versa *H(Y|X)*:

$$MI(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{6}$$

In studies of biological sequences, mutual information is computed between the columns of a multiple sequence alignment to evaluate if neighboring, or distant, positions along the sequence are independent or covary. To this end, the frequencies of co-occurrence of elements at potentially covarying positions are used for $P(x_i, y_j)$ in Equation (5) [32]. In other applications, mutual information is often difficult to use because the joint distribution (in Equation (5)) between the biological receiver and signals is difficult to obtain. For example, to compute mutual information between a protein binding event and a peptide sequence, even under the assumption that there are only two states (bound and unbound), the joint distribution includes all the unbound peptide sequences, which are, in practice, rarely known. In contrast, the KL divergence (Equation (2)) can be computed for the distributions between uncertain (before) and bound (after) states, which does not require the full joint distribution.

2.2. Sequence Entropy in Biological Macromolecules: The "Positional Information Paradigm"

In the 'vertical' applications, each position in a set of aligned biological sequences is regarded as an independent message, and relative entropy is therefore calculated at each position. The probability, *P*, is simply a multinomial distribution on the observed numbers of each type of symbol at that position. The relative entropy at each position can be computed with respect to the "maximum" entropy for that sequence type: $\log_2(4)$ for DNA and $\log_2(20)$ for proteins, or alternatively, relative to a defined background distribution (*Q* in Equation (2)). This can be, for instance, the overall distribution of nucleotides or amino acids in the aligned sequences [32]. The relative entropy at each position thus corresponds to information and forms the basis for the widely used "sequence logo" representation for biological sequences [33] (Figure 1A). In this representation, the heights of the letters in this representation are proportional to the information at each position [33]. Since information is additive, to obtain the information in the whole alignment, the information at each position is simply added.
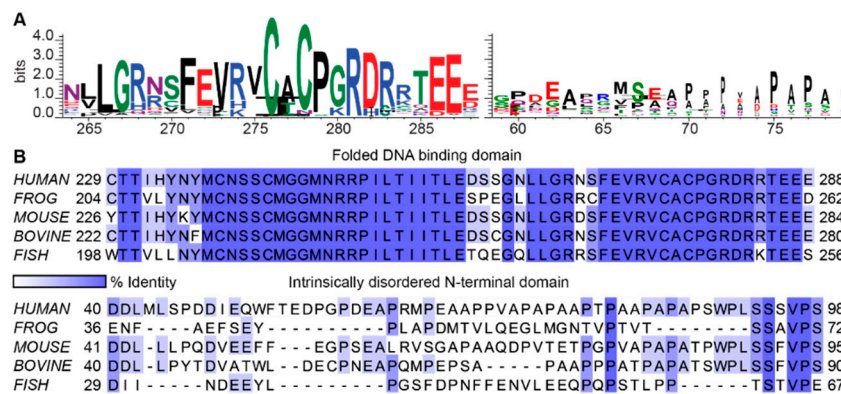
Figure 1. Entropy in alignment reveals information at position according to the positional paradigm. (**A**) *Left*, the 'sequence logo' generated for a portion of the DNA binding domain (DBD) of the tumor protein p53 based on the multiple sequence alignment of 69 vertebrate orthologues of p53. *Right*, the 'sequence logo' of the intrinsically disordered N-terminal domain (NTD) of p53 based on the same alignment. The numbered positions on the x-axis correspond to the positions in the human sequence. Both logos were generated using WebLogo [34]. (**B**) Multiple sequence alignment of a portion of the folded DNA binding domain (*top*), and intrinsically disordered N-terminal domain (*bottom*) of p53. The alignment is displayed for several vertebrate orthologues. In contrast to folded domains, IDRs display low positional conservation.

The information at each position in alignments has been found empirically and theoretically to be predictive of biological function, where information is typically associated with the concept of sequence conservation. In classical studies, information at each position in DNA sequences patterns has been found to predict protein binding in vitro and in vivo [35,36]. Similar results have been found for the information in protein sequences: positions with large amounts of information point to the functional residues [37,38]. In fact, protein folding has been proposed as 'a noiseless communication channel' [39] leading from sequence to structure, whereby all information in the protein structure is shared with the sequence (mutual information) [30,39]. The positional information and the correlations (mutual information) between positions are being used in predictions of protein structure [40–43], catalytic residues [44] and protein interaction interfaces [45].

2.3. Evolutionary Origin of Positional Information in Sequence Alignments

Biological sequences represent the products of the evolutionary process. Sequences are the "genotypes" that produce biological functions and phenotypes. Hence, if natural selection is acting to preserve a biological function, changes (mutations) in the sequence (genotype) that alter that function will be removed from the population over evolutionary time. It has been recognized since the early days of molecular evolution that natural selection would act to reduce the entropy and increase the information in the genotype [46]. Indeed, modern bioinformatics analyses have borne out this prediction: sequence positions with more information show less evolutionary variation [47] and stronger natural selection [48].

Thus, theoretical, empirical, and evolutionary arguments all support the application of information theory to the positions in alignments [30,31,49], and the use of information as a measure of the biological importance of that position. So appealing is this "positional information paradigm" that the arguments have been presented for a 1:1 relationship between sequence information and function [49]: if there were no detectable information at some positions of a sequence alignment, this could rule out the possibility of function for those positions. In other words, positions in sequences with no information have been considered to encode no function.

2.4. Intrinsically Disordered Regions Contain Little Positional Information, but Still Encode Function

Most positions in disordered regions are not readily aligned across long evolutionary distances [50]. When homology can be inferred based on surrounding folded domains, many of the positions in disordered regions show nearly maximum sequence entropy, as expected for random sequences (Figure 1). IDRs are known to contain molecular recognition features (MoRFs) that are characterized by disorder-to-order transition upon interactions, and short-linear peptide motifs (SLiMS) that mediate protein-protein interactions and are often associated with posttranslational modifications [8,51–54]. In some cases, these functional regions of IDRs show clear positional information [54]. However, the information is limited to short stretches of amino acid residues, typically three to fifteen amino acids long [55]. Hence, alignments of IDRs reveal little sequence information, as measured by the definitions introduced above, and the functional importance of most amino acid residues in disordered regions remains unknown. Thus, application of the positional information paradigm is consistent with the idea that most IDRs encode no biological function ("junk proteins" [56]) or serve as inert linkers or tethers that harbor interaction sites (MoRFs or SLiMs) needed to bring other components together [57].

Nonetheless, there is increasing appreciation of the critical biological function of disordered regions with little positional information [9,10]. How can we reconcile this contradiction? We argue that the positional approach to information primarily measures the spatial relationship between residues, i.e., functions specific to the exact position in a folded structure that a residue occupies. If a protein encodes a functional structure, then the metric coincides with function. In fact, residues in folded proteins evolve according to their spatial position in a protein structure. No two residues can occupy the same position in space, and when a position in a structure has a unique functional role no two residues can share that function. This uniqueness of positional information in folded proteins underlies most methods involved in positional analysis of sequence alignments (see Section 2.2). In contrast, if a protein functions as an ensemble of conformers (Section 3), where multiple residues can occupy the same functional positions in space dynamically over time, then the information on function remains undetectable when attempted to be revealed through a multiple sequence alignment. This consideration illustrates how the relationship between space and function in IDRs is significantly more complicated than the 1:1 relationship characteristic of folded proteins.

Given that most of an IDR amino acid sequence does not need to encode stable residue-specific interactions, what kind of, and how much information do IDRs contain? Studies of the sequence determinants of IDR function for individual proteins have revealed "cryptic" sequence features that are correlated with specific molecular functions [58]. This suggests that IDR sequences do contain some form of information, but that it is not detectable using sequence alignment-based techniques, which assume that each position carries information independently. Recently, we and others have begun to detect evolutionary signatures in disordered regions that are largely devoid of positional information [9,10,59,60]. This revealed that most of the apparently highly diverged IDRs in the yeast proteome contain multiple conserved molecular features in the form of physicochemical properties (e.g., net charge, isoelectric point, hydrophobicity, fraction of polar residues, etc.), amino acid repeats (e.g., RG, RGG, QQ), sequence complexity, and sequence motifs (e.g., phosphorylation motifs, proline-rich motifs, nuclear localization signals, mitochondrial localization motifs, etc.) [10]. The evolutionary signatures in IDRs establish how natural selection can preserve distinct molecular features predictive of function, and suggest that, in addition to sequence motifs, bulk properties of disordered regions encode functional information stored in IDRs [10].

A number of functional roles of IDRs have been identified that do not rely on the formation of stable, folded structure: proteinaceous detergents and solubility tags, scaffolding within interaction hubs, entropic springs, timers and linkers, allosteric transmission, and binding via bulk electrostatic or other physical property. Other important functions of IDRs that do not generally require stable structure are in cellular compartmentalization and biomaterial formation through the process of liquid-liquid phase separation, with consequences for regulation of enzymatic reactions and other biological processes [61–68]. The lack of stable structure exposes IDRs for interactions with multiple

binding partners, which can enable more efficient driving of macromolecular condensation. In addition to the dynamic multivalent interactions of sequence motifs within IDRs with modular binding domains of folded proteins, IDRs also contribute to phase-separation through their low-complexity (LC) regions, which can participate in multi-valent protein and nucleic acid binding [69–74]. A number of studies have assessed the extent to which conformational dynamics and secondary structural propensities of IDRs are affected by phase-separation [69,73,75–77]. This revealed that IDRs can remain highly disordered in the phase-separated droplets [68–70,75], while exhibiting significantly slower translational diffusion [69,75]. Inter-molecular contacts within the droplets could be detected [69,75,76], with an increase in intra-molecular contacts also reported in some cases [76]. It is interesting to note that sequence alterations and disease-associated mutations in IDRs can directly impact phase separation propensities [69,78,79]. For instance, disease-associated mutations in hnRNPA2 and FUS were found to promote aggregation [72,78,79], while some ALS-associated variants of TDP-43-LC disrupt phase separation by destabilizing the transient formation of a helical conformation [71].

2.5. Information in Low-Complexity IDR Sequences

As defined above, sequence entropy assumes that each symbol in a protein, where symbols correspond to 20 amino acids, carries independent information. In most proteins this assumption holds, as individual amino acids make stable inter- and intra-molecular contacts and the observed distribution of amino acids is the result of selection acting on random point mutations in the DNA sequences that encode them. Correlations between positions will reduce positional entropies due to functional associations between positions [30,31]. However, when DNA and protein sequences are the result of replication machinery errors (e.g., unequal cross-over, replication 'slippage') [80,81] each symbol no longer has the full range of possibilities, i.e., the next amino acid in the sequence can only be one of the previous ones that has been copied and inserted. These mechanisms have generated a large class of sequences in the genome known as "low-complexity", which show a reduced amino acid (or DNA) alphabet. These sequences are typically removed from standard bioinformatics analyses because they violate many of the assumptions of the methods [28,82,83]. Paradoxically, low-complexity, by definition, implies low entropy and therefore, when compared to the entropy of complete uncertainty, large information, which in this case is very different from that in globular folded protein domain sequences. We note that this "information" is likely to be a result of the different mutation processes that generate these sequences [56,80,81], and it is unclear in many cases what is, or whether there is, a "receiver" for this "information."

Use of "horizontal" sequence entropy metrics (across a protein sequence) revealed that IDRs often exhibit low-complexity, however, they are not limited to it [29]. Similarly, low-complexity sequences are not always disordered and detection of low-complexity alone is not predictive of any particular protein sequence property or function [66]. Nonetheless, categorization of low-complexity IDRs based on physicochemical characteristics can be used to classify these sequences based on their properties [66]. In this context, there is a considerable association of low-complexity IDR sequences with phase separation mediated through π-π interactions [84] or backbone beta interactions that can also lead to fiber formation [85,86]. Figure 2 shows that while low-complexity human protein regions are more often predicted to phase separate there is additional information content in factors such as specific composition and length, with even the total sum of glycine residues containing information not captured by Shannon entropy. Further developments in this direction are important given the enrichment of low-complexity IDRs in biomolecular condensates [87–89].

Similarly, functionally relevant insight can be gained from correlations between features of the primary amino acid sequence and the observed phase separation behavior of IDRs in vitro and in cells [70,84,90–93]. In addition, polymer physics models are being built to understand the competing enthalpic and entropic contributions to the free-energy changes that underlie the phase separation of IDRs, and the dependence of these contributions on the primary amino acid sequences and post-translational modifications (PTMs) [66,94–96]. Ongoing research in these areas, jointly with insights from cell biology

studies, will further clarify how information stored in IDR sequences encodes their phase-separation behavior, and how this is tuned by the changes in the sequence and/or environment.
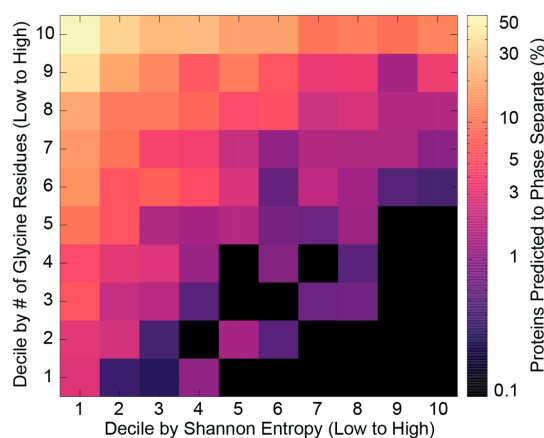

Figure 2. Sequence complexity and glycine content are predictive of phase-separation propensity, shown here for the human proteome across bins ranked by decile position values for glycine content and Shannon entropy of a full sequence. Complexity alone is not sufficient to explain phase-separation behavior; composition differences have additional information even when comparing proteins with similar complexity.

3. IDRs Feature High Conformational Entropy

The high positional ("vertical") entropy measured from multiple sequence alignments of IDRs (Figure 1) stems from their lack of stable three-dimensional structures, which in turn underlies their high conformational entropy [6,97–100] (Figure 3). The high conformational entropy of IDRs has been proposed as a thermodynamic reservoir of free energy (Box 3) that can be used to regulate their functions and interactions [15,101–104].
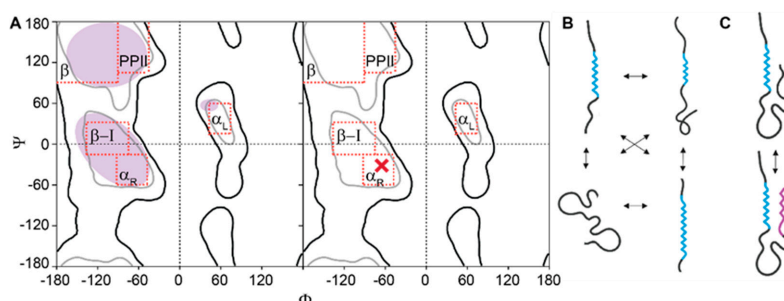

Figure 3. (**A**) The large conformational entropy of IDRs is evident in the broad distribution of the backbone dihedral Φ,Ψ angles that can be sampled by a single residue in the polypeptide chain (purple, left panel, see references [105,106]). Given the large extent of conformational sampling by every single residue, the total extent of conformational sampling of an entire IDR chain quickly reaches astronomical dimensions as the length of the chain increases. In contrast, a residue in a folded domain in a similar chemical environment, i.e., with identical neighboring amino acids, will typically sample a well-defined set of Φ, Ψ angles (red cross) that deviates within a very narrow distribution due to thermal fluctuations. Regions of Φ, Ψ angles that are associated with secondary structural elements are indicated with text on the Ramachandran plot, with favored and allowed areas respectively denoted by the grey and black contours. (**B,C**) An illustration of conformational plasticity of an IDR ensemble in the free state based on the works of Schneider et al. [107] and Iešmantavičius et al. [108]. An ensemble of IDRs can feature a range of transient secondary structure (**B**). Modulation of the secondary structural propensities in one part of an IDR can have a stabilizing effect on the secondary structure formation in a distant region of the IDR through transient tertiary contacts (**C**). Transiently formed α-helical segments are indicated in cyan and purple. Note that IDRs need not sample any transient secondary or tertiary structure to an appreciable degree in order to be functional (see text).

<div style="border:1px solid black;padding:10px;">

Box 3. Entropy in biophysics.

Although equilibrium thermodynamics and information theory share the mathematical expression for entropy (Equation (1)), the term is used in different contexts in the studies of proteins in bioinformatics (see above) and biophysics. In biophysics, entropy is defined, calculated, and measured on multiple levels. The total number of possible arrangements, i.e., different relative center-of-mass positions, of polypeptide chains in solution contributes to configurational entropy [95,109]. Configurational entropy encompasses conformational entropy, which is defined by the number of different conformations of an individual polypeptide chain. Conformational entropy itself can be split into two contributions: local, small-scale fluctuations about a well-defined structure (e.g., an α-helix), and larger-scale conformational differences (e.g., different protein conformations in a random coil ensemble) [110,111]. In an experimental system, the free energy change accompanying protein folding or an interaction with a protein or ligand is routinely measured using, e.g., isothermal titration calorimetry [112]. This measurement allows decomposition of the free energy change, $\Delta G_{tot} = \Delta H_{tot} - T\Delta S_{tot}$, into the total enthalpic (ΔH_{tot}) and entropic ($-T\Delta S_{tot}$) contributions. The entropic component encompasses contributions from all sources of entropy in the system, including the conformational entropy of the protein and its interacting partner (e.g., ligand), the contributions from solvent and the roto-translational degrees of freedom of all interacting partners, and any other, remaining contributions; $\Delta S_{tot} = \Delta S_{protein}^{conf} + \Delta S_{ligand}^{conf} + \Delta S_{solvent} + \Delta S_{r-t} + \Delta S_{other}$ [113]. Untangling the different entropic contributions to the free-energy change, however, remains non-trivial.

</div>

3.1. Conformational Entropy of IDRs is Difficult to Measure Precisely

In contrast to folded proteins that show functionally important conformational dynamics about a well-defined, energetically stable structure, IDRs display large conformational heterogeneity [8,114]. The ease of interconversion between different conformational states of IDRs is well captured by a rugged, relatively flat energy-landscape model with abundant minima separated by small energy barriers [8,101]. Thermodynamically, the structural heterogeneity of IDRs stems from the large contribution of conformational entropy to their free energy (G = H − TS). This entropy contribution becomes a critical component to the free energy change between the different states of IDRs, for instance, upon post-translational modification or complexation with a target biomolecule (Box 3).

Despite our abstract understanding of the conformational entropy as a defining characteristic of IDRs [6,97–100], it has proven tremendously difficult to quantify the full range of this thermodynamic component at IDRs' disposal. Some of the underlying reasons are the limited availability of experimental data to characterize the vast degrees of freedom that contribute to the conformational entropy of IDRs, the lack of understanding of the contributions of solvent, and a still evolving synergy between theory, experiment and simulations (Box 4). Therefore, our understanding of conformational entropy, and its change in a functional context (Figure 4), is limited to the degrees of freedom that can be measured experimentally; increasing contributions from a range of experimental techniques of varying resolutions are helping gain a more conclusive picture [11,115–117]. In addition, molecular dynamics simulations can often enhance the interpretability of experimental parameters [118–120] (Box 4).

The extent of backbone conformational sampling of an IDR in Ramachandran space can be obtained with reasonable precision from the local structural parameters [105] (Figure 3A); however, experimental information is much more sparse for correlations between the backbone angles of distant sites along the chain. Consequently, reconstructing the extent of the conformational sampling, i.e., conformational entropy, of an entire IDR quickly becomes a severely underdetermined problem that grows exponentially with the size of the IDR [105,106].

Box 4. Conformational entropy of IDRs.

Solution-state NMR spectroscopy is the primary experimental technique for the dynamical characterization of proteins at atomic resolution [121–125], which can serve as a proxy for conformational entropy. In the simplest terms, the conformational entropy of a protein can be described by the total number of conformations that are accessible to the protein under a defined set of macroscopic conditions (e.g., protein concentration, temperature, and volume). The full extent of the conformational space of a protein, however, is prohibitively large. Therefore, to compute conformational entropy approximately, conformations are discretized and defined using a combination of structural parameters. In experimental practice, only sparse and typically strictly local structural parameters can be measured for disordered proteins, e.g., distributions of the backbone Φ and Ψ angles (Figure 3A). These can be used to constrain molecular dynamics (MD) or Monte Carlo (MC) simulations and estimate the conformational entropy of a protein.

To understand the restriction of conformational entropy in a disordered polypeptide chain, both short- and long-range restraints are essential. Therefore, continuous efforts are being put forth community-wide [126] to improve the measurement of long-range restraints, correct for any noisy contributions to the restraints (e.g., the error in back-calculations of experimental observables from structures), and integrate information available from complementary, lower-resolution techniques such as small-angle X-ray scattering (SAXS) and single-molecule fluorescence (SMF) [115,116]. These efforts are aided by the continuously improving databases of reference random coil chemical shifts [127] and predictions of NMR observables from statistical coil models [11]. To generate adequate IDR ensemble representations, programs like ENSEMBLE [128] and ASTEROIDS [129] use statistical coil, structurally biased or MD-derived models to initially generate a large pool of conformers that are then sub-sampled to optimize the agreement with a combination of complementary local and global experimental restraints (including NMR chemical shifts, residual dipolar couplings, J-couplings, paramagnetic relaxation enhancements, and nuclear Overhauser effects, SAXS and SMF data). In parallel, a better understanding of the dynamics of inter-conversion between IDR conformers in free and bound-states using spin relaxation and chemical exchange NMR techniques is being sought [11,122,130,131]. Additional insights can be provided by molecular dynamics simulations restrained by experimentally available information [119,120,132]. Further technological advances in NMR spectroscopy in synergy with lower-resolution techniques and improved computational approaches [126] are expected to bring us closer to understanding conformational entropy of IDRs in isolation, and its changes upon PTMs [133,134], interactions with ligands [135,136], and formation of complexes [103].
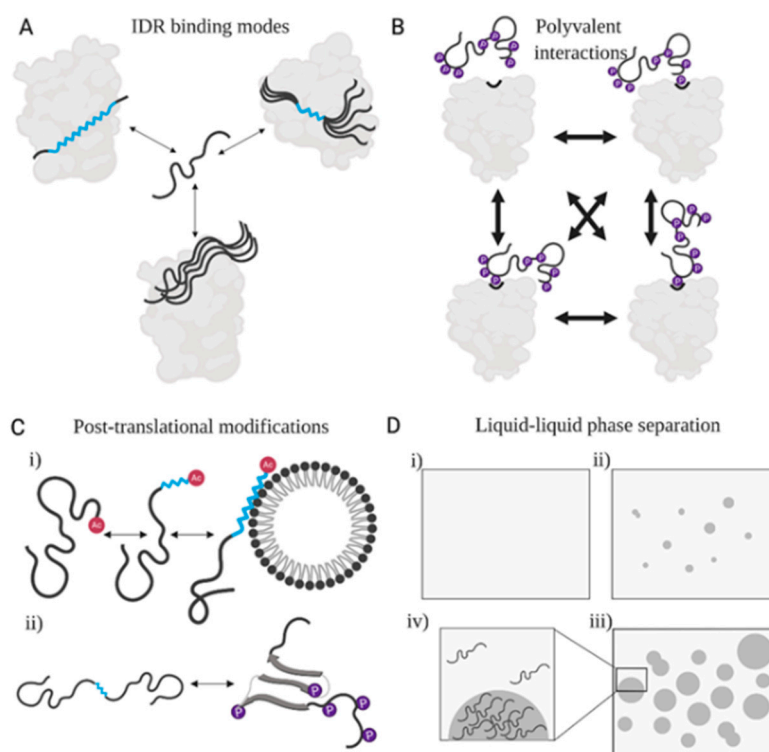

Figure 4. An illustration of functionally relevant processes that impact the configurational and conformational entropy of IDRs. (**A**) IDRs can undergo coupled folding and binding [137] (top left),

partial ordering of one part of the chain with high disorder retained in the rest of the chain [131] (top right), or remain highly disordered (bottom middle) in the complex [138]. The parts of the IDR that retain disorder are shown in black and α-helical regions are in cyan. (**B**) PTMs of SLiMs in an IDR can enable polyvalent interactions, whereby multiple sites on the IDR can dynamically exchange with a single binding site of the partner, as illustrated here based on the example of phosphorylated Sic1 in a dynamic complex with Cdc4 (see Mittag et al. [139]). The presence of multiple phosphorylated sites acts to fine-tune the binding affinity by modulating the charge of the IDR. Phosphate groups are shown as purple spheres. (**C**) (**i**) PTMs can induce transient sampling of secondary structure, which can be further reinforced upon functional interactions. The illustrated example is based on the work from Maltsev et al. [140] on α-synuclein. Acetylation of the N-terminus of α-synuclein leads to transient sampling of an α-helical conformation in the first 12 residues that increases the binding affinity for lipids, through enhancing the association rate. The helical conformation is further propagated in the complex of α-synuclein with lipids. (**ii**) PTMs can induce more significant structural transitions, as demonstrated for 4E-BP2 by Bah et al. [133]. Phosphorylation at particular sites leads to formation of a β-sheet, with the extent of phosphorylation impacting the stability of the formed structure. The formation of the β-strands encompasses a sequence motif that, prior to phosphorylation, transiently samples an α-helical conformation and engages in interactions with the binding partner eIF4E. The formation of the β-strand occludes the interaction site on 4E-BP2 thereby weakening the affinity for the binding partner and regulating the interaction to allow downstream functional consequences (i.e., translational initiation by eIF4E). (**D**) Liquid-liquid phase separation manifests as the separation, and the subsequent stable coexistence, of protein-dense (dark grey circles) and protein-dilute (light grey) phases (**ii,iii**) from an initially miscible protein solution (**i**). In vitro studies of phase-separation provide the basis for understanding the driving forces of the formation of membraneless organelles in cells [64]. The condensation into the dense phase leads to a decrease in the roto-translational freedom of an IDR (i.e., lower configurational entropy); however, IDRs can still retain high conformational entropy in the condensed state (**iv**) [68–70,75]. PTMs can either increase or decrease the propensity of an IDR to phase separate, depending on the overall physicochemical properties of the IDR and the nature of the PTM (see text). The illustration was created with BioRender.com.

3.2. Conformational Diversity and Information in Ensembles of IDRs

The degree of backbone flexibility, local secondary structure propensities, distributions of short- and long-range contacts, and the overall shape and size of IDR ensembles represent experimentally obtained information from IDRs in solution [11,115,116,141,142] (Box 4). On these bases, a picture emerges of the highly diverse conformational features of IDRs in their isolated states (Figure 3B). While some IDRs locally resemble a random-coil [106], but show some compaction due to transient long-range contacts [143,144], others are more compact and appreciably sample transient secondary structure in their free form [139].

The sampling of local transient secondary and tertiary structure in the free forms of IDRs (Figure 3B) is informative, as it can be predictive of their interactions with ligands [145,146] and protein binding partners [147–149], and has been found to impact the kinetics and lifetimes of these interactions [150,151]. For instance, differential propensity to form an α-helical conformation in the free state of disordered regions of two transcription factors (TFs) that bind to the same site on a transcriptional coactivator, was shown to result in different binding mechanisms [150]. The TF that rapidly sampled an α-helical conformation in its free state was shown to interact with the transcriptional activator through conformational-selection. In contrast, the second TF interacted with the transcriptional activator by an induced-fit mechanism [152]. These in vitro observations were proposed to directly relate to the biological functions of the two TFs, as one represents a constitutive transcriptional activator, whereas the other requires phosphorylation for high-affinity binding to the transcriptional coactivator [150]. Another example demonstrated the impact of a transient, partially helical conformation within the disordered N-terminal transactivation domain (TAD) of p53 [148]. The engineered increase in the transient occupancy of the helical conformation in p53-TAD enhanced

the binding affinity for its cellular binding partner and regulator E3 ubiquitin ligase Mdm2, which resulted in disrupted p53 signaling activity and downstream gene regulation in a cellular context [148]. Hence, amino acid substitutions in IDRs can affect transient, local structural propensities [148], and can also exhibit longer-range effects through the disruption of stabilizing tertiary interactions between distant secondary structural elements [131].

In conclusion, IDRs retain a high amount of conformational entropy in their native states, unlike the low entropy states occupied by stable folded proteins. In some cases, IDRs may have conformational entropy that is comparable in magnitude to that of an unfolded polypeptide chain. This conformational plasticity could aid IDRs in sampling a diverse range of distinct functional states or regulate the binding behavior of a single functional conformation. We used the examples above to illustrate how the formation of transient structure in IDRs can, in some cases, reduce the uncertainty about the functional conformational state(s), thereby comprising functionally relevant structural information. Such findings motivate the implicit search for, and utilization of, mutual information between the sequences of IDRs and their structural propensities [153]. Nonetheless, the absence of measurable structural propensities in the free states of IDRs, or in their bound states (see below), does not equal functional irrelevance.

3.3. IDRs Can Retain High Conformational Entropy in Complexes

The absence of transient secondary or tertiary structure in the free state of an IDR does not always equate to a complete absence of structural propensities in a complex (Figure 4A). IDRs engage in diverse complexes [154] featuring complete or partial folding upon binding [131,152,155,156], remaining highly disordered as in "extreme" disordered discrete complexes [157] or phase-separated states [68–70,75], exchange between multiple highly disordered states [138,157], or retention of some secondary and tertiary structural propensity while remaining disordered in the non-interacting regions [103] (Figure 4A). Reports of mixed ordering in some parts of an IDR with increased disorder in other segments are particularly interesting, as they suggest fine-tuning of the entropic loss upon complexation [101,139,158,159], while maintaining a biologically meaningful binding affinity [159]. It is also interesting to note that SLiMs in IDRs need not always rigidify upon binding, as might be thought due to the bias in X-ray structures. In contrast, SLiMs can exhibit fast dynamics on the surface of a binding partner [158], or dynamics on the intermediate timescale in the context of exchange with other binding elements within dynamic, multivalent complexes [139].

The parameters obtained from NMR dynamics measurements of fast (ps-ns) backbone and side-chain dynamics have been introduced as a proxy for contributions to changes in conformational entropy upon protein folding [160], and interactions of folded proteins with their binding partners [113,161,162] (Box 4). An approach for addressing conformational entropy changes in both IDRs and their binding partners upon interaction was proposed on these bases [159]. However, IDRs can often have slower conformational exchange on the NMR chemical shift timescale, e.g., cis-trans Pro isomerization leading to distinct resonances [163], and a range of stabilities of intramolecular interactions leading to some exchange-broadened resonances [164]. In addition, IDRs can exhibit slower conformational exchange in the context of dynamic complexes, which leads to linewidth broadening [165–167] and, while useful in some cases, can hinder attempts to obtain information on the conformational sampling in these protein regions. These features of IDRs challenge the assumption that their conformational entropy can be estimated from fast timescale motion alone. Finally, as IDRs engage in extensive protein-solvent interactions, a better understanding of the solvent contributions to the changes in entropy upon complexation and phase separation involving IDRs (see below) will be critical to fully understand the driving forces behind their biomolecular interactions.

3.4. Post-Translationally Modified Sites in IDRs Transmit Biological Information

The conformational plasticity of IDRs is also exploited for the regulation of their biological activities through post-translational modifications (PTMs) [8,102,134,168] (Figure 4). IDRs are the prevalent sites of PTMs perhaps because the lack of stable structure enables easier access to modifying enzymes, as

previously proposed [134,169–172]. The PTM sites represent high-information density regions in IDR sequences that offer vastly diverse options for regulating biological functions, such as modulation of subcellular localization, protein-protein interactions, and rates of protein-synthesis [134,169–172].

From the structural point of view, PTMs can act on a broad scale, from modulating secondary structural propensities [140,173] (Figure 4Ci) and transient tertiary intramolecular contacts, to completely changing the fold of a protein [133,134] (Figure 4Cii). For instance, N-terminal acetylation of α-synuclein was shown to increase the population of a transiently formed α-helix at the N-terminus and subsequently increase the binding affinity for lipids [140,173,174] (Figure 4Ci). Importantly, this result has been confirmed in living cells by NMR spectroscopy, demonstrating how a PTM can establish a structure-function link in an IDR [175,176]. A more drastic structural impact of PTM was shown on 4EBP2: upon multisite phosphorylation, a 40-residue region of the disordered protein transitioned to a folded structure (Figure 4Cii). This structural conversion leads to the sequestration of its eIF4E binding interface, thus providing mechanistic insight into its control of translational initiation [133].

In addition to their effect on specific interactions of IDRs with other cellular components, PTMs can also regulate liquid-liquid phase separation [68,70,76,77] (Figure 4D). For instance, phosphorylation was found to promote phase separation of some IDRs, such as Tau and FMRP [68,76], while inhibiting that of others, e.g., FUS-LC [77]. Similarly, arginine methylation was found to inhibit phase separation in some contexts [68,70,72], yet promote granule formation in others [177]. The modulation of phase separation by PTMs suggest a mechanism to control both the formation and dissolution of different membranelles organelles [68,70].

Post-translationally modified sequence motifs in IDRs are sometimes positionally conserved and can manifest as information due to the reduction in the positional sequence entropy. In other instances, PTM sites are not conserved at precise positions along the sequence, but as an aggregate property [178]. This illustrates how functional information in IDRs can be encoded in the absence of positional sequence conservation [9,10], as discussed in Section 2.4.

3.5. Functional Engagements of IDRs Alter Physical Entropy in All Directions

The favorability of functional engagements of IDRs with other cellular components is dictated not only by entropy, but also by the underlying changes in free energy. The entropic contribution to such changes can go in either direction, depending on the enthalpic components, and both increases and decreases in the overall entropy can be observed in a functional context. Hence, a decrease in physical entropy cannot be used directly to extract functionally relevant information in IDR-containing systems. When separately considering the constituents of the overall entropy, e.g., conformational entropy, it is often not straightforward to experimentally evaluate (or to simulate) its changes in a functional context. Information in the form of a loss of conformational entropy upon functional interactions is intuitively expected, and, indeed, numerous reports exist of IDRs that acquire a partial or complete fold in a complex or stabilize their pre-existing transient structural propensities.

However, there are also accumulating reports of biomolecular complexes in which IDRs remain highly disordered, or, where a loss of disorder in one part of an IDR is compensated by a gain of disorder in another (Figure 4A). Hence, functional context need not always result in a reduced conformational entropy of an IDR. Therefore, like their primary amino acid sequences, the conformational ensembles of IDRs demand additional metrics of functional information that are not strictly determined by structural propensities.

Similarly, post-translational modifications can restrict the conformational entropy of IDR ensembles, relative to their free state (Figure 4C), and the concentration of IDRs in phase-separated states can reduce the configurational entropy, which is typically compensated by a gain in entropy of the solvent (Figure 4D). However, a decrease in conformational entropy is not the rule, as there are reports of order-to-disorder transitions upon post-translational modifications [179,180] or changes in pH, temperature or redox state, which can support chaperoning activities [101,181]. Thus, when function is imposed as a condition (Y in Equation (6)), the physical entropy of IDR-containing systems,

or its constituents, does not always decrease. The question therefore poses itself: what is the proper entropy function to measure information that is relevant to the biological functions of IDRs? From the information-theoretic perspective, the function should: (i) increase with an increase in uncertainty about the state of an IDR, (ii) be additive with respect to different sources of the uncertainty, and (iii) decrease when IDRs undergo functional engagements, thus directly translating to functionally relevant information. Given the current reports in the literature, most available entropic measures fall short on the criterion (iii), thus demanding a better understanding of the receivers of the information contained in IDR ensembles, and expansion of the existing or derivation of novel metrics to extract that information.

4. Conclusions and Outlook

The apparent high entropy in the sequence space of IDRs points to the ineffectiveness of the positional sequence-structure paradigm for extracting functional information stored within IDRs. To fully appreciate the extent of information encoded in IDRs, we first must bypass the conventional measures of protein fold and search for new ways of extracting information. This task remains difficult, primarily because of our incomplete understanding of the receivers of the IDR-encoded information. The receivers undoubtedly exist, as IDRs have myriad cellular roles. Nonetheless, the receivers do not have a universal mechanism for decoding information from the amino acid sequences of IDRs. While some information is obtained from positional sequence conservation, such as the position of SLiMs and some PTM sites, positionally conserved regions represent only a small fraction of an IDR sequence. Until recently, understanding of the functional contributions from the remainder of the IDR sequence was limited to individual proteins or protein classes. However, a proteome-wide approach to detect evolutionarily conserved molecular features of IDRs now enables IDR functions to be revealed directly from their sequences [10].

If the translation of genetic information to biological function is determined by thermodynamic processes for folded proteins, as codified in structure-function relationships, then it stands to reason that the thermodynamic behavior of disordered proteins also translates genetic information into other conformational ensemble-function and "transient-structure"/dynamics-function relationships. What kind of information can be extracted from IDR ensembles themselves? Measuring structural propensities in free states allows for some functionally relevant information to be extracted for a subset of IDRs from a reduction of the conformational entropy of an IDR chain. This structural information can be predictive of the binding affinities and the conformational preferences of an IDR in a complex. More interesting is perhaps the information gained as a difference between the entropy of the "off" (free) state of IDRs and the "on" (bound, post-translationally modified, or phase-separated) states. Here, a decrease in conformational entropy can accompany complex formation and PTMs, while a decrease in configurational entropy can underly phase-separation. However, the picture is complicated by alternative reports of highly disordered states of IDRs in functional complexes, and of an increase in disorder upon PTMs. Given that functional information cannot always be acquired from a reduction in conformational entropy, new metrics of information and a better understanding of receivers are needed to approach the sequence-conformational ensemble-function relationship of IDRs.

It is interesting to note the proposed role of IDRs in information transfer as linkers, effectors, and sensors that enable complex regulatory behavior in molecular signaling [182–188]. In this framework, IDRs that connect folded domains can be used as linkers to allosterically propagate information between the sensor and effector domains. IDRs that adopt a structure upon interactions can act as effectors, directly impacting the signaling outcome. IDRs can also be effective sensors, as they can alter the distribution of conformers in their ensembles, both through external signals (e.g., environmental changes, small molecule binding) and internal signals (e.g., PTMs). Finally, IDRs that retain substantial disorder in complexes have been proposed to act as high-capacity information transfer channels, given that their high conformational disorder in the bound-state leads to a large number of interface configurations [182].

Finding correlations between the sequences and conformational and phase-separation propensities of IDRs in a functionally relevant context will continue to be important for understanding disease, as around 20% of disease-associated mutations are found in IDRs [189]. Recently, a proteomics study investigated the impact of disease-associated missense mutations in human IDRs on protein-protein interactions [190]. It was shown that the gain of di-leucine motifs in disordered cytosolic tails of transmembrane proteins can result in protein mistrafficking and might represent a general disease mechanism for the cytosolic IDRs of membrane proteins [190]. In addition to expanding the understanding of disease-associated mutations in SLiMs, it will also be important to investigate the impact of mutations that fall outside of these regions on the conserved molecular features of IDRs.

In the last decades, considerable efforts have been made to derive the sequence-conformational ensemble-function relationship for IDRs. As an increasing number of IDRs are being characterized and patterns in the complicated relationship are becoming clearer, continuous progress will be ensured by nurturing communication and synergy across physical and biological sciences.

Author Contributions: Conceptualization, I.P., A.M.M. and J.D.F.-K.; formal analysis, I.P., R.M.V., A.M.M.; investigation, I.P., R.M.V., A.M.M.; resources, A.M.M. and J.D.F.-K.; data curation, I.P., R.M.V., A.M.M.; writing—original draft preparation, I.P. and A.M.M.; writing—review and editing, I.P., R.M.V., A.M.M., and J.D.F.-K.; visualization, I.P. and R.M.V.; supervision, A.M.M. and J.D.F.-K.; funding acquisition, A.M.M. and J.D.F.-K.

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

References

1. Crick, F.H.C. On Protein Syntesis. *Symp. Soc. Exp. Biol. XII* **1958**, *12*, 139–163.

2. Ebeling, W.; Volkenstein, M.V. Entropy and the evolution of biological information. *Physica A* **1990**, *163*, 398–402. [CrossRef]

3. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230. [CrossRef] [PubMed]

4. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Struct. Bioinform.* **2000**, *28*, 235–242. [CrossRef] [PubMed]

5. Śledź, P.; Caflisch, A. Protein structure-based drug design: From docking to molecular dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93–102. [CrossRef] [PubMed]

6. Dunker, A.K.; Babu, M.M.; Barbar, E.; Blackledge, M.; Bondos, S.E.; Dosztányi, Z.; Dyson, H.J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; et al. What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disord. Proteins* **2013**, *1*, e24157. [CrossRef]

7. Walsh, I.; Giollo, M.; Di Domenico, T.; Ferrari, C.; Zimmermann, O.; Tosatto, S.C.E. Comprehensive large-scale assessment of intrinsic protein disorder. *Bioinformatics* **2015**, *31*, 201–208. [CrossRef]

8. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R.J.; Daughdrill, G.W.; Dunker, A.K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D.T.; et al. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114*, 6589–6631. [CrossRef]

9. Zarin, T.; Tsai, C.N.; Nguyen Ba, A.N.; Moses, A.M. Selection maintains signaling function of a highly diverged intrinsically disordered region. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E1450–E1459. [CrossRef]

10. Zarin, T.; Strome, B.; Nguyen Ba, A.N.; Alberti, S.; Forman-Kay, J.D.; Moses, A.M. Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *bioRxiv* **2019**, 578716. [CrossRef]

11. Milles, S.; Salvi, N.; Blackledge, M.; Jensen, M.R. Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *109*, 79–100. [CrossRef] [PubMed]

12. Tompa, P.; Davey, N.E.; Gibson, T.J.; Babu, M.M. A Million peptide motifs for the molecular biologist. *Mol. Cell* **2014**, *55*, 161–169. [CrossRef] [PubMed]

13. Beltrao, P.; Albanèse, V.; Kenner, L.R.; Swaney, D.L.; Burlingame, A.; Villén, J.; Lim, W.A.; Fraser, J.S.; Frydman, J.; Krogan, N.J. Systematic functional prioritization of protein posttranslational modifications. *Cell* **2012**, *150*, 413–425. [CrossRef] [PubMed]

14. Chong, P.A.; Forman-Kay, J.D. Liquid–liquid phase separation in cellular signaling systems. *Curr. Opin. Struct. Biol.* **2016**, *41*, 180–186. [CrossRef] [PubMed]

15. Forman-Kay, J.D.; Mittag, T. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure* **2013**, *21*, 1492–1499. [CrossRef] [PubMed]

16. Forman-Kay, J.D.; Kriwacki, R.W.; Seydoux, G. Phase Separation in Biology and Disease. *J. Mol. Biol.* **2018**, *430*, 4603–4606. [CrossRef] [PubMed]

17. Alberti, S.; Carra, S. Quality Control of Membraneless Organelles. *J. Mol. Biol.* **2018**, *430*, 4711–4729. [CrossRef]

18. Babu, M.M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **2016**, *44*, 1185–1200. [CrossRef]

19. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005; ISBN 9780471241959.

20. Ebeling, W. *Physical Approaches to Biological Evolution*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 191, pp. 142–143.

21. Müller, I. *A History of Thermodynamics: The Doctrine of Energy and Entropy*; Springer: Berlin/Heidelberg, Germany, 2007; ISBN 3540462260.

22. Gibbs, J.W. *Elementary Principles in Statistical Mechanics*; Gale: New York, NY, USA, 1902; ISBN 1108017029.

23. Boltzmann, L. Über die Beziehung eines allgemeinen mechanischen Satzes zum zweiten Hauptsatze der Wärmetheorie. In *Kinetische Theorie II: Irreversible Prozesse Einführung und Originaltexte*; Brush, S.G., Ed.; Vieweg + Teubner Verlag: Wiesbaden, Germany, 1970; pp. 240–247, ISBN 978-3-322-84986-1. (In German)

24. Shannon, C.E. *The Mathematical Theory of Communication*; The University of Illinois Press: Champagne, IL, USA, 1964; Volume 14, pp. 306–317.

25. Vinga, S. Information theory applications for biological sequence analysis. *Brief. Bioinform.* **2014**, *15*, 376–389. [CrossRef]

26. Konorski, J.; Szpankowski, W. What is information? In Proceedings of the 2008 IEEE Information Theory Workshop, Porto, Portugal, 5–9 May 2008; Volume 374, pp. 269–270.

27. Wootton, J.C.; Federhen, S. Statistics of local complexity in amino acid sequences and sequence databases. *Comput. Chem.* **1993**, *17*, 149–163. [CrossRef]

28. Wootton, J.C.; Federhen, S. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **1996**, *266*, 554–571.

29. Romero, P.; Obradovic, Z.; Li, X.; Garner, E.C.; Brown, C.J.; Dunker, A.K. Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* **2001**, *42*, 38–48. [CrossRef]

30. Adami, C. Information theory in molecular biology. *Phys. Life Rev.* **2004**, *1*, 3–22. [CrossRef]

31. Adami, C. The use of information theory in evolutionary biology. *Ann. N. Y. Acad. Sci.* **2012**, *1256*, 49–65. [CrossRef]

32. Durbin, R.; Eddy, S.R.; Mitchison, G.J. *Biological Sequence Analysis, Probabilistic Models of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, UK, 1998; ISBN 9780521629713.

33. Schneider, T.D.; Stephens, R.M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **1990**, *18*, 6097–6100. [CrossRef] [PubMed]

34. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef]

35. Berg, O.G.; von Hippel, P.H. Selection of DNA binding sites by regulatory proteins. *Trends Biochem. Sci.* **1988**, *13*, 207–211. [CrossRef]

36. Schneider, T.D.; Stormo, G.D.; Gold, L.; Ehrenfeucht, A. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* **1986**, *188*, 415–431. [CrossRef]

37. Oliveira, L.; Paiva, P.B.; Paiva, A.C.M.; Vriend, G. Identification of functionally conserved residues with the use of entropy-variability plots. *Proteins Struct. Funct. Genet.* **2003**, *52*, 544–552. [CrossRef] [PubMed]

38. Lawrence, C.E.; Altschul, S.F.; Boguski, M.S.; Liu, J.S.; Neuwald, A.F.; Wootton, J.C. Detecting subtle sequence signals: A gibbs sampling strategy for multiple alignment. *Science* **1993**, *262*, 208–214. [CrossRef]

39. Dewey, T.G. Algorithmic complexity and thermodynamics of sequence-structure relationships in proteins. *Phys. Rev. E* **1997**, *56*, 4545–4552. [CrossRef]

40. Atchley, W.R.; Wollenberg, K.R.; Fitch, W.M.; Terhalle, W.; Dress, A.W. Correlations among amino acid sites in bHLH protein domains: An information theoretic analysis. *Mol. Biol. Evol.* **2000**, *17*, 164–178. [CrossRef] [PubMed]

41. Marks, D.S.; Colwell, L.J.; Sheridan, R.; Hopf, T.A.; Pagnani, A.; Zecchina, R.; Sander, C. Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* **2011**, *6*, e28766. [CrossRef] [PubMed]

42. Martin, L.C.; Gloor, G.B.; Dunn, S.D.; Wahl, L.M. Using information theory to search for co-evolving residues in proteins. *Bioinformatics* **2005**, *21*, 4116–4124. [CrossRef] [PubMed]

43. Dunn, S.D.; Wahl, L.M.; Gloor, G.B. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* **2008**, *24*, 333–340. [CrossRef]

44. Capra, J.A.; Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics* **2007**, *23*, 1875–1882. [CrossRef]

45. Hopf, T.A.; Schärfe, C.P.I.; Rodrigues, J.P.G.L.M.; Green, A.G.; Kohlbacher, O.; Sander, C.; Bonvin, A.M.J.J.; Marks, D.S. Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* **2014**, *3*, e03430. [CrossRef]

46. Kimura, M. Natural selection as the process of accumulating genetic information in adaptive evolution. *Genet. Res.* **1961**, *2*, 127. [CrossRef]

47. Moses, A.M.; Chiang, D.Y.; Kellis, M.; Lander, E.S.; Eisen, M.B. Position specific variation in the rate of evolution in transcription factor binding sites. *BMC Evol. Biol.* **2003**, *3*, 19. [CrossRef]

48. Moses, A.M.; Durbin, R. Inferring selection on amino acid preference in protein domains. *Mol. Biol. Evol.* **2009**, *26*, 527–536. [CrossRef]

49. Koonin, E.V. The meaning of biological information. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150065. [CrossRef] [PubMed]

50. Colak, R.; Kim, T.H.; Michaut, M.; Sun, M.; Irimia, M.; Bellay, J.; Myers, C.L.; Blencowe, B.J.; Kim, P.M. Distinct Types of Disorder in the Human Proteome: Functional Implications for Alternative Splicing. *PLoS Comput. Biol.* **2013**, *9*, e1003030. [CrossRef] [PubMed]

51. Vacic, V.; Oldfield, C.J.; Mohan, A.; Radivojac, P.; Cortese, M.S.; Uversky, V.N.; Dunker, A.K. Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.* **2007**, *6*, 2351–2366. [CrossRef] [PubMed]

52. Cumberworth, A.; Lamour, G.; Babu, M.M.; Gsponer, J. Promiscuity as a functional trait: Intrinsically disordered regions as central players of interactomes. *Biochem. J.* **2013**, *454*, 361–369. [CrossRef] [PubMed]

53. Davey, N.E.; Van Roey, K.; Weatheritt, R.J.; Toedt, G.; Uyar, B.; Altenberg, B.; Budd, A.; Diella, F.; Dinkel, H.; Gibson, T.J. Attributes of short linear motifs. *Mol. Biosyst.* **2012**, *8*, 268–281. [CrossRef] [PubMed]

54. Nguyen Ba, A.N.; Yeh, B.J.; Van Dyk, D.; Davidson, A.R.; Andrews, B.J.; Weiss, E.L.; Moses, A.M. Proteome-wide discovery of evolutionary conserved sequences in disordered regions. *Sci. Signal.* **2012**, *5*, rs1. [CrossRef]

55. Gouw, M.; Michael, S.; Sámano-Sánchez, H.; Kumar, M.; Zeke, A.; Lang, B.; Bely, B.; Chemes, L.B.; Davey, N.E.; Deng, Z.; et al. The eukaryotic linear motif resource—2018 update. *Nucleic Acids Res.* **2018**, *46*, D428–D434. [CrossRef]

56. Lovell, S.C. Are non-functional, unfolded proteins ('junk proteins') common in the genome? *FEBS Lett.* **2003**, *554*, 237–239. [CrossRef]

57. Good, M.C.; Zalatan, J.G.; Lim, W.A. Scaffold proteins: Hubs for controlling the flow of cellular information. *Science* **2011**, *332*, 680–686. [CrossRef]

58. Ravarani, C.N.; Erkina, T.Y.; De Baets, G.; Dudman, D.C.; Erkine, A.M.; Babu, M.M. High-throughput discovery of functional disordered regions: Investigation of transactivation domains. *Mol. Syst. Biol.* **2018**, *14*, e8190. [CrossRef]

59. Daughdrill, G.W.; Narayanaswami, P.; Gilmore, S.H.; Belczyk, A.; Brown, C.J. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J. Mol. Evol.* **2007**, *65*, 277–288. [CrossRef] [PubMed]

60. Lemas, D.; Lekkas, P.; Ballif, B.A.; Vigoreaux, J.O. Intrinsic disorder and multiple phosphorylations constrain the evolution of the flightin N-terminal region. *J. Proteom.* **2016**, *135*, 191–200. [CrossRef] [PubMed]

61. Banani, S.F.; Lee, H.O.; Hyman, A.A.; Rosen, M.K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 285–298. [CrossRef] [PubMed]

62. Protter, D.S.W.; Rao, B.S.; Van Treeck, B.; Lin, Y.; Mizoue, L.; Rosen, M.K.; Parker, R. Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep.* **2018**, *22*, 1401–1412. [CrossRef] [PubMed]

63. Mittag, T.; Parker, R. Multiple Modes of Protein–Protein Interactions Promote RNP Granule Assembly. *J. Mol. Biol.* **2018**, *430*, 4636–4649. [CrossRef] [PubMed]

64. Boeynaems, S.; Alberti, S.; Fawzi, N.L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **2018**, *28*, 420–435. [CrossRef] [PubMed]

65. Chong, P.A.; Vernon, R.M.; Forman-Kay, J.D. RGG/RG Motif Regions in RNA Binding and Phase Separation. *J. Mol. Biol.* **2018**, *430*, 4650–4665. [CrossRef]

66. Martin, E.W.; Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **2018**, *57*, 2478–2487. [CrossRef]

67. Uversky, V.N. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.* **2017**, *44*, 18–30. [CrossRef]

68. Tsang, B.; Arsenault, J.; Vernon, R.M.; Lin, H.; Sonenberg, N.; Wang, L.Y.; Bah, A.; Forman-Kay, J.D. Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4218–4227. [CrossRef]

69. Brady, J.P.; Farber, P.J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T.J.; Chan, H.S.; Baldwin, A.J.; Forman-Kay, J.D.; et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E8194–E8203. [CrossRef] [PubMed]

70. Nott, T.J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T.D.; Bazett-Jones, D.P.; Pawson, T.; Forman-Kay, J.D.; et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **2015**, *57*, 936–947. [CrossRef] [PubMed]

71. Conicella, A.E.; Zerze, G.H.; Mittal, J.; Fawzi, N.L. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **2016**, *24*, 1537–1549. [CrossRef] [PubMed]

72. Ryan, V.H.; Dignon, G.L.; Zerze, G.H.; Chabata, C.V.; Silva, R.; Conicella, A.E.; Amaya, J.; Burke, K.A.; Mittal, J.; Fawzi, N.L. Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol. Cell* **2018**, *69*, 465–479.e7. [CrossRef] [PubMed]

73. Xiang, S.; Kato, M.; Wu, L.C.; Lin, Y.; Ding, M.; Zhang, Y.; Yu, Y.; McKnight, S.L. The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei. *Cell* **2015**, *163*, 829–839. [CrossRef] [PubMed]

74. Banani, S.F.; Rice, A.M.; Peeples, W.B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M.K. Compositional Control of Phase-Separated Cellular Bodies. *Cell* **2016**, *166*, 651–663. [CrossRef] [PubMed]

75. Burke, K.A.; Janke, A.M.; Rhine, C.L.; Fawzi, N.L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* **2015**, *60*, 231–241. [CrossRef]

76. Ambadipudi, S.; Biernat, J.; Riedel, D.; Mandelkow, E.; Zweckstetter, M. Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun.* **2017**, *8*, 275. [CrossRef]

77. Murray, D.T.; Kato, M.; Lin, Y.; Thurber, K.R.; Hung, I.; McKnight, S.L.; Tycko, R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **2017**, *171*, 615–627.e16. [CrossRef]

78. Murakami, T.; Qamar, S.; Lin, J.Q.; Schierle, G.S.K.; Rees, E.; Miyashita, A.; Costa, A.R.; Dodd, R.B.; Chan, F.T.S.; Michel, C.H.; et al. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **2015**, *88*, 678–690. [CrossRef]

79. Patel, A.; Lee, H.O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M.Y.; Stoynov, S.; Mahamid, J.; Saha, S.; Franzmann, T.M.; et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162*, 1066–1077. [CrossRef] [PubMed]

80. Mar Albà, M.; Santibáñez-Koref, M.F.; Hancock, J.M. Amino acid reiterations in yeast are overrepresented in particular classes of proteins and show evidence of a slippage-like mutational process. *J. Mol. Evol.* **1999**, *49*, 789–797. [CrossRef] [PubMed]

81. Albà, M.; Tompa, P.; Veitia, R. Amino acid repeats and the structure and evolution of proteins. *Genome Dyn.* **2007**, *3*, 119–130. [PubMed]

82. Morgulis, A.; Gertz, E.M.; Schäffer, A.A.; Agarwala, R. A Fast and Symmetric DUST Implementation to Mask Low-Complexity DNA Sequences. *J. Comput. Biol.* **2006**, *13*, 1028–1040. [CrossRef] [PubMed]

83. Boratyn, G.M.; Camacho, C.; Cooper, P.S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T.L.; Matten, W.T.; McGinnis, S.D.; Merezhuk, Y.; et al. BLAST: A more efficient report with usability improvements. *Nucleic Acids Res.* **2013**, *41*, W29–W33. [CrossRef] [PubMed]

84. Vernon, R.M.; Chong, P.A.; Tsang, B.; Kim, T.H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J.D. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **2018**, *7*, e31486. [CrossRef] [PubMed]

85. Kato, M.; McKnight, S.L. Cross-β polymerization of low complexity sequence domains. *Cold Spring Harb. Perspect. Biol.* **2017**, *9*, a023598. [CrossRef]

86. Boeynaems, S.; Bogaert, E.; Van Damme, P.; Van Den Bosch, L. Inside out: The role of nucleocytoplasmic transport in ALS and FTLD. *Acta Neuropathol.* **2016**, *132*, 159–173. [CrossRef]

87. Kato, M.; Han, T.W.; Xie, S.; Shi, K.; Du, X.; Wu, L.C.; Mirzaei, H.; Goldsmith, E.J.; Longgood, J.; Pei, J.; et al. Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **2012**, *149*, 753–767. [CrossRef]

88. Hennig, S.; Kong, G.; Mannen, T.; Sadowska, A.; Kobelke, S.; Blythe, A.; Knott, G.J.; Iyer, S.S.; Ho, D.; Newcombe, E.A.; et al. Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **2015**, *210*, 529–539. [CrossRef]

89. Franzmann, T.M.; Alberti, S. Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *J. Biol. Chem.* **2019**, *294*, 7128–7136. [CrossRef] [PubMed]

90. Wang, J.; Choi, J.M.; Holehouse, A.S.; Lee, H.O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 688–699.e16. [CrossRef] [PubMed]

91. Hughes, M.P.; Sawaya, M.R.; Boyer, D.R.; Goldschmidt, L.; Rodriguez, J.A.; Cascio, D.; Chong, L.; Gonen, T.; Eisenberg, D.S. Atomic structures of low-complexity protein segments reveal kinked b sheets that assemble networks. *Science* **2018**, *359*, 698–701. [CrossRef] [PubMed]

92. Lancaster, A.K.; Nutter-Upham, A.; Lindquist, S.; King, O.D. PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **2014**, *30*, 2501–2502. [CrossRef] [PubMed]

93. Bolognesi, B.; Gotor, N.L.; Dhar, R.; Cirillo, D.; Baldrighi, M.; Tartaglia, G.G.; Lehner, B. A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep.* **2016**, *16*, 222–231. [CrossRef]

94. Brangwynne, C.P.; Tompa, P.; Pappu, R.V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **2015**, *11*, 899–904. [CrossRef]

95. Lin, Y.H.; Forman-Kay, J.D.; Chan, H.S. Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. *Biochemistry* **2018**, *57*, 2499–2508. [CrossRef]

96. Lin, Y.H.; Song, J.; Forman-Kay, J.D.; Chan, H.S. Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *J. Mol. Liq.* **2017**, *228*, 176–193. [CrossRef]

97. Garner, E.; Cannon, P.; Romero, P.; Obradovic, Z.; Dunker, A.K. Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization. *Genome Inform. Ser. Workshop Genome Inform.* **1998**, *9*, 201–213.

98. Uversky, V.N. The alphabet of intrinsic disorder. *Intrinsically Disord. Proteins* **2013**, *1*, e24684. [CrossRef]

99. Romero, P.; Obradovic, Z.; Kissinger, C.R.; Villafranca, J.E.; Garner, E.; Guilliot, S.; Dunker, A.K. Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* **1998**, 437–448.

100. Meng, F.; Uversky, V.N.; Kurgan, L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell. Mol. Life Sci.* **2017**, *74*, 3069–3090. [CrossRef] [PubMed]

101. Flock, T.; Weatheritt, R.J.; Latysheva, N.S.; Babu, M.M. Controlling entropy to tune the functions of intrinsically disordered regions. *Curr. Opin. Struct. Biol.* **2014**, *26*, 62–72. [CrossRef] [PubMed]

102. Latysheva, N.S.; Flock, T.; Weatheritt, R.J.; Chavali, S.; Babu, M.M. How do disordered regions achieve comparable functions to structured domains? *Protein Sci.* **2015**, *24*, 909–922. [CrossRef] [PubMed]

103. Mittag, T.; Kay, L.E.; Forman-Kaya, J.D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **2010**, *23*, 105–116. [CrossRef]

104. Heller, G.T.; Sormanni, P.; Vendruscolo, M. Targeting disordered proteins with small molecules using entropy. *Trends Biochem. Sci.* **2015**, *40*, 491–496. [CrossRef]

105. Mantsyzov, A.B.; Shen, Y.; Lee, J.H.; Hummer, G.; Bax, A. MERA: A webserver for evaluating backbone torsion angle distributions in dynamic and disordered proteins from NMR data. *J. Biomol. NMR* **2015**, *63*, 85–95. [CrossRef]

106. Mantsyzov, A.B.; Maltsev, A.S.; Ying, J.; Shen, Y.; Hummer, G.; Bax, A. A maximum entropy approach to the study of residue-specific backbone angle distributions in α-synuclein, an intrinsically disordered protein. *Protein Sci.* **2014**, *23*, 1275–1290. [CrossRef]

107. Schneider, R.; Maurin, D.; Communie, G.; Kragelj, J.; Hansen, D.F.; Ruigrok, R.W.H.; Jensen, M.R.; Blackledge, M. Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. *J. Am. Chem. Soc.* **2015**, *137*, 1220–1229. [CrossRef]

108. Iešmantavičius, V.; Jensen, M.R.; Ozenne, V.; Blackledge, M.; Poulsen, F.M.; Kjaergaard, M. Modulation of the intrinsic helix propensity of an intrinsically disordered protein reveals long-range helix-helix interactions. *J. Am. Chem. Soc.* **2013**, *135*, 10155–10163. [CrossRef]

109. Huggins, M.L. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, USA, 1953; Volume 76.

110. Karplus, M.; Kushick, J.N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981**, *14*, 325–332. [CrossRef]

111. Karplus, M.; Ichiye, T.; Pettitt, B.M. Configurational entropy of native proteins. *Biophys. J.* **1987**, *52*, 1083–1085. [CrossRef]

112. Leavitt, S.; Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **2001**, *11*, 560–566. [CrossRef]

113. Wand, A.J.; Sharp, K.A. Measuring Entropy in Molecular Recognition by Proteins. *Annu. Rev. Biophys.* **2018**, *47*, 41–61. [CrossRef] [PubMed]

114. Dyson, H.J.; Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208. [CrossRef] [PubMed]

115. Cordeiro, T.N.; Herranz-Trillo, F.; Urbanek, A.; Estaña, A.; Cortés, J.; Sibille, N.; Bernadó, P. Structural characterization of highly flexible proteins by small-angle scattering. In *Advances in Experimental Medicine and Biology*; Springer Nature: New York, NY, USA, 2017; Volume 1009, pp. 107–129.

116. Schuler, B.; Soranno, A.; Hofmann, H.; Nettels, D. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annu. Rev. Biophys.* **2016**, *45*, 207–231. [CrossRef] [PubMed]

117. Le Breton, N.; Martinho, M.; Mileo, E.; Etienne, E.; Gerbaud, G.; Guigliarelli, B.; Belle, V. Exploring intrinsically disordered proteins using site-directed spin labeling electron paramagnetic resonance spectroscopy. *Front. Mol. Biosci.* **2015**, *2*, 21. [CrossRef] [PubMed]

118. Allison, J.R. Using simulation to interpret experimental data in terms of protein conformational ensembles. *Curr. Opin. Struct. Biol.* **2017**, *43*, 79–87. [CrossRef]

119. Best, R.B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2017**, *42*, 147–154. [CrossRef]

120. Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining Experiments and Simulations Using the Maximum Entropy Principle. *PLoS Comput. Biol.* **2014**, *10*, e1003406. [CrossRef]

121. Sormanni, P.; Piovesan, D.; Heller, G.T.; Bonomi, M.; Kukic, P.; Camilloni, C.; Fuxreiter, M.; Dosztanyi, Z.; Pappu, R.V.; Babu, M.M.; et al. Simultaneous quantification of protein order and disorder. *Nat. Chem. Biol.* **2017**, *13*, 339–342. [CrossRef] [PubMed]

122. Sekhar, A.; Kay, L.E. An NMR View of Protein Dynamics in Health and Disease. *Annu. Rev. Biophys.* **2019**, *48*, 297–319. [CrossRef] [PubMed]

123. Schneider, R.; Blackledge, M.; Jensen, M.R. Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Curr. Opin. Struct. Biol.* **2019**, *54*, 10–18. [CrossRef] [PubMed]

124. Jensen, M.R.; Zweckstetter, M.; Huang, J.; Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* **2014**, *114*, 6632–6660. [CrossRef] [PubMed]

125. Jensen, M.R.; Ruigrok, R.W.H.; Blackledge, M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.* **2013**, *23*, 426–435. [CrossRef]

126. Bhowmick, A.; Brookes, D.H.; Yost, S.R.; Dyson, H.J.; Forman-Kay, J.D.; Gunter, D.; Head-Gordon, M.; Hura, G.L.; Pande, V.S.; Wemmer, D.E.; et al. Finding Our Way in the Dark Proteome. *J. Am. Chem. Soc.* **2016**, *138*, 9730–9742. [CrossRef] [PubMed]

127. Nielsen, J.T.; Mulder, F.A.A. POTENCI: Prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. *J. Biomol. NMR* **2018**, *70*, 141–165. [CrossRef]

128. Krzeminski, M.; Marsh, J.A.; Neale, C.; Choy, W.Y.; Forman-Kay, J.D. Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* **2013**, *29*, 398–399. [CrossRef]

129. Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M.R.; Blackledge, M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918. [CrossRef]

130. Salvi, N.; Abyzov, A.; Blackledge, M. Atomic resolution conformational dynamics of intrinsically disordered proteins from NMR spin relaxation. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *102–103*, 43–60. [CrossRef]

131. Charlier, C.; Bouvignies, G.; Pelupessy, P.; Walrant, A.; Marquant, R.; Kozlov, M.; De Ioannes, P.; Bolik-Coulon, N.; Sagan, S.; Cortes, P.; et al. Structure and Dynamics of an Intrinsically Disordered Protein Region That Partially Folds upon Binding by Chemical-Exchange NMR. *J. Am. Chem. Soc.* **2017**, *139*, 12219–12227. [CrossRef] [PubMed]

132. Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355–360. [CrossRef] [PubMed]

133. Bah, A.; Vernon, R.M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L.E.; Forman-Kay, J.D. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2015**, *519*, 106–109. [CrossRef] [PubMed]

134. Bah, A.; Forman-Kay, J.D. Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.* **2016**, *291*, 6696–6705. [CrossRef] [PubMed]

135. Heller, G.T.; Aprile, F.A.; Bonomi, M.; Camilloni, C.; De Simone, A.; Vendruscolo, M. Sequence Specificity in the Entropy-Driven Binding of a Small Molecule and a Disordered Peptide. *J. Mol. Biol.* **2017**, *429*, 2772–2779. [CrossRef] [PubMed]

136. Heller, G.T.; Aprile, F.A.; Vendruscolo, M. Methods of probing the interactions between small molecules and disordered proteins. *Cell. Mol. Life Sci.* **2017**, *74*, 3225–3243. [CrossRef]

137. Wright, P.E.; Dyson, H.J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38. [CrossRef]

138. Fuxreiter, M. Fuzziness in Protein Interactions—A Historical Perspective. *J. Mol. Biol.* **2018**, *430*, 2278–2287. [CrossRef]

139. Mittag, T.; Orlicky, S.; Choy, W.-Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L.E.; Tyers, M.; Forman-Kay, J.D. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17772–17777. [CrossRef]

140. Maltsev, A.S.; Ying, J.; Bax, A. Impact of N-terminal acetylation of α-synuclein on its random coil and lipid binding properties. *Biochemistry* **2012**, *51*, 5004–5013. [CrossRef]

141. Marsh, J.A.; Singh, V.K.; Jia, Z.; Forman-Kay, J.D. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci.* **2006**, *15*, 2795–2804. [CrossRef] [PubMed]

142. Camilloni, C.; De Simone, A.; Vranken, W.F.; Vendruscolo, M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **2012**, *51*, 2224–2231. [CrossRef] [PubMed]

143. Bernadó, P.; Bertoncini, C.W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. Defining long-range order and local disorder in native α-synuclein using residual dipolar couplings. *J. Am. Chem. Soc.* **2005**, *127*, 17968–17969. [CrossRef] [PubMed]

144. Allison, J.R.; Varnai, P.; Dobson, C.M.; Vendruscolo, M. Determination of the free energy landscape of α-synuclein using spin label nuclear magnetic resonance measurements. *J. Am. Chem. Soc.* **2009**, *131*, 18314–18326. [CrossRef]

145. Iešmantavičius, V.; Dogan, J.; Jemth, P.; Teilum, K.; Kjaergaard, M. Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angew. Chemie Int. Ed.* **2014**, *53*, 1548–1551. [CrossRef]

146. Kim, D.-H.; Han, K.-H. Transient Secondary Structures as General Target-Binding Motifs in Intrinsically Disordered Proteins. *Int. J. Mol. Sci.* **2018**, *19*, 3614. [CrossRef]

147. Marsh, J.A.; Dancheck, B.; Ragusa, M.J.; Allaire, M.; Forman-Kay, J.D.; Peti, W. Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. *Structure* **2010**, *18*, 1094–1103. [CrossRef]

148. Borcherds, W.; Theillet, F.X.; Katzer, A.; Finzel, A.; Mishall, K.M.; Powell, A.T.; Wu, H.; Manieri, W.; Dieterich, C.; Selenko, P.; et al. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat. Chem. Biol.* **2014**, *10*, 1000–1002. [CrossRef]

149. Krieger, J.M.; Fusco, G.; Lewitzky, M.; Simister, P.C.; Marchant, J.; Camilloni, C.; Feller, S.M.; De Simone, A. Conformational recognition of an intrinsically disordered protein. *Biophys. J.* **2014**, *106*, 1771–1779. [CrossRef]

150. Arai, M.; Sugase, K.; Dyson, H.J.; Wright, P.E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9614–9619. [CrossRef]

151. Crabtree, M.D.; Borcherds, W.; Poosapati, A.; Shammas, S.L.; Daughdrill, G.W.; Clarke, J. Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein: Target Affinity by Altering the Lifetime of the Bound Complex. *Biochemistry* **2017**, *56*, 2379–2384. [CrossRef] [PubMed]

152. Sugase, K.; Dyson, H.J.; Wright, P.E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007**, *447*, 1021–1025. [CrossRef] [PubMed]

153. Sormanni, P.; Camilloni, C.; Fariselli, P.; Vendruscolo, M. The s2D method: Simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. *J. Mol. Biol.* **2015**, *427*, 982–996. [CrossRef] [PubMed]

154. Uversky, V.N. Multitude of binding modes attainable by intrinsically disordered proteins: A portrait gallery of disorder-based complexes. *Chem. Soc. Rev.* **2011**, *40*, 1623–1634. [CrossRef] [PubMed]

155. Dyson, H.J.; Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60. [CrossRef]

156. Gianni, S.; Dogan, J.; Jemth, P. Coupled binding and folding of intrinsically disordered proteins: What can we learn from kinetics? *Curr. Opin. Struct. Biol.* **2016**, *36*, 18–24. [CrossRef] [PubMed]

157. Borgia, A.; Borgia, M.B.; Bugge, K.; Kissling, V.M.; Heidarsson, P.O.; Fernandes, C.B.; Sottini, A.; Soranno, A.; Buholzer, K.J.; Nettels, D.; et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555*, 61–66. [CrossRef] [PubMed]

158. Delaforge, E.; Kragelj, J.; Tengo, L.; Palencia, A.; Milles, S.; Bouvignies, G.; Salvi, N.; Blackledge, M.; Jensen, M.R. Deciphering the Dynamic Interaction Profile of an Intrinsically Disordered Protein by NMR Exchange Spectroscopy. *J. Am. Chem. Soc.* **2018**, *140*, 1148–1158. [CrossRef] [PubMed]

159. Lindström, I.; Dogan, J. Dynamics, Conformational Entropy, and Frustration in Protein-Protein Interactions Involving an Intrinsically Disordered Protein Domain. *ACS Chem. Biol.* **2018**, *13*, 1218–1227. [CrossRef] [PubMed]

160. Yang, D.; Kay, L.E. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: Application to protein folding. *J. Mol. Biol.* **1996**, *263*, 369–382. [CrossRef] [PubMed]

161. Frederick, K.K.; Marlow, M.S.; Valentine, K.G.; Wand, A.J. Conformational entropy in molecular recognition by proteins. *Nature* **2007**, *448*, 325–329. [CrossRef] [PubMed]

162. Tzeng, S.R.; Kalodimos, C.G. Protein activity regulation by conformational entropy. *Nature* **2012**, *488*, 236–240. [CrossRef] [PubMed]

163. Alderson, T.R.; Lee, J.H.; Charlier, C.; Ying, J.; Bax, A. Propensity for cis-Proline Formation in Unfolded Proteins. *ChemBioChem* **2018**, *19*, 37–42. [CrossRef] [PubMed]

164. Baker, J.M.R.; Hudson, R.P.; Kanelis, V.; Choy, W.Y.; Thibodeau, P.H.; Thomas, P.J.; Forman-Kay, J.D. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat. Struct. Mol. Biol.* **2007**, *14*, 738–745. [CrossRef] [PubMed]

165. Kragelj, J.; Palencia, A.; Nanao, M.H.; Maurin, D.; Bouvignies, G.; Blackledge, M.; Jensen, M.R. Structure and dynamics of the MKK7–JNK signaling complex. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3409–3414. [CrossRef] [PubMed]

166. Martinez, A.I.C.; Weinhäupl, K.; Lee, W.K.; Wolff, N.A.; Storch, B.; Zerko, S.; Konrat, R.; Kozminski, W.; Breuker, K.; Thévenod, F.; et al. Biochemical and structural characterization of the interaction between the siderocalin NGAL/LCN2 (Neutrophil Gelatinase-associated lipocalin/lipocalin 2) and the N-terminal domain of its endocytic receptor SLC22A17. *J. Biol. Chem.* **2016**, *291*, 2917–2930. [CrossRef]

167. Ferreon, J.C.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13260–13265. [CrossRef]

168. Martin, E.W.; Holehouse, A.S.; Grace, C.R.; Hughes, A.; Pappu, R.V.; Mittag, T. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *J. Am. Chem. Soc.* **2016**, *138*, 15323–15335. [CrossRef]

169. Oldfield, C.J.; Dunker, A.K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **2014**, *83*, 553–584. [CrossRef]

170. Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049. [CrossRef]

171. Radivojac, P.; Vacic, V.; Haynes, C.; Cocklin, R.R.; Mohan, A.; Heyen, J.W.; Goebl, M.G.; Iakoucheva, L.M. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 365–380. [CrossRef] [PubMed]

172. Pejaver, V.; Hsu, W.L.; Xin, F.; Dunker, A.K.; Uversky, V.N.; Radivojac, P. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci.* **2014**, *23*, 1077–1093. [CrossRef] [PubMed]

173. Kang, L.; Moriarty, G.M.; Woods, L.A.; Ashcroft, A.E.; Radford, S.E.; Baum, J. N-terminal acetylation of α-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer. *Protein Sci.* **2012**, *21*, 911–917. [CrossRef]

174. Alderson, T.R.; Markley, J.L. Biophysical characterization of α-synuclein and its controversial structure. *Intrinsically Disord. Proteins* **2013**, *1*, 18–39. [CrossRef] [PubMed]

175. Theillet, F.X.; Binolfi, A.; Bekei, B.; Martorana, A.; Rose, H.M.; Stuiver, M.; Verzini, S.; Lorenz, D.; Van Rossum, M.; Goldfarb, D.; et al. Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature* **2016**, *530*, 45–50. [CrossRef] [PubMed]

176. Binolfi, A.; Limatola, A.; Verzini, S.; Kosten, J.; Theillet, F.X.; May Rose, H.; Bekei, B.; Stuiver, M.; Van Rossum, M.; Selenko, P. Intracellular repair of oxidation-damaged α-synuclein fails to target C-terminal modification sites. *Nat. Commun.* **2016**, *7*, 10251. [CrossRef] [PubMed]

177. Arribas-Layton, M.; Dennis, J.; Bennett, E.J.; Damgaard, C.K.; Lykke-Andersen, J. The C-Terminal RGG Domain of Human Lsm4 Promotes Processing Body Formation Stimulated by Arginine Dimethylation. *Mol. Cell. Biol.* **2016**, *36*, 2226–2235. [CrossRef] [PubMed]

178. Landry, C.R.; Freschi, L.; Zarin, T.; Moses, A.M. Turnover of protein phosphorylation evolving under stabilizing selection. *Front. Genet.* **2014**, *5*, 245. [CrossRef]

179. Johnson, L.N.; Lewis, R.J. Structural basis for control by phosphorylation. *Chem. Rev.* **2001**, *101*, 2209–2242. [CrossRef]

180. Darling, A.L.; Uversky, V.N. Intrinsic disorder and posttranslational modifications: The darker side of the biological dark matter. *Front. Genet.* **2018**, *9*, 158. [CrossRef]

181. Alderson, T.R.; Roche, J.; Gastall, H.Y.; Dias, D.M.; Pritišanac, I.; Ying, J.; Bax, A.; Benesch, J.L.P.; Baldwin, A.J. Local unfolding of the HSP27 monomer regulates chaperone activity. *Nat. Commun.* **2019**, *10*, 1068. [CrossRef] [PubMed]

182. Arbesú, M.; Iruela, G.; Fuentes, H.; Teixeira, J.M.C.; Pons, M. Intramolecular Fuzzy Interactions Involving Intrinsically Disordered Domains. *Front. Mol. Biosci.* **2018**, *5*, 39. [CrossRef] [PubMed]

183. Tompa, P. Multisteric regulation by structural disorder in modular signaling proteins: An extension of the concept of allostery. *Chem. Rev.* **2014**, *114*, 6715–6732. [CrossRef] [PubMed]

184. Hilser, V.J.; Thompson, E.B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8311–8315. [CrossRef] [PubMed]

185. Motlagh, H.N.; Hilser, V.J. Agonism/antagonism switching in allosteric ensembles. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 4134–4139. [CrossRef] [PubMed]

186. Li, J.; Hilser, V.J. Assessing Allostery in Intrinsically Disordered Proteins with Ensemble Allosteric Model. *Methods Enzymol.* **2018**, *611*, 531–557. [PubMed]

187. Zhang, L.; Li, M.; Liu, Z. A comprehensive ensemble model for comparing the allosteric effect of ordered and disordered proteins. *PLoS Comput. Biol.* **2018**, *14*, e1006393. [CrossRef] [PubMed]

188. Follis, A.V.; Llambi, F.; Kalkavan, H.; Yao, Y.; Phillips, A.H.; Park, C.G.; Marassi, F.M.; Green, D.R.; Kriwacki, R.W. Regulation of apoptosis by an intrinsically disordered region of Bcl-xL. *Nat. Chem. Biol.* **2018**, *14*, 458–465. [CrossRef]

189. Vacic, V.; Markwick, P.R.L.; Oldfield, C.J.; Zhao, X.; Haynes, C.; Uversky, V.N.; Iakoucheva, L.M. Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder. *PLoS Comput. Biol.* **2012**, *8*, e1002709. [CrossRef]

190. Meyer, K.; Kirchner, M.; Uyar, B.; Cheng, J.Y.; Russo, G.; Hernandez-Miranda, L.R.; Szymborska, A.; Zauber, H.; Rudolph, I.M.; Willnow, T.E.; et al. Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. *Cell* **2018**, *175*, 239–253.e17. [CrossRef]