

Introducing Protein Intrinsic Disorder

Johnny Habchi,¹ Peter Tompa^{2,3*}, Sonia Longhi^{1*} and Vladimir N. Uversky^{4,5*}

¹*CNRS and Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques (AFMB) UMR 7257, Marseille, France*

²*VIB Department of Structural Biology, Vrije Universiteit Brussel, Brussels, Belgium*

³*Institute of Enzymology; Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary*

⁴*Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida, USA*

⁵*Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Russia*

* corresponding authors

Table of Contents

1. Introducing protein intrinsic disorder phenomenon	2
2. Historical overview: Role of bioinformatics in establishing the IDP field	3
3. Computational means for IDPs analysis	3
3.1. Sequence peculiarities of IDPs	3
3.2. Computational tools for disorder prediction	4
4. Multiparametric approaches for assessing intrinsic disorder: Experimental view on conformational ensembles and induced folding transitions	7
4.1. NMR.....	7
4.2. Further quantitative techniques for characterizing structural disorder.....	9
4.3. Complementary methods for studying protein disorder.....	10
4.4. Analyzing disorder in a living cell	11
4.4.1. In vivo existence of protein disorder	11
4.4.2. Survival of IDPs in the cell	12
5. Structural and conformational characteristics of IDPs	13
5.1. IDP structure from the polymeric physics viewpoint.....	13
5.2. Evaluating residual structure in IDPs.....	14
5.3. IDPs as conformational ensembles	14
5.4. Electrostatic and charge effects of compaction of the structural ensemble of IDPs ...	15
5.5. IDPs, their environment and functional misfolding	16
6. Intrinsic disorder and protein function	17
6.1. Functional diversity of IDPs through the prism of structural plasticity	17
6.2. Short recognition motifs in the interactions of IDPs. Interaction networks, one-to-many and many-to-one binding.....	19

6.3. Variability of binding modes attainable by disordered proteins: folding before or after binding.....	21
6.4. Dynamic or fuzzy complexes	23
6.5. IDPs and posttranslational modifications	23
6.6. Alternative splicing in IDP function and dysfunction	24
6.7. Intrinsic disorder and allosteric coupling in proteins.....	25
6.8. Multi-valent interactions and phase-separation of IDPs	26
7. Cellular biology of IDPs. An overview.....	26
7.1. Abundance of intrinsic disorder in various proteomes.....	26
7.2. Involvement of structural disorder in biological processes	27
8. Distinctive features of IDP evolution.....	28
9. Intrinsic disorder and human diseases.....	29
9.1. Correlation of disorder and diseases	29
9.2. IDPs as drug targets.....	30
Abbreviations	30

1. Introducing protein intrinsic disorder phenomenon

Proteins are the major component of the living cell. They play crucial roles in the maintenance of life, and their dysfunctions are known to cause different pathologies. One of the best-understood function of proteins is catalysis (i.e. enzymatic activity), which attracted the major attention in the early days of protein science and led to the elaboration of the “lock and key” model by Fisher. Central to this model is the notion that the correct shape of the substrate can fit into the active site of the enzyme for enabling an efficient and specific catalysis, as observed for enzymes that hydrolyze β - but not α -glycosidic bonds.¹ Throughout the 20th century, tens of thousands of structures have been solved and deposited in the Protein Data Bank (PDB), supporting again the necessity of a 3-D structure for functionality. The explanatory power of these 3-D structures continued to reinforce the static view of protein structure that remained unquestioned. In contrast to this view, already in 1958, Koshland suggested the “induced-fit” model based on the observations that some enzymes could act on differently shaped substrates and hence a degree of flexibility is inevitable in function.²

Furthermore, questions such as “what does the missing electron density in most of the deposited structures in the PDB correspond to?”, “why some proteins are highly sensitive *in vitro* to proteolysis?”, and “why some proteins possess a particular behavior during the purification process?” further opened eyes towards flexibility and focused attention on proteins basically distinct from well-known globular proteins. A combined answer to these questions ruled out flexible proteins as artifacts and enlightened the “dark side” of Structural Biology, that of disordered proteins (i.e. proteins that lack 3-D structure). This review provides a detailed description of those proteins that broke the protein rules by their inherent flexibility and peculiar features, now generally termed “intrinsically disordered” proteins (IDPs) or regions (IDRs). This review is intended as a general introduction to the series of thematic reviews in this special issue of Chemical Reviews on IDPs. Our key message is that IDPs possess no well-defined 3-D structure but rather adopt an ensemble of

conformations in solution, yet they are functional. A literature survey is provided by summarizing the main aspects of IDPs that led to an exponential increase of interest in these proteins. Moreover, through the different parts of this review, biochemical and biophysical approaches that are frequently used to assess intrinsic disorder are detailed.

2. Historical overview: Role of bioinformatics in establishing the IDP field

It is now half a century since the first crystal structure of a protein (e.g. the atomic structure of myoglobin determined in 1958 by John Kendrew³) was published leading to the “birth” of Structural Biology. Soon after, the field grew rapidly with a dozen of atomic structures of proteins being published by the early 1970’s, leading to the establishment of the predecessor of PDB, the Brookhaven Data Bank.^{4,5} A well-folded, albeit dynamic, structure was thought to be the hallmark of protein function. This view was also built on the success of the previous half a century of the lock-and-key view explaining the specificity of enzymes (Fig. 1A) and complementarity of antibody to antigen structure. The classical structure-function paradigm is centered on the idea that protein function depends on a well-defined three-dimensional (3-D) structure.

Indeed, enzymes have a rather well-defined binding pocket for the formation of an enzyme-substrate (ES) complex (Fig. 1A), which ensures a tight fit between the binding pocket and the substrate (Fig. 1B). The fact that conditions causing denaturation of proteins (treatment by acid, alkali or urea) lead to the loss of enzyme activity lent strong support to the protein structure-function paradigm. Consequently, the notion that many proteins or regions of proteins could be not ordered, but intrinsically disordered was unacceptable for a long time. Indeed, computational studies carried out as early as in 2000 predict that ordered proteins and domains cover only about half of the sequence space in various proteomes,⁶ which has led to “breaking the protein rules”.⁷

The re-adaptation of the structure-function paradigm started as early as 1998 when Romero et al. predicted that more than 15,000 proteins in the Swiss Protein (SwissProt) database contain intrinsically disordered regions (IDRs) of at least 40 consecutive amino acids.⁸ Soon after, Wright and Dyson unveiled in 1999 that a large proportion of gene sequences are likely to be unfolded in solution, under native, functional conditions.⁹

Through the last years, the terminology seem to have settled on “intrinsically disordered”,¹⁰ for proteins/regions that exist as dynamic ensembles, within which atom positions and backbone Ramachandran angles exhibit extreme temporal fluctuations without specific equilibrium values.^{8,11-16} Their detailed characterization suggests that our traditional views of protein structure and function need to be re-assessed.

3. Computational means for IDPs analysis

3.1. Sequence peculiarities of IDPs

The frequency of amino acids in disordered proteins significantly differs from that of ordered proteins.^{11,12,16} Amino acids frequencies plotted as a function of the flexibility index of residues show a distinctive pattern¹⁷ (Fig. 2): IDPs are often enriched in charged and structure-breaking residues (Pro and Gly) and in Ala. This collection of amino acids has been called “disorder promoting amino acids”. On the other hand, these proteins are depleted in hydrophobic and aromatic residues and also possess fewer Cys and Asn residues, collectively termed “order promoting amino acids”.^{18,19} It is generally thought that the absence of structure is encoded by the amino acid sequence. The ensuing

low mean hydrophobicity and a high net charge represents an important prerequisite for the absence of a compact structure in proteins under physiological conditions,¹⁵ because it provides less driving force for protein compaction while also contributing to charge-charge repulsion. This simple principle can be applied to predict IDPs by plotting the absolute net charge as a function of the mean normalized hydrophobicity, a plot denoted as charge-hydropathy (CH) plot or Uversky plot (Fig. 3).¹⁵ In this plot, IDPs cluster in the high net charge - low net hydrophobicity half of the plane (Fig. 3), and the distance from the separating line (the formula of the linear function is: $\langle R \rangle = 2.743 \langle H \rangle - 1.109$) may carry information on the extent and type of disorder for the whole chain (Fig. 3).^{20,21}

Protein disorder is also related to low sequence complexity (i.e. repetitiveness and limited diversity in amino acid composition).^{22,23} Accordingly, protein disorder abounds in certain types of charged and Pro-rich repetitive regions.²⁴ 25% of all amino acids in the SwissProt database are found in such regions, and 34% of all proteins have at least one such segment.¹¹ It is to be noted though that some special cases of low-complexity sequences are found in proteins with a certain amino acid periodicity, yet ordered such as coiled-coils (collagen, for example). This biased composition together with these peculiar sequence features enabled the development of reliable computational tools for the analysis of structural disorder.

3.2. Computational tools for disorder prediction

The recognition of disordered regions from sequence facilitates the functional annotation of proteins and is instrumental in delineating protein domains amenable to crystallization.^{25,26} To this end, many disorder predictors have been developed.²⁵⁻²⁷ (cf. also the table in the Wikipedia article found at http://en.wikipedia.org/wiki/Intrinsically_unstructured_proteins). The reliability of disorder prediction benefits from the use of several methods based on different concepts, different physicochemical parameters, or different implementations.²⁵⁻²⁹ In principle, this problem can be approached from three distinct directions:

1) Simple amino acid propensities reflecting some basic physical or sequence features. Such propensity-based predictors rely on simple statistics of amino acid propensity, the physical/chemical features of amino acids or on a preliminary concept on the physical background of disorder. The CH-plot¹⁵ with its derivative FoldIndex,³⁰ PreLink,³¹ and GlobPlot,³² to name few, fit into this category.

2) Machine-learning algorithms (MLAs) are trained on datasets of disordered regions. Presently three databases of experimentally characterized disordered proteins are available: the Database of Protein Disorder, DisProt (<http://www.disprot.org/>),³³ IDEAL (<http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/>),³⁴ and MobiDB (<http://mobidb.bio.unipd.it/>),³⁵ with the former being the largest publicly available database, with about 1500 different regions. PONDR,²³ DisEMBL,³⁶ DISOPRED2,³⁷ and RONN,³⁸ to name few, fall in this category of predictors since they have been trained on disorder data coming from either short regions of missing density from the PDB or from nuclear magnetic resonance (NMR) or from the DisProt database.³³ Regions of missing electron density in the PDB are generally short, as long regions prevent crystallization.²⁵ As such, short disorder is over-represented in the database of disordered regions, and hence these predictors tend to perform better in predicting short disorder than long disorder.

3) The tendency of amino acids to make or avoid contacts with each other. The underlying idea is that IDPs are unfolded because they cannot make sufficient inter-residue contacts to overcome the large decrease in configurational entropy during folding. A simple statistical analysis of residue contact numbers is performed by FoldUnfold³⁹ that calculates the expected average number of contacts per residue from the amino acid sequence alone. IUPred^{40,41} evaluates the energy resulting from inter-residues interactions, based on their inter-residues contacts in globular proteins.^{25,42} Since

these predictors are not trained on data sets of disordered regions they avoid the shortcomings and biases associated with these disordered datasets. Therefore, they are expected to perform better than the former methods on disordered proteins presently under-represented in training datasets (i.e., fully or mostly disordered proteins).

Another non-automated method that is very useful for unveiling unstructured regions is hydrophobic clusters analysis (HCA).⁴³ HCA provides a two-dimensional helical representation of protein sequences in which hydrophobic clusters are plotted along the sequence.⁴³ As such, HCA is not *stricto sensu* a predictor, nevertheless, disordered regions are recognizable as they are depleted (or even devoid) in hydrophobic clusters (an HCA plot is presented in Fig. 4). The power of HCA analysis relies also on its ability to highlight coiled-coils, regions with a biased composition and regions with potential for induced folding (i.e. a disorder-to-order transition upon binding to a partner or ligand) (see also section 6.3).

Predictors can also be classified based on the binary nature of the prediction. Examples of binary predictors are the CH-plot and the cumulative distribution function (CDF) analysis.^{20,44} The CDF analysis summarizes the per-residue predictions by plotting predicted disorder scores against their cumulative frequency, which allows ordered and disordered proteins to be distinguished based on the distribution of prediction scores.^{20,44}

The CH-CDF plot (see Fig. 5) is an analytical tool combining the outputs of two binary predictors, the CH-plot and the CDF plot.^{45,46} The vertical distance on the CH-plot from the location of the protein to the boundary line is a scale of disorder (or structure) tendency of the protein (Fig. 5A). This distance is referred to as the CH-distance. In CDF-plots, curves of ordered proteins tend to stay on the upper left half, whereas curves of disordered proteins tend to locate at the lower right half of the plot. An approximately diagonal boundary line separating the two groups can be identified and the average distance of the CDF curves from this boundary (CDF-distance) is a measure of the disorder (order) status of a given protein (Fig. 5B). The CH-CDF plot is based on both the CH-distance and the CDF-distance (Fig. 5C).^{45,46} It provides very useful information on the general disorder status of a given protein. After setting up boundaries at CH=0 and CDF=0, the entire CH-CDF plot can be split into four quadrants. Starting from the upper right quadrant, by taking the clockwise sequence, the four quadrants are named Q1 (upper right), Q2 (lower right), Q3 (lower left), and Q4 (upper left). Proteins in Q1 are structured by CDF, but disordered by CH; proteins in Q2 are predicted to be structured by both CDF and CH; proteins in Q3 are disordered by CDF but structured by CH; and proteins in Q4 are predicted to be disordered by both methods. The location of a given protein in this CH-CDF plot gives information about its overall physical and structural characteristics. Fig. 5D shows the application of this tool for the analysis of ribosomal proteins.

In general, low sequence complexity, low predicted secondary structure content and high sequence variability constitute an indicator of flexibility of proteins or protein domains.²⁶ It should be stressed that it is difficult and maybe impractical to establish the “best” predictor at the moment. Some predictors perform better on short disordered regions (i.e. DISOPRED2 and PreLink) while other predictors (IUPred for instance) perform well in predicting long disordered segments and finally some predictors, such as PONDR, GlobPlot and FoldIndex, have been trained on both short and long disorder and provide a balanced performance.²⁵ Therefore, to avoid pitfalls, different predictors should be combined, as performed by metapredictors that seek a consensus of the scores of different predictors relying on different principles (PONDR-FIT for instance).⁴⁷ Alternatively, one could simply use publicly available servers allowing multiple predictions, such as “MeDor” (MEtaser of DisORder), a server that does not provide a consensus prediction but is intended to speed up the analysis of protein disorder thanks to the simultaneous submission and retrieval of various disorder predictions, the prediction of secondary structures and a HCA plot (see Fig. 4).⁴⁸ (For

a general procedure to predict structural disorder, see ref.²⁵; for a detailed description of metapredictors, see ref.²⁷)

An extreme extension of the combined use of different predictors is the combination of *in silico* and experimental approaches with the ultimate goal of inferring as many structural information as possible while limiting the experimental characterization to relatively low-demanding experiments. An illustration of such an approach can be found in ref.⁴⁹ where a spectroscopic and computational analysis were combined. In that study, the authors plotted the ratio between the Θ_{222} and Θ_{200} ($\Theta_{222}/\Theta_{200}$) of a set of IDPs under study, along with the $\Theta_{222}/\Theta_{200}$ ratio of a set of well-characterized random coil-like and premolten globule-like proteins.¹³ The authors then set an arbitrary threshold of the $\Theta_{222}/\Theta_{200}$ ratio that allows discrimination between random coil-like IDPs and IDPs adopting a premolten-like conformation. Then, they generated a plot in which the distance of each IDP under study from this threshold was plotted as a function of its CH-distance in the CH plot. This analysis was intended to combine, and hence extend, two previous methods^{13,15} so as to allow random coil-like forms to be readily and easily distinguished from premolten globule-like forms among proteins predicted to be intrinsically disordered by the hydrophathy/charge method. In the resulting plot, increasingly negative CH distances designate proteins with increasing disorder, while increasingly positive $\Theta_{222}/\Theta_{200}$ distances designate IDPs becoming progressively more collapsed, as a consequence of an increased content in regular secondary structure. Thus, the left bottom quadrant is expected to correspond to IDPs adopting a random coil-like conformation, while the right bottom quadrant is supposed to designate IDPs adopting a premolten globule-like conformation.

Although the number of known disordered sequences based on the analysis of amino acid sequence parts missing in the crystal structure of proteins and analysis of properties of already characterized IDPs and their sequences is not too large, it is not too small either. In fact, the available data are sufficient to create reliable computational tools (i.e., the existing data are enough to generate both training and testing sets needed for the development of predictors). These tools are able to rather accurately find sequence features similar to those used in training. The progressively increasing number of disorder predictors is an obvious reflection of these facts. The situation is further complicated by the fact that the structural spectrum of intrinsic disorder is very broad (in fact, it is much broader than that of ordered proteins). This means that existing computational tools and their various combinations are doomed to be sensitive to some specific (and rather limited) features and would not be able to cover the whole spectrum of disorder. We consider the realm of computational tools for disorder prediction to be similar to the multiparametric experimental approach for structural characterization of proteins. In that approach, although each technique is able to see only a part of picture, the whole picture still can be restored by looking to the structure at different angles (i.e., using different techniques sensitive to the different levels of protein structure). Obviously, it would be more appropriate to say that the generalized criteria for disorder are still evolving and that only the availability of a reasonably large number of experimentally characterized IDPs will provide a solid basis for justified protein disorder prediction. On the other hand, this does not mean that these predictors should not be developed before such large number of experimentally characterized IDPs will be collected. On the contrary, the existing experimentally validated criteria of intrinsic disorder can be and have to be used for the creation of corresponding computational tools. These algorithms represent a very important addition to the researcher's arsenal and open possibility for large scale analysis. One should remember that the interest to protein intrinsic disorder and appreciation of its importance for protein function was generated via the recognition that IDPs are not rare exceptions, but a new rule. Furthermore, this recognition was based on the bioinformatics and computational analyses, where various predictors of protein intrinsic disorder were used.

4. Multiparametric approaches for assessing intrinsic disorder: Experimental view on conformational ensembles and induced folding transitions

Structural disorder can be detected and analyzed by many (bio)physical techniques, some indirect, others more quantitative in providing structural data. Among them, one of the most powerful in providing quantitative information is NMR.

4.1. NMR

NMR is widely recognized as the predominant, quantitative technique in studies of IDPs. This section will provide a description of the most broadly used NMR experiments and of the most recent NMR advancements that have enabled a transition from a qualitative to a quantitative description of IDPs.⁵⁰

A global characterization of IDPs can be inferred by either simply recording a one-dimensional spectrum (1-D) and deriving the information from the low spread of the proton resonance frequencies due to the presence of all the residues in the same environment, or by performing a wide-line NMR relaxation experiment that provides information about the presence of a large hydrate layer in agreement with a disordered, open and extended state. Additionally, using pulsed-field gradient NMR, one can also measure the diffusion coefficient of proteins from which hydrodynamic parameters can be determined.⁴² Although these latter are less direct indicators of disorder as compared to chemical shifts, they provide useful hints because IDPs possess peculiar hydrodynamic parameters as compared to structure proteins. Heteronuclear single quantum coherence (HSQC) is one of the most frequent experiments in the IDP literature. It represents the starting point of resonance assignment (i.e. the process of identifying which resonance belongs to which residue of the protein), which is essential for a meaningful interpretation of more advanced NMR experiments. Indeed, high-field NMR spectrometers together with uniformly and specifically labeled proteins have overcome the problem of proton signal overlap of IDPs. Resonance assignments can be made using the dispersion of ¹⁵N and ¹³C nuclei, which are more sensitive to local amino acid sequence.⁵¹ Recently, methods exclusively based on ¹³C detection have been reported, which afford an independent strategy to simplify crowded spectra as well as to perform sequence-specific assignment.⁵² HSQC experiments are also useful for detecting interactions with unlabeled binding-proteins, because a change in the chemical environment and in relaxation due to the interaction make NMR parameters of residues directly involved shift or even disappear from the spectrum.⁵³ Once a resonance assignment has been achieved, a variety of NMR parameters can be determined to characterize the structural and dynamic behavior at the residue level, thus obtaining sequence-specific structural information (Fig. 6).

In order to determine local secondary structural preferences, many methods have been described in the literature such as chemical shifts (which is the most powerful tool for achieving this goal), coupling constants (which are directly related to bond torsion angles) and short range nuclear Overhauser effect (NOEs) (which can be used to corroborate secondary structure propensities detected through chemical shifts).⁵⁴

The primary observable in NMR studies of unfolded and partly folded proteins is the chemical shifts that are sensitive to polypeptide backbone torsion angles. Deviation of chemical shifts from random coil values (termed secondary chemical shift, SCS, or chemical shift index, CSI) for ¹H α , ¹³C α , ¹³C β , and ¹³CO are sensitive to local secondary structure (i.e. population and location) and hence provide important insights into the structures populated in the conformational ensemble in IDPs.^{51,55,56} One potential problem associated with this kind of approach concerns incorrect frequency referencing,

which can result in systematic errors in the secondary shifts. In order to address this problem, the $C\alpha$ and $C\beta$ chemical shifts (that shift in opposing directions for α -helical segments) can be used simultaneously to estimate the level of secondary structure in disordered proteins.⁵⁷ In Fig. 6B, an example taken from⁵⁸ is provided showing the $^{13}C\alpha$ chemical shifts of the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (N_{TAIL}) both in its free and bound (i.e. in complex with the C-terminal domain of the phosphoprotein, XD) states. This enabled to reveal that the protein has a propensity to form an α -helix even in the free state.

More detailed information about local conformational sampling can be derived from inter proton NOEs. However, quantitative interpretation of NOEs is complicated by the strong sensitivity of the interaction on the range of dynamic timescales commonly encountered in unfolded proteins.⁵⁹ In addition, long range NOEs characteristic of tertiary interactions are usually not observed in IDPs due to fluctuations and heterogeneity of the structure.⁴² Another interesting parameter that one could infer from HSQC-based experiments is relaxation rates (such as ^{15}N relaxation experiments) that reflect backbone and side-chains dynamics and could provide information about local disorder. Analysis of the relaxation data is particularly informative on local structural preferences, such as hydrophobic clusters, secondary structure elements and transient long-range contacts.⁵⁴ In fact, variations in backbone mobility can be correlated with intrinsic properties of the amino acids in the sequence. Clusters of small amino acids such as glycine and alanine show increased backbone mobility, compared to the average. In contrast, local hydrophobic interactions that persist in urea cause some restrictions of backbone motions, hence supporting the fact that a RC never exists, not even under the harshest denaturing conditions (i.e., in solutions containing 8M urea or 6 M guanidinium chloride).⁶⁰

The above-listed traditional NMR parameters only provide information about the local dynamics and structural state. This situation has changed by the advent of methods based on residual dipolar couplings (RDCs) and paramagnetic relaxation enhancements (PREs), which yield a high number of well-defined quantitative parameters reporting on long-range interactions (PRE) and global structure (PRE, RDC).^{54,59,61,62}

PRE has been most successful in detecting long-range contacts in disordered protein ensembles. The method relies on the introduction of a paramagnetic nitroxide spin label at a specific position within the amino acid sequence.⁵⁴ HSQC spectra are then recorded with the spin-label in the paramagnetic (oxidized) and diamagnetic (reduced) states, and differences in one of several parameters, such as line-width, relaxation rate, or intensity are determined. Limitations of this approach include the need of a number of cysteine-carrying mutants of the protein, and possible interference of non-native side-chain and/or the radical with native long-range contacts. Nevertheless, these measurements are extremely powerful, because they provide unambiguous evidence of the presence of fluctuating tertiary structure that can be very difficult to identify by other techniques.⁵⁹ An example taken from ref.⁵⁴ illustrating how a long-range distance could be inferred from PRE measurements is provided in Fig. 7. Moreover, many examples in the literature have used PRE distance information to constrain MD simulations and finally arrive at a reasonably confined distribution of structural states in the ensemble of IDPs.^{61,63}

As compared to chemical shifts and NOEs, RDCs present a powerful tool that provides a far more extensive and quantitative description of local, and possibly, global order in the unfolded state.⁵⁵ Indeed, RDCs measured between pairs of nuclei in partially aligned proteins, are very sensitive probes of time and ensemble-averaged conformational equilibria exchanging on timescales up to the millisecond and can therefore be used to characterize both the structure and the dynamics of unfolded proteins.⁵⁹ An example taken from ref.⁶² showing how RDCs have been used to probe the level and nature of residual structure in the C-terminal region of the MeV nucleoprotein (N_{TAIL}) is provided in Fig. 8A.

The unique power of NMR in characterizing IDPs is that the combination of different approaches enables to describe the real structure of IDPs; i.e., their structural ensemble, either alone or in the bound form.⁶⁴ Indeed, recently the disordered C-terminal domain of the measles virus nucleoprotein has been studied by a combination of NMR spectroscopy, SANS (small angle neutron scattering) and electron microscopy (EM), providing an *in situ* ensemble model describing the conformational behavior of the disordered region at the surface of the nucleocapsid (see Fig. 8)⁶² (For further description, see section 5.3)

In all, NMR provides quantitative residue-level information on structure and dynamics of IDPs, which enables to develop structural ensembles, the current best representation of the structure of IDPs. It should not be forgotten, however, that NMR has some inherent limitations, due to which it is best complemented by other quantitative biophysical techniques. Examples of such limitations, are averaging of fast exchange events, low signal/noise ratio, difficulty in defining ensemble shape and poor ability to provide information on global crossing between meta-stable sub-states, not to mention the fact that NMR is highly demanding in terms of protein amounts, that isotopic labeling is rather expensive and that some proteins are very poorly expressed in minimal media.

4.2. Further quantitative techniques for characterizing structural disorder

The limitations of NMR call upon complementing and/or combining its results with some other biophysical techniques which provide additional insights into the structural details of IDPs, and also enable approximations not yet amenable for NMR. Among these, we can cite for instance site-directed spin-labeling EPR spectroscopy,⁶⁵⁻⁶⁷ an approach powerful in documenting and mapping structural transitions in IDPs, which is currently being expanded thanks to the introduction of a new generation of spin labels targeting residues other than cysteine.^{68,69} Concomitantly to experimental techniques, *in silico* approaches have also been developed that enable physical quantification of folding and binding energy landscapes of IDPs at atomistic level. Among them, we cite those established by the group of Jin Wang who developed an atomic hybrid model for IDPs by integrating local physics-based interactions and non-local structure-based interactions.⁷⁰⁻⁷² In their recent approach, they have quantitatively uncovered the intrinsic energy landscape (not the free energy landscape often discussed in the field) of folding and binding. To this end, they have explored the underlying density of states, which is a statistical distribution in microcanonical ensemble extracted from the conventional canonical ensemble binding–folding dynamics using a structure-based model both with and without considering energetic roughness.⁷¹ To visualize the landscape, they chose the energy and the structural similarity of generated conformers to native structure defined as the fraction of native contacts as reaction coordinates. The authors showed that the energy landscape of folding and binding is funneled towards the native state. The energy landscape topography representing the degree of the funnelness is quantified by the dimensionless ratio of the slope and the roughness modularized by the entropy (size) of the funnel. The authors showed that this landscape topography ratio determines the thermodynamics, kinetics, and the association mechanism of the binding–folding dynamics. Using this newly developed IDP model in combination with physics-based simulations (with solvent), Wang *et al.* recently described the unbound ensembles of the α -MoRE (residues 484-504) located within the intrinsically disordered C-terminal domain of the measles virus (MeV) nucleoprotein (N_{TAIL}) and deciphered the mechanism by which the α -MoRE recognizes the X domain (XD) of the MeV phosphoprotein.⁷³

Among experimental techniques providing quantitative information, SAXS and single-molecule approaches, which allow the observation of transient intermediates, have enabled a better understanding of the behavior of IDPs. In particular, SAXS studies, combined with ensemble

optimization methods (EOM), have led to low-resolution description of IDPs as realistic conformational ensembles (see ref.⁷⁴ and references therein cited).

In-bulk methodologies to studying IDPs, however, have two main intrinsic limitations. The first limitation is connected to the limited capability to single out the monomeric state of the investigated IDPs. In the experimental conditions, the monomeric state of the protein under study might easily co-exist along with soluble oligomers that can form quickly in solution at the concentrations required by the sample preparations. Single-molecule methods operating at low concentrations or with immobilized molecules allow the properties of the monomeric species to be monitored without any interference by aggregation already started. The second limitation of the traditional in-bulk methodologies is connected to their time and ensemble averaging. They can observe only average properties of 10^{14} – 10^{17} molecules at a time. Even in the case in which no oligomers are present in the sample, the intrinsic ensemble averaging of in-bulk methodologies might prevent to single out poorly populated states. Single-molecule (SM) approaches, such as high-speed atomic force microscopy (HS-AFM),^{75,76} AFM-based force spectroscopy (FS),⁷⁷ and fluorescence resonance energy transfer (FRET),⁷⁸⁻⁸⁰ therefore represent a powerful tool to investigate the structure and the dynamics of IDPs. While SM-FRET reports changes in the mean distance between two residues (steady-state FRET) and intramolecular distance distributions (time-resolved FRET) enabling to study conformational equilibria in timescales shorter than a few milliseconds, AFM-based SM-FS is particularly sensitive to the formation of secondary structures, and probes timescales from milliseconds to seconds, much longer than those of SM-FRET.

HS-AFM allows direct observation of structural dynamics and dynamic processes of biological molecules, with a sub-second to sub-100 ms temporal resolution.^{75,81} Various dynamic protein processes have been successfully visualized by this approach. HS-AFM is not only applicable to well-structured proteins but also to IDPs. The group of Toshio Ando was the first to use HS-AFM to directly visualize IDPs attached on a mica surface.⁷⁶

Altogether, conformational and spectroscopic studies carried out thus far have unveiled that IDPs do not represent a homogeneous structural class, and range from fully extended (random-coil like) to compact (molten-globule like) states. To accommodate all these states in a functional framework, Keith Dunker elaborated the Protein Trinity Hypothesis,⁸² which posits that a native protein can be in one of three states, the ordered state, the collapsed-disordered (molten globule, MG) state and the extended-disordered state (random coil, RC), and that function can arise from any of the three states or from transitions between them.⁸² This model was subsequently expanded to include the pre-molten globule (PMG) state, which corresponds to an intermediate state between the RC and the MG (Fig. 9)¹³ (see also section 5.1).

4.3. Complementary methods for studying protein disorder

Certain low-tech approaches have provided ample early insight into the phenomenon of structural disorder, and paved the way for the general acceptance of this concept. Even today they may provide important insight and/or fast information on the structural state of a protein or its region.

IDPs can be recognized by their peculiar biochemical behavior. Indeed, due to their amino acid sequence bias and the lack of a tightly packed hydrophobic core, the presence of disordered proteins is often apparent during the purification process. Owing to their unique amino acid composition, IDPs frequently bind less sodium dodecyl sulfate (SDS) molecules than globular proteins leading to an aberrant mobility in SDS polyacrylamide gel electrophoresis (SDS-PAGE). Their apparent molecular mass determined by this technique is usually 1.2-1.8-fold higher than expected from the amino acid sequence measured by mass spectrometry.¹¹ Although this abnormal behavior has been ascribed in

the past to enrichment in acidic residues, recent findings suggest that other additional parameters, such as extension in solution, can be responsible for the abnormal electrophoretic migration of IDPs in SDS-PAGE.⁴⁹

An abnormal mobility is also observed in gel filtration (GF) or size-exclusion chromatography (SEC) experiments, in which IDPs have unusually high apparent molecular mass.¹²

Furthermore, flexible regions are known to be highly susceptible to proteolysis. Thus, limited proteolysis can be used to distinguish ordered and disordered proteins, as an increased proteolytic degradation *in vitro* argues for increased flexibility. Early observations on the curiously protease sensitive behavior of proteins were made on bacteriophage λ N,⁸³ dsp16⁸⁴ and p21,^{85,86} as outlined in general in refs.^{11,16,87}

A further special behavior of IDPs is their insensitivity to high temperatures and their stability towards acidic treatment, as observed early for MAP2,⁸⁸ α -synuclein,⁸⁹ involucrin,⁹⁰ and caldesmon,⁹¹ for example. These conditions usually causing the precipitation and/or denaturation of globular proteins,⁵³ as often exploited in the purification of IDPs.⁹²⁻⁹⁴ These peculiarities of IDPs provide the first line of indirect evidence of their unusual structural state. Direct structural information, however, can be obtained by a wide variety of experimental approaches (reviewed in ref.⁹⁵).

4.4. Analyzing disorder in a living cell

4.4.1. In vivo existence of protein disorder

A major challenge in the field of IDPs is to assess to what extent *in vitro* observations on the structural state and function of IDPs can be extrapolated to the living cell. Indeed, concentrations of proteins and other macromolecules reach 300-400 mg/ml in the cell, giving rise to a crowding effect^{96,97} that could affect the structural state of IDPs. Crowding might force IDPs to assume compact or even folded states caused by excluded volume effects, making this issue critical with respect to their physiological structural state and function. The presence of their binding partners might suggest that IDPs are partner-bound (and folded) most of the time. There are two direct approaches to address these issues: 1) mimicking crowding conditions in the test tube and 2) characterizing IDPs directly in the living cell.

An ideal *in vitro* crowding agent combines three different effects: an excluded volume effect, an increase in viscosity and an influence on the solvation/hydration of proteins. To mimic these, accepted approaches are to apply high molecular-weight (Mw) polymers, such as Dextran and Ficoll 70, or small molecule osmolytes, such as sucrose, trifluoroethanol (TFE) or trimethylamine N-oxide (TMAO). The high-Mw polymers primarily exert excluded volume effect, whereas osmolytes primarily act upon solvation of the protein backbone.⁹⁸ Osmolytes tend to stabilize function-related secondary structure of IDPs, as usually demonstrated by CD spectroscopy.⁵³ Crowding by Dextran and Ficoll-70 does elicit some compaction but not folding of IDPs,^{42,99,100} as demonstrated on different IDPs under different conditions and using a variety of techniques.¹⁰¹ It appears that the formation of local secondary structural elements of IDPs is promoted by crowding and thus preformed structural elements (PSEs) may receive even more credit than suggested by *in vitro* studies.⁴²

To obtain direct information from the cell, two approaches could be used: assessing proteasomal degradation and in-cell NMR studies. In the first approach, the ability of the 20S proteasome to degrade unfolded proteins is exploited. This degradation “by default” (i.e. ubiquitin-independent

degradation) provides evidence on the physiological structural status of proteins. It was used to characterize *in vivo* several IDPs including p53,¹⁰² and p21^{Cip1},¹⁰³ for which the default degradation by the proteasome argues for their disordered state in the cell.

The application of in-cell NMR requires the deposition of labeled proteins within the cell either by over-expressing under labeling conditions, by covalently linking cell-penetrating peptides to the protein, or by microinjecting it into the cytoplasm of the cell.^{42,104} An example is the tau protein that has been studied by in-cell NMR in *Xenopus* oocytes: while its microtubule-binding region became ordered, its long projection domain remained largely disordered.¹⁰⁵ Very recently, the group of Michele Vendruscolo has reported a study where in-cell NMR spectroscopy was used to observe directly the structure and dynamics of α -synuclein within *E. coli* cells. In agreement with previous studies by the group of Philipp Selenko,¹⁰⁶ the results indicate that, at least within the bacterial cytosol, α -synuclein populates a highly dynamic state that, despite the highly crowded environment, and it has the same characteristics as the disordered monomeric form observed in aqueous solution.¹⁰⁷ α -synuclein was observed to be largely disordered in the periplasmic space of *E. coli*.¹⁰⁸

4.4.2. Survival of IDPs in the cell

These observations raise the issue of *in vivo* degradation of IDPs, given their extreme proteolytic sensitivity *in vitro*.¹¹ The work by Belle et al. in 2006, in which *in vitro* half-lives of 3,750 yeast proteins have been determined, made the analysis of *in vivo* protein degradation possible.¹⁰⁹ By observing a limited correlation between half-life and predicted disorder, it was established that intrinsic disorder serves as a weak signal for intracellular degradation and the in-cell degradation is a multi-factorial process to which several signals including physical features, and regulatory mechanisms contribute.¹¹⁰ Further, it was found that the half-lives of proteins depend primarily on the length of the polypeptide chain, and hence on the number of potential cleavage sites. The observation that ubiquitinated proteins require disordered initiation sites for degradation,¹¹¹ and that disordered proteins may be even directly degraded by the 20S proteasome without prior ubiquitination (i.e. degradation by default),¹¹² is in line with this notion. Interestingly however, PEST regions (i.e. sequences of 10-50 amino acid residues that are enriched in Pro, Glu, Ser, and, Thr and that tend to turn over rapidly), destruction boxes and KEN-boxes (i.e. signal for a more specific degradation mechanism), are only poorly correlated with short half-lives.

A corollary of this study is that IDPs might be well-protected *in vivo* by accessory proteins, such as chaperones. To address the role of chaperones in IDPs function *in vivo*, pairwise interaction data from high-throughput interaction studies were analyzed for the correlation between intrinsic disorder and chaperone binding.¹¹³ The finding that partners of chaperones tend to be ordered proteins suggested that IDPs need no assistance for folding and/or protection against degradation, but possibly only for assembly into complexes and against aggregation.¹¹³

A related study by M. Babu and colleagues on the yeast and human proteomes showed that rather than a single regulatory mechanism, multiple mechanisms during transcription and translation control the availability of IDPs, and that IDPs do undergo a faster turnover as compared to globular proteins. Furthermore, the residence time (i.e. abundance, rate of synthesis, and half-lives of mRNA and proteins) of IDPs, in different organisms including human cells, was found rather low due to increased transcript clearance and proteolytic degradation.¹¹⁴ The need of tight regulation of IDPs is probably explained by their noted "dosage sensitivity" which denotes genes that are harmful when overexpressed. Intrinsic protein disorder was found to be an important determinant for proteins encoded by these genes, and explained by their inclination to be involved in promiscuous molecular interactions.¹¹⁵ Regulation of dosage-sensitive genes occurs at the transcriptional, RNA, and protein levels.¹¹⁶

Recent studies by Suskiewicz and colleagues proposed several possible further mechanisms by which IDPs could be regulated *in vivo*,¹¹⁷ where IDPs would be protected from degradation by default (see Fig. 10) through interactions with partners. In fact, forming a functional complex *in vivo* provides a way to avoid degradation either by promoting order or by masking IDRs (Fig. 10A). Apparently, some binding partners interact with IDPs principally to prevent their proteolysis. Those partners could be proteasome gatekeepers (Fig. 10B) interacting with the 20S proteasome or nannies (Fig. 10C) that are usually more specific for particular IDPs. Moreover, IDPs can bind to DNA (Fig. 10D), (for instance IDPs functioning as transcription factors) or also to “decoy” binding sites (Fig. 10E) that can compete with functional regions for binding to transcription factors. On the other hand, IDPs make interactions with themselves also (Fig. 10F). In order to avoid proteolysis initiated by a disordered N-terminus, an IDP could be stabilized by either a folded N-terminal domain (Fig. 10G), or by the interaction with ribosomal proteins (Fig. 10H) until a longer sequence is synthesized that could be stabilized by either a suitably nanny (Fig. 10C) or by a self-organization into a loopy structure (Fig. 10F).^{117,118} These proposed mechanisms actually relieve chaperones from the duty of guarding IDPs.

5. Structural and conformational characteristics of IDPs

5.1. IDP structure from the polymeric physics viewpoint

By definition, IDPs are devoid of stable secondary and/or tertiary structures under physiological conditions (please note that although collapsed IDPs, which are classified as native MG (see below), could contain some stable secondary structure elements, they still will be devoid of tertiary structure). Indeed, they consist of dynamic structures that interconvert on a number of timescales. However, a majority of them has some residual transient secondary structures that are required for function. Hence, it should be borne in mind that most proteins are neither fully ordered nor fully disordered but contain ordered and disordered regions at different ratios (Fig. 11).^{42,119,120} Indeed, only about 32% of the crystal structures of the PDB are completely devoid of disorder^{121,122} and the degree or the depth of disorder (as it is manifested by the absence of the electron density) vary widely, suggesting that the fully ordered protein depicted in Fig. 11 is apparently the exception rather than the rule.¹¹⁹ In addition, numerous experimental examples in the literature argue for a continuum of disorder rather than a unique disordered state, as suggested in ref.¹²³

Depending on the content of residual structure and environment, IDPs may exist in at least three separate functional conformations: molten globule (MG), pre-molten globule (PMG) and random coil-like (RC-like).¹³ In fact, IDPs can attain extended conformations (RC and PMG) or remain globally collapsed (MG) (Fig. 12, see also section 4.3)

The collapsed disorder comprises the MG state, in which native secondary structures exist although the protein molecule lacks well-packed core. MGs are characterized not only by the well-developed secondary structure, but also by the presence of some topology, i.e. relatively fixed mutual positioning of the secondary structure elements, and a compact state close to that of ordered globular proteins.¹³ The considerable increase in the affinity for hydrophobic fluorescence probes (ANS, for instance), the accessibility to proteases and the increase in the hydrodynamic radius which corresponds to a volume increase of approximately 50% with respect to that of the native globular state, are diagnostic markers of the MG state.¹³ Recent analyses of IDPs by the group of Pappu indicated sequence polarity and distribution of charged residues as determinants of IDP compaction: indeed polar IDPs were found to favor collapsed ensembles in water despite the absence of hydrophobic groups.^{124,125}

Extended disordered proteins also should not always be considered random coils. For practical reasons, they can be divided into two separate sub-families,¹³ PMG and RC conformations. The PMG state can effectively interact with the hydrophobic fluorescent probe ANS, though more weakly than a MG, which suggests that some hydrophobic clusters are formed in this state. Therefore, the PMG probably represents a “squeezed” and partially ordered form of the coil with some residual secondary structure.¹²⁶ The RC, or rather, RC-like, state has hydrodynamic dimensions typical of considerably unfolded polypeptide chains that are devoid of any ordered secondary structure.¹³

In general, PMGs are somewhat enriched in hydrophobic residues and have regions corresponding to residual secondary structure, which can be pointed out by HCA analysis indicating regions with a potential for induced folding (see section 3.2). It should be noted however that the content in regular secondary structure is not a major determinant of protein compaction.^{49,127,128}

5.2. Evaluating residual structure in IDPs

The hydrodynamic volume of a single-domain globular protein in the MG, PMG and the unfolded state, typically increases 1.5, ~3 and ~12 times, respectively, in comparison with that of the native state.^{129,130} Hence, several techniques such as GF, small angle X-ray scattering (SAXS), analytical ultracentrifugation and dynamic light scattering (DLS), to name a few, may help in determining whether a protein is collapsed or extended due to the difference in its hydrodynamic properties (Fig. 12).¹³ An empirical formula connecting the Stokes radii (R_s) to the apparent molecular masses (appMM) of proteins in different conformational states deduced from GF analysis enable to distinguish between the various protein conformations (Fig. 12).^{12,13,95} Residual structure in IDPs can also be unveiled by measuring the increase in hydrodynamic properties using DLS or SEC upon heating or chemical denaturation.⁹⁵

The residual structure in IDPs can also be unveiled by spectroscopic techniques such as circular dichroism (CD).¹³¹ The unfolded polypeptide chain is characterized by an intensive minimum in the vicinity of 190 and 200nm and an ellipticity close to zero in the vicinity of 222 nm (Fig. 13A).⁵³ Consequently, the estimation of the secondary structure content allows MGs, with stable secondary structures, to be distinguished from RCs or PMGs (Fig. 13A). To discriminate between the extended states (i.e. PMG or RC) a “double wavelength” plot plotting the mean residue ellipticity at 222 nm as a function of the mean residue ellipticity at 200 nm, was developed by Uversky (Fig. 13B).^{13,95} Discrimination can also be improved by applying denaturants and/or heating the protein, and by the simultaneous use of CD and hydrodynamic techniques or by the combined use of near-UV CD.¹³

Interactions between multivalent IDPs may lead to the formation of higher-order structural organization, as demonstrated by sharp liquid-liquid-demixing phase separations, generating micrometre-sized liquid droplets in aqueous solution. This macroscopic transition corresponds to a molecular transition between small complexes and large, dynamic supramolecular polymers, with the valency of the interacting species being directly related to the concentrations needed for phase transition. Importantly, phase transitions can be related with changes in protein activity and can be governed by posttranslational modifications (PTMs), as shown in the case of the actin-regulatory neural Wiskott-Aldrich syndrome protein (N-WASP).¹³² The widespread occurrence of multivalent systems suggests that similar phase transitions may be used to spatially organize and biochemically regulate information in biological processes.¹³³

5.3. IDPs as conformational ensembles

The full structural description of IDPs is only feasible by ensemble approaches, primarily based on NMR observables, such as chemical shifts (CSs), Residual Dipolar Couplings (RDCs), ^{15}N R_2 relaxation rates, Paramagnetic Relaxation Enhancement (PRE) distance restraints, J-couplings, pulsed field gradient (PFG)-derived R_h values, and ^1H - ^{15}N heteronuclear Nuclear Overhauser Effects (hetNOEs), complemented by SAXS, which all together yield detailed information on short- and long-range contacts and backbone flexibility.¹³⁴⁻¹³⁶ By all this information combined, one can describe the structure of an IDP as an ensemble of conformations. To this end, two broad approaches have been developed: either through restrained molecular dynamics simulations, or the generation of a large number of conformations and selection of a subset that fits the data.⁶⁴

In molecular dynamics simulations, the conformational space accessible to a given protein is sampled by penalizing deviations between the experimental data and the corresponding values back-calculated from the structures sampled during the simulations.⁶³ It should be noted that there is no guarantee that the generated structural ensemble would be consistent with experimental data not used as restraints. This is a general consequence that we cannot measure sufficient number of restraints at the moment.^{64,137}

The second approach starts with generating a very large pool of conformations, which may be completely random or constrained by experimental or theoretical data (e.g. Ramachandran Ψ/Φ angles taken from coil regions of PDB structures). Programs such as Flexible-Meccano,¹³⁸ Ensemble Optimization Method,^{135,139} and TRaDES^{140,141} are currently used for this purpose. If not provided by the initial algorithm, side chains can be modeled in with an algorithm such as SCCOMP¹⁴² or SCRWL.¹⁴³ Expected experimental observables are then back-calculated for each conformation in the pool (with CRY SOL for SAXS,¹⁴⁴ or ShiftX or SPARTA for NMR¹⁴⁵).

From the pool, a subset of structures (on the order of 50-100) is selected through step-wise optimization until the best fit is achieved. The ensemble describes the conformational state of the IDP, to the limit of the amount and quality of experimental data. Obviously, one of the most important issues in IDP research is to define the number (or range) of conformers for such ensembles to be relevant. Although this can only be addressed by performing the corresponding computational experiments, we expect the size of the ensemble to be roughly proportional to the extent of disorder in a given protein. It is also possible that the size of the ensemble can depend on the particular type of IDP function. Recent analyses revealed that in most cases 50-100 structures in the ensemble suffice in describing experimental observations⁵⁹, and even in large (such as α -synuclein and tau) and highly extended (p53TAD) disordered proteins, their conformational ensemble can be reliably described by about 200-400 members.^{146,147}

Ensembles can be deposited in the newly developed protein ensemble database (pE-DB, <http://pedb.vib.be>), which hosts primary experimental data, algorithms used for calculation and ensembles for IDPs.¹⁴⁸ Because ensemble are not unequivocally defined by structural data available, one protein can have several ensembles, and it is the major aspiration of the IDP field to find out how relevant ensembles are, i.e. if the function of IDPs can be interpreted by the ensembles. If we achieve this goal, it will signal the dawn of a new era in unstructural biology.¹⁴⁹

5.4. Electrostatic and charge effects of compaction of the structural ensemble of IDPs

The structural ensembles of IDPs have many features, local and global, which are critical for the functioning of the protein.¹⁴⁹ Local features such as the presence of pre-structured motifs might be involved in mediating interactions with partner proteins, whereas global hydrodynamic features might be important in regulating accessibility/exposure of motifs and also long-range communication

between different parts of the protein. Recent advance in describing IDPs by polymer theories offers a much improved insight into the sequence determinants of their global structural organization.

These theoretical and experimental approaches are derived from recognizing that despite a paucity of hydrophobic residues, polar/charged residues in IDPs prefer their collapsed structures in aqueous milieu.¹⁵⁰ Many of these analyses suggest that water at ambient temperatures is a poor solvent for generic polypeptide backbones, which makes IDPs in general to compact, with electrostatic interactions having a basic influence on the final state.¹²⁴ A study of highly charged IDPs by molecular simulations and fluorescence approaches showed that the charge content and pattern modulate the intrinsic preference of polypeptide backbones for collapsed structures. It was found that net charge per residue serves as the major discriminating order parameter, also in line with previous suggestions that net charge is a primary determinant not only of protein disorder¹⁵ but of the extension of IDPs.¹⁵¹ In addition, although overall amino acid composition represents a good proxy for polymeric properties, fine details of conformational ensembles depend on specific sequence attributes.

This is particularly true for the net charge of the polypeptide chain. Most IDPs are polyampholytes, i.e. their sequences include both positively and negatively charged residues. The fraction of charged residues discriminates between weak and strong polyampholytes, and weak polyampholytes tend to collapse into compact states.¹²⁵ The conformational preferences of strong polyampholytes, on the other hand, are determined by both the fraction of charged residues and the sequence distribution of oppositely charged residues. In the case of well-mixed sequences, electrostatic repulsions and attractions within the chain are counterbalanced, leading to a preference to a self-avoiding random walk or generic Flory-type random coil conformational state. When oppositely charged residues are segregated within the sequence, hairpin-like conformations emerge because long-range electrostatic attractions are preferred. It was also observed that the naturally occurring polyampholytes are usually well mixed, with a preference for random-coil ensembles. These inferences have been further born out by FRET experiments, which show that IDPs exhibit expansion at low ionic strength that correlates with their net charge.¹⁵² Charge-balanced polypeptides exhibit an additional collapse at low ionic strength conditions, yet they show low internal friction, which may have a significant impact on reconfiguration time.¹⁵³ Reconfiguration time is particularly relevant for the kinetics of conformational reorganization, which is directly linked with the dynamics of the interactions of IDPs with cellular binding partners.

5.5. IDPs, their environment and functional misfolding

Noteworthy, many examples in the literature pointed out that the structure of IDPs is very sensitive to their environment, often displaying an inside-out behavior, i.e. a partial folding of IDPs with increasing temperature or increasing/decreasing pH.^{13,154-156} Indeed, this behavior makes perfect sense: increasing the temperature can increase the strength of the hydrophobic interactions leading to a stronger hydrophobic driving force for partial folding, whereas variations in pH minimize the large net charge present at neutral pH and consequently decrease the charge/charge intramolecular repulsion and thus shift the conformational ensemble towards more compact states.^{13,53}

Intriguingly, some regions cross the disorder/order boundary or assume different structures as a result of slight environmental changes. This “ambiguous” behavior has been addressed via two concepts, Chameleon Sequences¹⁵⁷ and Dual-Personality (DP) sequences.¹⁵⁸ While chameleon sequences fold in a template-dependent manner and could change conformation within the same protein as a result of point mutations, ligand binding or a change in pH, DP fragments are close to the order/disorder boundary. Owing to DP fragments, identical proteins solved in different conditions

and functional states can be captured in both conformational states: ordered and disordered. DP sequences possess an amino acid composition intermediary between ordered globular and disordered proteins. DP fragments are thus defined as a distinct structural category.⁴² When IDPs appear to have the capacity to adapt to the structure of distinct partners and to carry out different, occasionally opposing functions, the concept is denoted as “moonlighting”.¹⁵⁹

Finally, how does Nature prevent highly promiscuous IDPs from undergoing unwanted interactions with non-native partners? Interesting findings by Uversky¹⁶⁰ suggest that IDPs escape unwanted interactions via functional misfolding (i.e. via the formation of non-native intramolecular interactions). In fact, since binding of IDPs to partners is frequently driven by hydrophobic residues organized in patches, the occurrence of PSEs with a non-polar continuous surface will surely promote the mutual attraction of these regions within the IDP and the formation of a structure where these elements are oriented to each other according to their interacting surfaces. In other words, the formation of native-like functionally relevant secondary structure elements containing sites for specific binding of native partners has to be accompanied by partial compaction of a polypeptide chain due to non-native intramolecular interactions (i.e. misfolding), thus native binding sites will no longer be exposed to interact with non-native partners.

6. Intrinsic disorder and protein function

6.1. Functional diversity of IDPs through the prism of structural plasticity

The ultimate goal of the structural description of IDPs is to elucidate and rationalize the types and modes of functions they play. Functions of proteins can be classified according to the cellular processes they are involved in (Biological Process, BP) and their mode of action at the molecular level (Molecular Function, MF). Here we will outline the basic molecular modes of IDP functions, from a mechanistic perspective. Initially, IDPs were suggested to fall into 28 different mechanistic types.¹⁶¹ In this classification, biological processes and molecular functions were not distinguished, and it was later shown that the actual functional modes can be rationalized in terms of five¹¹ to six broad functional modes, as illustrated in Fig. 14.

Entropic chain constitutes the first class where function involves no coupled binding and folding, rather it directly depends on the flexibility and the plasticity of the backbone. Indeed, the function of linkers and spacers, entropic clocks, entropic springs and entropic bristles stems directly from the disordered state. The microtubule-associated protein2 (MAP2), which functions as a spacer in microtubule architecture¹⁶² and FG Nups, which function as size-dependent filtering devices of the nuclear pore¹⁶³ are appropriate examples for this functional mode.

In the other classes, IDPs function via molecular recognition and hence they bind to a target, which could be DNA, RNA, another protein or a small ligand. When binding of the IDP alters the action of the partner (as, for instance, inhibitors or activators), the IDP belongs to the second class, termed effectors. Well-characterized IDPs in this class are p21^{Cip1} and p27^{Kip1} that inhibit the cyclin-dependent-kinase 2 (Cdk2) by binding to the Cyclin A-Cdk2 complex^{164,165}, and FlgM, the inhibitor of sigma28 transcription factor in bacteria¹⁶⁶. p21^{Cip1} and FlgM were the first IDPs to be studied in detail, which has led to concepts such as folding induced upon binding, pre-existence of binding-competent secondary structure and structural adaptability. Similar observations have been made with inhibitor 2 (I2) binding to, and inhibiting protein phosphatase 1.^{167,168} Effector functions of IDPs are probably the most straightforward to interpret in terms of the classical structure-function paradigm because the resulting complexes are often found in the PDB. When the effector has both

activities, activator and inhibitor, sometimes with the same partner, the term “moonlighting”¹⁶⁹ is used (see also section 5.5).

The third class is that of scavengers, in which the IDP stores and/or neutralizes small ligands. In fact, the open and extended structure of IDPs is particularly suited to bind small ligands, such as ions and organic compounds, either for disposal or for sequestration and later release upon the need of the organism.⁴² For example, salivary proline-rich glycoproteins, which constitute about two-thirds of the total protein in saliva, are scavengers that neutralize polyphenolic plant compounds (i.e. tannins).¹⁷⁰ Biom mineralization proteins storing and delivering calcium/phosphate to the site of the buildup of the mineral phase in bone and teeth also exemplify this mode of action.¹⁷¹

The fourth class constitutes assemblers, which assemble, stabilize and regulate large multiprotein complexes such as ribosome, cytoskeleton and chromatin.¹¹ In agreement, both prediction and structural data show statistically significant correlations between disorder and either the number of protein subunits of complexes^{42,172} or established scaffold function.^{173,174} In addition, high disorder also appears in hub proteins.^{175,176} In interactomes, a few proteins possess a large number of connections (i.e. hubs), whereas most others have very few, often only one, connection. This arrangement suggests a functional specialization in which hubs are preferentially involved in organizing the network. Interestingly, a range of bioinformatics studies suggests that hub proteins have an elevated level of disorder (reviewed in ref.¹⁷⁷). For example, Elisar Barbar and colleagues¹⁷⁸ describe how dynein light chain (LC8) interacts with multiple partners, many of which are not connected with dynein or microtubule-based transport, the function that was first attributed to this protein.¹⁷⁸

A special case of molecular recognition is exhibited by the fifth class, display sites, which mediate regulatory PTMs, such as phosphorylation or limited proteolysis. PTM of proteins has three structural requirements: an appropriate local sequence, structural exposure, and flexibility of the site so that it can be productively accommodated by the active site of the modifying enzyme; these requirements are met by structural disorder. Indeed, computational and experimental evidence support the abundance of PTM such as phosphorylation¹⁷⁹ and ubiquitination¹⁸⁰ sites in disordered regions of proteins. In addition, a recent study suggests that IDPs in yeast are more likely substrates of a large number of kinases.¹¹⁴ The power of DisPhos (disorder-enhanced phosphorylation) in predicting phosphorylation sites¹⁷⁹ revealed that phosphorylation sites primarily occur in regulatory, cancer-associated and cytoskeletal proteins, as opposed to proteins involved in degradation, biosynthesis, and metabolism, a feature associated with protein disorder from early on.¹⁷⁹

A recent addition to the list of functional modes of IDPs is chaperone function. A bioinformatics analysis showed that a high proportion of residues in RNA (54.2%) and protein (36.7%) chaperones fall into locally disordered regions.¹⁸¹ In addition, certain chaperone proteins are disordered along their entire length, e.g. α -synuclein,¹⁸² β -casein,¹⁸³ or LEA proteins,^{184,185} or chaperone activity is localized to disordered segments, such as in the small heat-shock protein (Hsp) α -crystallin.^{186,187} Disordered regions in chaperones can bind different partners (i.e. promiscuity), enable an enhanced speed of interactions and uncouple specificity from binding strength, which makes sense as chaperones need to bind a wide range of unrelated misfolded substrates. Moreover, the highly hydrophilic character of IDRs affords a solubilizing effect, preventing their aggregation^{42,181} as demonstrated by fusing disordered segments to a range of aggregation-prone proteins in expression studies.¹⁸⁸ To accommodate all different aspects of the mechanism of action of disordered chaperones, an “entropy transfer” model has been suggested, in which transient binding, solubilization and entropic exclusion elements are blended with transient interaction with the misfolded region of the substrate (client).¹⁸¹

The role of disorder in chaperone action has been most convincingly demonstrated in the recent example of the redox-regulated chaperone Hsp33.¹⁸⁹ This chaperone, which protects organisms against oxidative stress, is inactive in its fully folded state, but it undergoes oxidative unfolding within its redox-sensor domain upon stress. The IDR thus generated is used to discriminate between unfolded and folded partners. Upon returning to normal conditions, Hsp33 returns to the folded state and transfers bound clients to ATP-dependent foldases, i.e. its activity depends on internal order-to-disorder transitions.

6.2. Short recognition motifs in the interactions of IDPs. Interaction networks, one-to-many and many-to-one binding

As also apparent from the foregoing classification, IDPs often function via molecular recognition, when they bind partner molecules in and induced folding process. Their mode of binding is thought to confer many advantages, and is often mediated by short recognition elements (motifs).^{190,191}

Different concepts regarding short recognition elements have been extensively discussed in the literature. Indeed, depending on whether the idea is approached from a structural point of view or defined at the sequence level, a short motif could be denoted as “Molecular Recognition Element” (MoRE) / “Molecular Recognition Feature” (MoRF)¹⁹² or “Linear Motif” (LM), respectively (LM is also denoted as “Eukaryotic Linear Motif” (ELM) or “Short Linear Motif” (SLiM)).¹⁹³

Moreover, Fuxreiter et al. in 2004 elaborated the concept of preformed structural elements (PSEs) correlating the probable structural preferences of IDPs in the unbound state.¹⁹⁴ The structure of the recognition element of an IDP in complex with its partner is often known, and comparison of structures adopted in the bound form and that observed (or predicted) in the free form suggested that IDPs have rather strong intrinsic preferences for the conformation they attain when bound to their partners, with these preferences being highest for α -helical conformations. This may be interpreted in terms of partial pre-formation of their recognition segments in the free state.¹⁹⁴ Indeed, often a similar structure in the unbound and bound states is observed when the IDP is characterized by NMR, which formed the basis of the related concept of Pre-structured Motifs (PreSMos).¹⁹⁵

In a similar vein, MoREs/MoRFs has been established as short regions of (predicted) disorder that became structured and thus deposited into PDB, in complex with their partner. MoREs/MoRFs can adopt α -helix, β -strand or irregular secondary structure conformations (leading to α -, β - and I-MoREs, respectively) or a combination of different secondary structural forms (complex-MoREs) (Fig. 15).¹⁹² Although an exact relationship has not yet been systematically tested,¹⁹⁶ it is appealing to assume that MoREs/MoRFs could correspond to PSEs. Repartition of secondary structures, which suggests significant differences between structures of IDPs in the bound state and those of globular proteins, supports this similarity. Indeed, secondary structure statistics on 26 IDRs in complex with their partners and on a reference dataset of globular proteins revealed that helices are almost equally populated in both datasets. On the other hand, β -structures in IDPs are about 50% less frequent than in globular proteins. The most significant difference is the increased level of coil conformation in bound IDPs. The analysis of the distribution of secondary structures in MoREs/MoRFs in the PDB is in complete agreement with these findings: such short binding elements have 27% of their residues in α -helical conformation, 12% in β -strands and approximately 48% in irregular conformations (13% missing from the atomic coordinates).^{20,42,197,198}

Molecular recognition by short recognition elements (motifs) can - and has been historically - approached from the completely different direction of short sequence patterns determining functional interactions and enzymatic modification, i.e. LMs, ELMs, and SLiMs.^{190,199,200} These motifs

are linear in the sense that 3-D organization is not required to bring distant segments of the molecule together to build up the recognizable unit.¹⁹³ Disorder prediction of LMs and their flanking regions for the experimentally characterized examples of the ELM database (a database gathering Eukaryotic linear motifs) (<http://elm.eu.org>)¹⁹⁹ suggests that LMs and their flanking regions are segments of intrinsic disorder within a more ordered environment.¹⁹¹ Their inherent flexibility probably allows them to adopt various conformations and to bind to multiple partners. LMs, however, are enriched in hydrophobic residues (Trp, Leu, Cys and Tyr), charged Arg and Asp, and Pro residues, and they are depleted in Gly and Ala. This could be explained by the fact that their restrictive sites (RS) directly involved in recognition are either hydrophobic and rigid, or charged and flexible, whereas at the non-restrictive sites (NRS), excessive flexibility, even exceeding average IDPs, can be observed.^{42,190,191} Overall, the unique amino acid composition suggests a mixed nature of LMs, with a few specificity-determining residues strongly favoring order grafted on a completely disordered carrier sequence flanking the region critical for interaction.⁴²

Although not fully substantiated,²⁰¹ the three concepts of short recognition motifs (PSEs, LMs, MoREs) can probably be considered as manifestations of the same underlying principle of binding of an ordered partner by a short segment within a disordered region, which undergoes a disorder-to-order transition or induced folding upon binding.²⁰² All the recognition motifs are short, on the order of 3 to 20 residues in the case of LMs,¹⁹¹ 4 to 15 for PSEs, and less than 30 residues in the case of MoREs.^{20,159} Moreover, they possess an amino acid composition that resembles that of IDPs with some notable deviations (i.e. specificity-determinant residues favoring order) and they can even in some cases be already preformed in the free state (PSEs, PreSMOs).^{191,193-195}

It is to be noted, however, that binding regions within IDPs might correspond more to a domain than a motif, with their length exceeding 20-30 residues.²⁰³ Indeed, these regions possess typical characteristics of domains: 1) they are structurally and functionally independent of the remainder of the protein molecule, 2) they can be recognized by homology due to evolutionary conservation of sequences,²⁰⁴⁻²⁰⁶ and 3) they possess at least one specific function. In fact, about 14% of Pfam domains (i.e. conserved evolutionary domains gathered in the Pfam database) selected from the SwissProt database possess more than 50% of predicted disorder and 4% are fully disordered.

Due to their functional importance, predicting recognition elements (motifs) from sequence is of great interest. Although the information content of short sites is very limited, due to their enrichment in hydrophobic residues, indirect techniques have some success in delineating them. As mentioned, for example, the HCA plot⁴³ (see section 3.2), can indeed unveil such binding sites.²⁵ The PONDR VL-XT,²³ ANCHOR (<http://anchor.enzim.hu>)²⁰⁷ and DynaMine²⁰⁸ disorder predictors are sensitive to local tendency of ordering, and are thus informative in highlighting potential induced folding regions. In the case of PONDR VL-XT, it was shown that there is a correlation between downward spikes in disorder scores and binding regions within IDRs.¹⁹² On the other hand, ANCHOR relies on the pairwise energy estimation approach (e.g. the approach of IUPred, see section 3.1) which identifies segments that are in disordered regions and cannot form enough favorable intra-chain interactions to fold on their own and are likely to gain stabilizing energy by interacting with a globular protein partner.²⁰⁹ Although motif discovery from sequence alone is fraught with extremely high false positive rates, inclusion of additional data, such as enrichment in functionally related proteins, can significantly increase its confidence, as demonstrated in the SLIMDisc approach (for an example see ref.²⁰⁰). Furthermore, an ELM server (<http://elm.eu.org/>) has been developed in view of investigating short functional sites in modular Eukaryotic proteins.¹⁹⁹ The recently developed dynamics predictor, DynaMine,²⁰⁸ shows that recognition motifs have characteristic patterns of local dynamics, which could be used for improving the accuracy of motif prediction.

6.3. Variability of binding modes attainable by disordered proteins: folding before or after binding

It is currently accepted that the crux of the binding of IDPs is binding-induced folding, which can occur between an IDP and a globular partner, but also between two IDPs, with the resulting complexes often being amenable to structural studies.²⁰² It is of significant interest whether folding occurs before, after or concomitantly to binding since experimental evidence seems to support all these varieties. NMR, molecular dynamics (MD) and site-directed mutagenesis are approaches to study the folding and binding events.⁴² Although conformational selection requires that the IDP populates at least partly the bound state in the free form, which therefore presages the final bound conformation,^{194,195} it is important to emphasize that the pre-existence of folded structures in IDPs does not necessarily imply a conformational selection mechanism.^{73,210,211} To further complicate the scenario, many IDPs show a high degree of adaptability with respect to the partner (Fig. 16), i.e. they can fold into different structures on binding to different partners. This plasticity raises questions about the role of PSEs in the recognition process, and suggests that the conformation sampled in solution will not necessarily be favored upon binding to a partner;^{212,213} i.e., there seems to be no general mechanism for the binding-induced folding of IDPs.⁴² Altogether, and in spite of recent breakthroughs, our understanding of the molecular features that encode the binding efficiency of IDPs is far from being elucidated and still matter of controversies (see ref.²¹⁴ for a review on this topic).

Generally, one should note that a major functional consequence of induced folding is a decoupling between specificity and affinity (i.e. binding strength) to enable weak and reversible interactions (Fig. 17).¹⁹² In contrast with interactions involving globular proteins, where specificity and affinity are often coupled,¹⁹² when binding involves disorder-to-order transitions, entropy becomes a main factor in the free energy of binding, and the large decrease in conformational entropy uncouples binding strength from specificity, which makes highly specific interactions reversible.^{159,192} To limit unfavorable effects on binding and ensure kinetically and energetically “facile” interactions between IDPs and their physiological partner(s),¹⁹⁴ the “conformational selection” model was proposed, in which pre-existent native-like conformations dominate the binding process.^{215,216} It is however to be expected that some degree of conformational adaptation will be required even when a preformed structural element in an IDP binds to its target.²¹² On the other hand, unlike “conformational selection”, much experimental evidence in the literature proposes that folding of IDPs can only take place upon binding and hence contacts (usually hydrophobic) with the binding-partner constitute the major driving force for the folding.^{202,212} In fact, recently, Phi-value analysis developed to study the mechanism of folding of globular proteins^{217,218} has been implemented in the IDP field. By this technique, residue-specific structural information is inferred by comparing the kinetics of the reaction (folding and/or binding) of the wild- type protein with a series of conservative single mutants, probing the presence of native-like structure in the transition state. Detailed kinetic and thermodynamic analysis showed that IDPs may bind both by conformational selection and induction of folding in the bound state.²¹⁹⁻²²²

Primary binding via an extended, disordered state, is the key element of the so-called “fly-casting” binding mechanism.^{223,224} It was assumed that IDPs can have an enhanced capture radius for a specific binding site, which confers an effective enhanced ability to search for a partner (also termed protein fishing).²²⁴ This mechanism is closely related to another concept of recognition by short-motifs, that of “Primary Contact Sites” (PCS), which derives from the observation that IDPs can often attain the bound state very rapidly, suggesting that certain regions within their fluctuating structural ensemble might be exposed for initiating a productive interaction with the partner.^{42,225}

It appears that the three mechanisms “conformational selection”, “induced folding” and “fly-casting” can be merged into a single “synergistic model”, in which all these elements synergistically participate. A brief description of such a consensus mechanism from ref.²²⁶ is provided (Fig. 18): the first step can be described as the non-specific reeling of the IDP by its target molecule via a “fly-casting” mechanism. Once the IDP is close enough to its target, specific encounter is facilitated by PSEs, and “conformational selection” comes into play. Full formation of the complex, largely dominated by the folding coupled to binding, constitutes the third and last step of the consensus mechanism. The driving force of binding is provided by both the favorable intermolecular interaction energy and restriction of conformational entropy by binding conformationally restricted PSEs.²²⁶ An excellent example studied by NMR and MD experiments is the pKID and CBP complex formation.^{227,228}

It is of note, however, that in some complexes the binding does not induce folding. In such “random complexes”, the IDP interacts with its partner via transient contacts, and it behaves similar to its free form. This mechanism seems to operate in the C-terminal domain of the full length *E. coli* SSB that is disordered even when bound to DNA,²²⁹ and in T-cell receptor cytoplasmic tails forming homooligomers without assuming a folded state.^{230,231} A similarly high flexibility in the bound form was also observed in the complex between the C-terminal disordered N_{TAIL} domain of the nucleoprotein and the X domain (XD) of the phosphoprotein from Hendra virus, where the resonance behavior suggests α -helical fraying of the MoRE of N_{TAIL} on the surface of XD.²³² This binding via a disordered state has been termed “fuzziness”^{159,233} (see section 6.4).

The chemical nature of the interface of IDP complexes also constitutes an area of interest. Indeed, it has been shown that the chemical and physical features of interfaces of IDPs are distinct from those of globular proteins in four major aspects.^{198,234}

1) Geometric particularities. IDPs use a large proportion of their surface for interaction with their partner (50% for IDPs whereas globular proteins use 5%-15% only).²³⁴ In terms of absolute values, however, the interfaces of IDPs are slightly smaller than those of ordered complexes ($1141 \pm 110 \text{ \AA}^2$ for IDPs²³⁴ versus $1600 \pm 400 \text{ \AA}^2$ for globular proteins,²³⁵) which stems from distinct interaction topologies, i.e. from the fact that in 70% of the cases the binding surface of the IDP represents a sequentially continuous segment, and it never contains more than three separate segments. Ordered proteins, on the other hand, hardly ever use a single segment for binding to their partner, and their segmentation number may occasionally even exceed 10. It has been inferred that these differences might also be relevant for the weaker interactions of IDPs that might also entail less specificity.^{236,237}

2) *Chemical nature of interfaces: buried versus exposed residues.* The ratio of buried-to-exposed area of IDPs is much smaller for both polar and hydrophobic residues, which suggests that IDPs keep even their very few hydrophobic residues exposed for contact with the partner. In addition, interfaces of complexes involving IDPs were found to be enriched in hydrophobic residues, with these latter contributing significantly to the stability of the interaction interface.²³⁴ In a recent study, although enrichment in hydrophobic residues was confirmed, polar interactions were found to play a larger role in these complexes than in structured protein complexes, with interfaces being more complementary with respect to electrostatics than interfaces of globular proteins.²³⁸

3) *Interaction energies at the interface.* Ordered proteins tend to establish more stabilizing interactions within their polypeptide chains, whereas IDPs derive more stabilization from the interaction with the partner.²³⁴ These features are also in line with the noted chemical nature of the interface, the higher ratio of hydrophobic amino acids and tighter fit, which compensate for the larger unfavorable decrease in entropy that occurs in folding induced upon binding.

4) *Conservation of interfaces*. The regions that become ordered upon complex formation in IDPs are the most conserved,²³⁴ in contrast with the general tendency of evolutionary variability of IDPs.²³⁹ In fact, interface positions show the most pronounced tendency to be conserved, in complete line with their functional importance, which helps their identification from sequence as linear motifs.²⁴⁰

All these findings are in line with the exposure of hydrophobic amino acids in ELMS¹⁹¹ and MoRFs¹⁹⁸ and their tight interactions due to a better adaptation to the structure of the partner enabled by their induced folding. Recently, it has been confirmed that aromatic-aromatic intermolecular interactions in complexes implying IDPs occur frequently at the interface of the complexes, which provide both specificity and stability to the complexes and are enthalpically favorable.²⁴¹

6.4. Dynamic or fuzzy complexes

In spite of the insight provided by the folding coupled to binding concept, ordering of IDPs is very often not complete even in the presence of the partner, i.e. the protein remains partially disordered in the bound state (Fig. 19). This led Tompa and Fuxreiter to coin the term “fuzziness”,^{159,233} which represents an extension of the concept of disorder in the bound state.^{159,242}

Apparently, this phenomenon presents a widespread novel paradigm of protein structure and function as supported by a dozen of IDP-complex examples gathered and described in detail,^{201,233} pointing to the abundance of fuzziness in IDP-complexes. Fuzziness may mean only static structural promiscuity, when the IDP has more than one stable bound states, or it may encompass dynamic/disordered parts of the bound IDP. A linker region connecting two binding regions, or a flanking region might remain disordered, sometimes even the entire IDP fails to acquire a stable structure, as observed in the case of the cytoplasmic domain of T-cell receptors.^{230,231}

Many functional advantages can result from “fuzziness”, including interactions with alternative partners and simultaneous interactions with different partners.⁴² Indeed, the residual plasticity often allows adaptation of the same motif to different partners, or a variable arrangement of the recognition motifs, which can mediate interactions with alternative partners (i.e. promiscuity). Moreover, the disordered tails in complexes serve for partner fishing with nonspecific, transient contacts, the first step in the above-proposed model for IDPs interactions.²²⁴ Fuzzy parts of the complexes can harbor regulatory PTM sites or they can mediate interactions with additional partners. They can even directly or indirectly interfere with (promote or inhibit) binding of the part that undergoes folding transition. An illustrative example is provided by a proline-rich segment of the myelin basic protein (MBP) that can bind to several SH3 domain-containing proteins.²⁴³ In addition, fuzziness provides a way to reduce entropic penalty thereby affording enhanced affinity.

6.5. IDPs and posttranslational modifications

PTMs of proteins are reversible or irreversible chemical changes of a polypeptide chain that occur after translation. PTMs range from the enzymatic cleavage of peptide bonds to the covalent additions of particular chemical groups, lipids, carbohydrates or even entire proteins to amino acid side chains. PTMs extend the range of amino acid structures and properties, and diversify structures and functions of proteins.²⁴⁴ As many as 300 post-translational modifications of proteins are known to occur physiologically,²⁴⁵ and because of the various PTMs, the actual number of chemically different amino acids in proteins is believed to increase from 20 to more than 140.²⁴⁴ In general, proteomes are significantly more complex than one can expect from the analysis of the encoding genomes mostly due to two mechanisms, PTMs and alternative splicing of their mRNAs (see below).²⁴⁴

Some PTMs (e.g., phosphorylation, acetylation or palmitoylation) are readily reversible by the action of specific deconjugating enzymes. The interplay between modifying and demodifying enzymes allows for rapid and economical control of protein function. A similar control by protein degradation and *de novo* synthesis would take much longer time and cost much more energy. PTM can occur at any stage of the protein's life. Some proteins are modified shortly after their translation is completed and prior to the final steps of their folding. These early PTMs might affect the protein folding efficiency, protein conformational stability and even determine the fate of the nascent protein via directing it to distinct cellular compartments. Other proteins are modified after their folding and localization are completed. Here, PTMs can activate or inactivate catalytic activity or otherwise influence the biological activity of a protein.²⁴⁴

Some proteins require multiple different posttranslational modifications for their function. For such multi-PTM proteins, modified sites in proteins can not only mediate individual functions, but can also function together to fine-tune molecular interactions and to modulate overall protein activity and stability.^{38,246} One illustrative example of such proteins is provided by a family of nuclear IDPs, histones, which require methylation, acetylation, phosphorylation, ubiquitylation, ADP-ribosylation, and SUMOylation at different stages of their action, with different modifications affecting histone–DNA interactions and also histone–histone interfaces, thus providing the capacity to disrupt intra-nucleosomal interactions and to alter nucleosome stability.²⁴⁷ Although the N-terminal domains of the core histones are known to contain an extraordinary number of sites that can be subjected to PTM, over 30 histone modifications have been recently identified in the core domains too.²⁴⁸

Based on the peculiarities of the conformational state of the site where the PTM would take place, PTMs can be grouped into two major classes, namely modifications mostly affecting ordered proteins and modifications occurring in disordered proteins or disordered regions.^{154,249} Among disorder-targeting PTMs are phosphorylation, acetylation, acylation, adenylation, ADP ribosylation, amidation, carboxylation, formylation, glycosylation, methylation, sulfation, prenylation, ubiquitination, and Ubl-conjugation (i.e., covalent attachment of ubiquitin-like proteins, including SUMO, ISG15, Nedd8, and Atg8).¹⁵⁴ Many sites of protein phosphorylation are located within intrinsically disordered regions,^{161,179,250} and several other types of PTMs, such as acetylation, protease digestion, ubiquitination, fatty acid acylation, and methylation, also abound in regions of intrinsic disorder.^{161,180,250,251} It has been suggested that sites of the enzyme-catalyzed PTMs in eukaryotic cells exhibit a preference for intrinsically disordered regions.^{129,252} In fact, the modifying enzyme binds and modifies similar sites in a wide variety of proteins, which is facilitated by disorder of flanking regions.^{129,252}

6.6. Alternative splicing in IDP function and dysfunction

Alternative splicing (AS) is a process by which two or more mature mRNAs are produced from a single precursor pre-mRNA by the inclusion and omission of different segments.^{253,254} AS is commonly observed mostly in multicellular eukaryotes.^{255,256} For humans and other mammals, multiple proteins are often produced from a single gene, and 40 – 60% the genes yield proteins *via* AS.²⁵⁷⁻²⁶⁰ It was hypothesized that AS very likely provides an important mechanism for enhancing protein diversity in multicellular eukaryotes,²⁶¹ playing a crucial role in regulation of protein-protein interactions, ligand binding, and enzymatic activity,²⁶²⁻²⁶⁵ and yielding protein isoforms with different, even opposing, functions.

The existence of a strong prevalence of intrinsic disorder in protein regions affected by alternative splicing of their corresponding pre-mRNAs has been noticed long ago.²⁶⁶ Such commonness of intrinsic disorder in protein regions affected by alternative splicing is used by nature to avoid

structural complications. In fact, removal of a piece of sequence from a structured protein can lead to protein missfolding, causing protein aggregation and loss of function. Such structural catastrophe can be avoided when AS maps to intrinsically disordered regions.²⁶⁶ On the other hand, since molecular recognition is often achieved via relatively short amino acid segments located within disordered regions, modification of such functions can be easily accomplished by AS affecting the corresponding disordered regions. In this way, diverse functional profiles can be generated for the transcribed alternatively spliced products of the same gene. Furthermore, since regulatory and signaling elements in disordered regions can be comprised of just a few more or less continuous amino acids, and since a high density of functionally important segments can be located within disordered regions, functionality of IDPs can be completely rewired via AS.^{266,267}

AS is known to occur in almost all human genes, and therefore alterations of this process are intimately connected to the pathogenesis of various human diseases, ranging from cancer,^{268,269} to neurodegenerative²⁷⁰⁻²⁷² and cardiovascular diseases.²⁷³ Many of the pathology-related proteins affected by pathology-related AS are intrinsically disordered. For example, aberrant splicing events are described for a number of familial cancer syndromes and affect the tumor suppressor genes such as *BRCA1*, *APC*, *p53*, *FHIT*, and *LKB1*,^{269,274-276} many of which encode intrinsically disordered proteins. One of the major players in Alzheimer's disease pathology, microtubule-associated protein tau, is a typical IDP,²⁷⁷ the function and dysfunction of which is controlled via multiple AS events affecting its gene *MAPT*.²⁷¹ Similarly, six genes, including *PARK2*, *SNCAIP*, *LRRK2*, *SNCA*, *SRRM2*, and *MAPT*, are involved in aberrant AS events in Parkinson's disease patients.²⁷⁰

6.7. Intrinsic disorder and allosteric coupling in proteins

The level of structural disorder is high in signaling proteins, which are most often subjected to regulation of their activity. The classic mechanism of signaling regulation is allostery, in which a signal at "another site" ("allos stereos" – another object in Greek) modifies the activity/function of the protein. The classical models of allostery (or cooperativity, as observed in the classical case of hemoglobin) rely on the traditional structure-function paradigm, assuming conformational changes elicited by ligand binding in oligomeric proteins.^{278,279}

With the advent of IDPs, it has been recognized that structural disorder - in particular folding transitions - might play a role in optimizing intramolecular site-to-site communication.^{280,281} The basic assumption in this novel ensemble allosteric model (EAM) is that binding of a regulatory ligand (effector) promotes induced folding, which couples binding of adjacent domain(s). This coupling promotes binding of the ligand at the other domain, resulting in positive, but also in negative, type of allostery. The model predicts that site-to-site coupling is maximized when intrinsic disorder is present, overcoming the level of coupling based on conformational changes in classical allostery. Thus, the key mechanistic element of EAM is not the pathway (network) of residues through which coupling occurs, rather the disorder-to-order transition that energetically optimizes coupling. This mechanism seems to operate in several well-studied cases, such as the toxin/antitoxin pair Doc/PhD,²⁸² glucocorticoid receptor,²⁸³ and the ternary complex between the adenovirus early region 1A (E1A) oncoprotein, the CREB binding protein and the retinoblastoma protein.⁷⁹ Notably, it may even rely on the opposing structural transition, i.e. induced unfolding, as observed in the case of PUMA BH3-only protein, which disrupts the interaction between cytosolic p53 and BCL-xL to enable the apoptotic function of p53.²⁸⁴

Finally, in a slightly different though related mechanism, the discovery of regulatory sites that can enhance or dampen the interaction with a partner while being located outside the primary binding site, provides an example illustrating the extreme allostery (i.e. the long-range nature of the effects of amino acid substitutions) that typifies IDPs.²⁸⁵

6.8. Multi-valent interactions and phase-separation of IDPs

IDPs are often repetitive in nature,²⁸⁶ and are thus capable of mutual multi-valent binding which results in their separation from solution state in a second-order phase transition generating a hydrogel-like state. This phase transition might represent a novel and special aspect of the functionality of IDPs.¹³³ The original observation was made by Li and co-workers,¹³² who studied the interactions of IDPs with tandem copies of SH3 domains and SH3 interaction motifs, and showed that the proteins engaged in multi-valent interactions producing sharp sol-gel phase transition and phase separation. Not only is phase separation a physical phenomenon, but it also causes an increase in actin-polymerization activity in the system made up of NCK–nephrin–N-WASP proteins. Further, the sol-gel transition is controlled by the level and stoichiometry of the proteins and post-translational modification, such as phosphorylation.

A similar physical phenomenon is induced by low-complexity regions in RNA-binding proteins, which can form cellular RNA granules.²⁸⁷ This mechanism may be important in the physical regulation of mRNA localization and translation, through reversible phase transitions with a highly dynamic hydrogel state. In addition, hydrogels formed by one disordered RNA-binding protein can bind IDRs of other members of the protein family, which, through the formation of heterogenous systems, might drive regulatory cross-talk between distinct sub-systems.

Yet another system is represented by cytoskeletal components neurofilaments (NFs, intermediate filaments in non-neuronal cells). NFs have three components of different molecular weight, all three having a rod region that forms the body of the cytoskeletal structure, and intrinsically disordered projection domains of variable length. These highly repetitive (low-complexity) disordered domains can attract each-other and form transient cross-bridges supporting a liquid crystal-like hydrogel structure.²⁸⁸ Depending on external pressure, the hydrogel has several stable structures, which may have different effect on the flexibility (physical stiffness) of the cytoskeleton.

These and other cases¹³³ suggest an interesting novel regulatory mechanism in the cell by multivalent IDPs. These proteins can generate an isolated physical phase of unique physical and biochemical properties, which might also bridge diverse length scales in the cell stemming from macromolecules to organelles.

7. Cellular biology of IDPs. An overview

7.1. Abundance of intrinsic disorder in various proteomes

Intrinsic disorder is abundant in nature, and many proteins are disordered along their entire length. Indeed, predictions on representative genomes from the three kingdoms of life (i.e. bacteria, archaea and eukaryotes) confirmed the prevalence of disorder although the disorder amount differs significantly between them.^{82,289} Apparently, there is an overall increase of disorder with the complexity of an organism.^{82,289,290} In agreement, earlier predictions revealed a high disorder percentage for eukaryotes (with 30% of IDRs longer than 40 consecutive amino acids) compared to bacteria (4%) and archaea (2%).²⁸⁹ This point is convincingly illustrated by differences between the genome and the proteome of two model organisms, *S. cerevisiae* and *E. coli*,²⁹¹ which underscore a “large jump in putatively disordered proteins in multi-celled, rather than single-celled, organisms”.⁸²

It is to be noted, though, that the level disorder in both prokaryotes²⁹² and eukaryotes²⁹³ varies widely, and comparing grand averages is of limited information only. In both kingdoms the level of disorder seems to represent a strong adaptive trait that shows correlation with the environment²⁹² or pathogenic lifestyle.²⁹³ In all, the signaling functions linked with the complexity of the organism

seem to represent a major factor determining the general prevalence of disorder in an organism, on which various adaptive changes reflecting the environment and lifestyle of the organism, operate.

It has been proposed that motifs embedded in disordered regions of viral proteins are employed by viruses to interact with host proteins, in order to hijack and manipulate these proteins for their purposes.²⁹⁴ In line with this suggestion, it was shown that viral proteins, and in particular RNA virus proteins, are enriched in disordered regions.²⁹⁵ In that study, the authors propose that beyond affording a broad partnership, the wide occurrence of disordered regions in viral proteins could also be related to the typical high mutation rates of RNA viruses, representing a strategy for buffering the deleterious effects of mutations.²⁹⁵ Taking into account these considerations, as well as the correlation between overlapping genes and disorder,²⁹⁶⁻²⁹⁸ it has been proposed²⁹⁹⁻³⁰² that the main advantage of the abundance of disorder within viruses would reside in pleiotropy and genetic compaction. Indeed, disorder provides a solution to reduce both genome size and molecular crowding, where a single gene would (i) encode a single (regulatory) protein product that can establish multiple interactions via its disordered regions and hence exert multiple concomitant biological effects, and/or (ii) would encode more than one product by means of overlapping reading frames. In fact, since disordered regions are less sensitive to structural constraints than ordered ones, the occurrence of disorder within one or both protein products encoded by an overlapping reading frame can represent a strategy to alleviate evolutionary constraints imposed by the overlap. As such, disorder would confer to viruses the ability to "handle" overlaps, thus further expanding the coding potential of viral genomes.

In agreement with these predictions, and following the first and seminal observations that disorder is abundant in proteins of the replicative complex of paramyxoviruses,³⁰³⁻³⁰⁵ an increasing number of experimental evidence has been gathered in the last decade pointing out the abundance of disorder in viral proteins (for reviews see refs.³⁰⁶⁻³⁰⁹). In fact, a recent bioinformatics analysis of the abundance of intrinsic disorder in various proteomes from viruses and the three domains of life revealed that viruses are characterized by the widest spread of proteome disorder content, with the percentage of disordered residues ranging from 7.3% in human coronavirus NL63 to 77.3% in Avian carcinoma virus.³⁰⁹

7.2. Involvement of structural disorder in biological processes

As suggested above, there are many functional categories that are positively or negatively correlated with structural disorder. In a comprehensive bioinformatics study carried out by Xie et al.^{249,251} a positive correlation between the functional annotation of the SwissProt database and predicted intrinsic disorder has been found (238 out of 710 SwissProt functional keywords), whereas a negative correlation was found in 310 out of 710 functional keywords (the remainder 162 functional keywords yielded ambiguity).

In some detail, the major functional categories which are significantly enriched in structural disorder are: differentiation, transcription, transcription regulation, spermatogenesis, DNA condensation, cell cycle, mRNA processing (splicing), mitosis, apoptosis, protein transport and meiosis, whereas major categories depleted in structural disorder are: GMP biosynthesis, amino acid biosynthesis, transport, electron transport, lipid biosynthesis, aromatic hydrocarbon catabolism, glycolysis, purine and pyrimidine biosynthesis, carbohydrate metabolism, and biosynthesis of branched-chain amino acids and lipopolysaccharide biosynthesis. These preferences are in line with a large-scale, unbiased analysis of the proteome in mouse embryo fibroblasts that were under- and over-represented in heat-stable proteins, followed by a rigorous analysis of GO terms (biological process, molecular function, and cellular localization).³¹⁰ In all, the structural preferences of these categories underscore

the old wisdom in the IDP field that structural disorder is enriched in proteins involved in signaling and regulatory functions and depleted in enzymes.³¹¹

8. Distinctive features of IDP evolution

Tracking the evolutionary history of a protein is very closely related to understanding the molecular basis of its function. As it was already mentioned (section 7.1), the level of disorder increases with the complexity of organisms,^{8,289,312} and it can undergo rapid changes in adaptation to external conditions.^{292,293} These rapid evolutionary changes raise the question as to how genes encoding for IDPs can arise. There are several possible mechanisms, such as *de novo* generation (i.e. from random pieces of genomic DNA), lateral gene transfer and horizontal gene transfer.⁴² A possible mechanism that has been confirmed through several examples in the literature is the generation of disordered domains by gene duplication and module exchange (i.e. attaching a disordered domain to an already existing protein). Indeed, many protein families exist that have common disordered domains, as for instance the kinase inhibitory domain (KID) of Cdk inhibitors p21^{Cip1}, p27^{Kip1}, and p57^{Kip2},³¹³ the catenin-binding domain (CBD) in T-cell factors LEF-1, Tcf3, and Tcf4,³¹⁴ and the WASP-homology domain 2 (WH2) in actin-binding proteins thymosin- β 4, ciboulot, and WASP³¹⁵, as outlined in⁴² The possible generality of disorder spreading by domain duplications and exchange between genes is also underscored by the observation that about 14% of all Pfam domains are predicted to be mostly disordered.²⁰³

Another possible mechanism is the expansion of internal repeat regions (i.e. regions coding micro and mini-satellites). Indeed, an analysis of 126 known IDPs showed that the percentage of proteins with tandemly repeated segments is much higher in IDPs (39%) than in SwissProt (14%), or human (28%) proteins²⁸⁶ and short sequence repeats often occur in IDPs.²⁴ The genetic instability of repetitive regions as witnessed by large interspecies variability and polymorphism, combined with their frequent involvement in the function of IDPs, confers to IDPs an exceptional functional variability and might represent a prevalent mechanism of the evolutionary spread of disorder^{239,289,316} and/or rapid adaptive genomic sweeps.^{292,293}

By aligning sequences in which a globular domain and a disordered region are present, one could compare the rates of their evolutionary change. Accordingly, Celeste Brown and colleagues,³¹⁷ performed a sequence alignment of homologs of 26 such families and calculated their pairwise genetic distances, confirming a high evolutionary rate of IDPs with respect to globular ones in 19 families (with similar rates in 5 families and slower rates in 2 families).³¹⁷ The likely explanation for the generally faster rates of evolution in IDPs is the lack of structural constraints (i.e. lack of a fixed 3-D structure) that enables many substitutions.³¹⁷

Despite the general tendency of IDPs to evolve rapidly, certain IDPs are rather resistant to evolutionary changes and hence they are conserved. Those regions were termed conserved disorder predictions (CDPs) and are usually short. CDPs can be found in all kingdoms of life, but long ones are almost 10 times more frequent in viruses and eukaryotes than in bacteria and archae.²⁰⁵ In a recent comparative analysis, disorder was found to segregate into three distinct behavioral categories, flexible disorder (conserved disorder with quickly evolving sequences), constrained disorder (conserved disorder with conserved amino acid sequence, such as ribosomal proteins) and non-conserved disorder.³¹⁸

These findings have to be elucidated in terms of the subtle balance between the disadvantage of the effect of a mutation on impairing activity/function, and the activity of change in generating evolution variability.^{319,320} In other words, evolutionary variability provides the raw material for selection

among functional variants, but too much variability works against retention of function. In general, in the case of functions that stem directly from the disordered state (i.e. entropic chains), the amino acid sequences are rather resistant to mutations as long as the mutations do not bring a major transition in the conformational ensemble. This was formally demonstrated in the case of replication protein A (RPA70), which contains two domains and a disordered linker. It was shown by NMR that the dynamics of the linker, despite great evolution variability, is preserved as a key functional feature.³²¹

In the case of molecular recognition functions, which require motifs of specificity-determinant residues interspersed in highly disordered regions (see section 6.2),¹⁹¹ conservation poses a serious challenge with respect to the generally high sequence variability observed in IDPs. Apparently, only the very little fraction of direct recognition residues is subjected to evolutionary constraints. The same applies to PTM sites, changes of which might have a deleterious effect on signaling and regulation events.²³⁹ In agreement, Chen et al.³²² showed that phosphorylated Ser and Thr residues tend to evolve more slowly than the non-phosphorylated ones.

9. Intrinsic disorder and human diseases

9.1. Correlation of disorder and diseases

A general correlation between intrinsic disorder and various diseases, such as cancer, diabetes, amyloidoses, neurodegenerative and cardiovascular diseases, is supported by several bioinformatics analyses (Fig. 20) and by detailed studies on individual proteins (see also section 6.6). These findings suggest that protein conformational diseases may result not only from protein misfolding, but also from misidentification, missignaling, and unnatural or non-native folding, which lead to the D^2 concept, i.e. disorder in disorders.³²³

In fact, many human diseases known as protein conformational or protein-misfolding diseases arise from the failure of a specific peptide or protein to adopt its final conformational state (i.e. protein misfolding), which leads to protein aggregation and/or fibril formation, loss of normal function, and gain of toxic function. An increase in the propensity to misfold can be the result of interactions with endogenous factors, such as chaperones, intracellular and extracellular matrixes, other proteins and small molecules, or point mutations, or also impaired PTMs.³²³ The conversion of specific proteins from their soluble functional states into stable, highly ordered, filamentous protein aggregates (i.e. amyloid fibrils) and the deposition of these aggregated materials into a variety of tissues and organs lead to a high number of misfolding diseases (as for instance, neurodegenerative diseases like Parkinson and Alzheimer diseases). The fact that IDPs are key players in the development of such pathological conditions is also in agreement with their involvement in key cellular functions (i.e. signaling and regulation).³²⁴ It should be noted, however that the capacity to form fibrils is not unique to IDPs: globular proteins can also form such aggregates (transthyretin for instance³²⁵) and it has been suggested that practically all proteins have the capacity to form fibrils.³²⁶

Based on the high abundance of IDPs in all kingdoms of life, their important functions and their ability to undergo a misfolding (i.e. amyloids), a novel view of the fate of a polypeptide chain inside the cell has been reported by Uversky (see Fig. 21).³²⁷ A polypeptide chain can either fold to gain a unique structure necessary for catalytic and transport activities, it can stay substantially non-folded, or can misfold and, under some circumstances, can form amyloid-like fibrils.

9.2. IDPs as drug targets

Protein-protein interactions are attractive targets for drug development. If the 3-D structure of the complex is available, small molecules can be designed that will bind to interfaces and interfere with protein-protein interactions. The low affinity of molecules to their targets and their low specificity towards specific interfaces (i.e. all active sites of proteases are similar), however, often limits the biological use of these molecules.¹²⁹

In this regard, protein-protein interactions involving IDPs are of great interest to develop new strategies for drug design. The “druggability” of IDPs is based on different observations: 1) the lack of flat surfaces in IDP-complexes which are typically found in interactions implying globular proteins 2) the weakness of interactions involving IDPs due to the use of energy in organizing the disordered partner thus facilitating tighter binding by a small drug, 3) the frequent involvement of hydrophobic clusters in folding coupled to binding and the involvement of only one (or few) contiguous segments making the interaction area small and compact, and 4) the ease of mimicking MoREs by appropriate small molecules.³²⁸

In IDP-partner interactions, drugs can be designed either to target the binding site of the partner through mimicking the disordered motif,³²⁸ or the IDP itself. As an example of the first approach, nutlins have been identified to bind Mdm2 and hence block the binding of a disordered region of p53.^{329,330} These compounds have the capacity to stabilize and activate the p53 pathway in cancer cells, leading to cell-cycle arrest, apoptosis, and growth inhibition of human tumor xenografts in nude mice.^{329,330}

In the inverse scenario, inhibitors bind to the disordered regions and interfere with their interactions. c-Myc-Max complex provides an interesting example. c-Myc is a transcription factor involved in many types of cancer. In order to bind to its targets, c-Myc must dimerize with Max by mutual folding when their zipper domains interact to form a helical coiled-coil. Several inhibitors were found through high throughput screening which are able to bind to a disordered region within the monomeric c-Myc. Inhibitor binding induces only local conformational changes, preserves the overall disorder of c-Myc, and inhibits dimerization with Max.^{129,331}

Abbreviations

1-D	one-dimensional
3-D	three-dimensional
AFM	atomic force microscopy
appMM	apparent molecular mass
AS	alternative splicing
CD	circular dichroism
CDF	cumulative distribution function
CH-plot	charge-hydrophathy plot
CDP	conserved disorder prediction
CSI	chemical shift index
DLS	dynamic light scattering
EAM	ensemble allosteric model
ELM	eukaryotic linear motif
EM	electron microscopy
ES	enzyme-substrate
FRET	fluorescence resonance energy transfer

FS	force spectroscopy
GF	gel filtration
HCA	hydrophobic clusters analysis
HS	high speed
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
KID	kinase inhibitory domain
LC8	dynein light chain
LM	linear motif
MAP2	microtubule-associated protein 2
MD	molecular dynamics
MeV	measles virus
MG	molten globule
MLA	machine-learning algorithms
MoRE	molecular recognition élément
MoRF	molecular recognition feature
MW	molecular-weight
NMR	nuclear magnetic résonance
NOE	nuclear Overhauser effect
N-WASP	actin-regulatory neural Wiskott-Aldrich syndrome protein
N _{TAIL}	C-terminal domain of measles virus nucleoprotein
PDB	protein data bank
pE-DB	protein ensemble database
PCS	primary contact site
PMG	pre-molten globule
PRE	paramagnetic relaxation enhancement
PreSMo	pre-structured motif
PSE	performed structural élément
PTM	posttranslational modification
RC	random coil
RDC	residual dipolar coupling
RPA70	replication protein A
SAXS	small angle X-ray scattering
SCS	secondary chemical shift
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SLiM	short linear motif
SM	single-molecule
TFE	trifluoroethanol
TMAO	trimethylamine N-oxide
XD	X domain

Acknowledgements

This work was carried out with the financial support of the Odysseus grant G.0029.12 from Research Foundation Flanders (FWO) to P.T., and of the Agence Nationale de la Recherche, specific programs "Physico-Chimie du Vivant", ANR-08-PCVI-0020-01, and "ASTRID", ANR-11-ASTR-003-01, to S.L.. The work was also supported in part by the Russian Academy of Sciences "Molecular and Cellular Biology" Program to V.N.U. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figure legends

Figure 1. A well-defined 3-D structure is required for enzyme activity. (A) The classical model of lock-and-key was formulated by Emil Fisher in 1894 to explain the stereo-specificity of enzyme catalysis.¹ (B) The model assumes that the substrates fits tightly to the binding site on the enzyme as a key into its lock.⁴²

Figure 2. Amino acid composition of disordered proteins. The differences between the amino acid compositions of disordered datasets (DisProt 1.0 and Disprot 3.4) and that of an ordered dataset (Globular-3D) were plotted as a function of the B-factor estimates of flexibility of residues. IDPs possess a tendency to be depleted in order-promoting amino acids and enriched in disorder-promoting amino acids.¹⁷

Figure 3. Charge-hydrophathy plot of protein disorder. Net charge vs. mean hydrophobicity is plotted for disordered (red circles) and ordered (blue squares) proteins. The two sets are separated by a straight line $\langle R \rangle = 2.743 \langle H \rangle - 1.109$ shown as a green line. Adapted with permission from ref. ¹⁷ Copyright 2008 Springer.

Figure 4. MeDor graphical output of the N-terminal domain of the measles virus phosphoprotein (PNT) (DisProt accession number DP00133). Predicted secondary structure elements as provided by the Pred2ary predictor, and the HCA plot, are shown above and below the amino acid sequence, respectively. Arrows below the HCA plot correspond to regions of predicted disorder. The highlighted region with enrichment in hydrophobic clusters encompassing the first 50 amino acids corresponds to an experimentally characterized α -MoRE.³⁰³

Figure 5. Binary predictors of intrinsic disorder. (A) CDF analysis. The dark blue curve located above the boundary represents the CDF curve of an ordered protein (β -glycosidase, UniProt ID : P07986), whereas the dark red line located below the boundary corresponds to the CDF curve of an IDP (secretogranin-1, UniProt ID: P23389). Here, δ_i and d_j (where i and j range from 1 to 7) are attributed to the ordered and disordered protein, respectively, and represent the distances of points at the CDF curve from the corresponding boundary points. The averaged distance of a given CDF curve from a

boundary line is calculated as $\Delta CDF = \frac{\sum_{i=1}^7 \delta_i}{7}$ or $\Delta CDF = \frac{\sum_{j=1}^7 d_j}{7}$. (B) CH-plot analysis. The dark red circle

located above the boundary corresponds to the same IDP as in (A), namely secretogranin-1 (UniProt ID: P23389), the dark blue square located below the boundary represents the same ordered protein as in (A), namely β -glycosidase (UniProt ID: P07986). (C) CH-CDF analysis. The dark red circle corresponds to the disordered protein, secretogranin-1 (UniProt ID: P23389), whereas the dark blue square represents the ordered protein β -glycosidase (UniProt ID: P07986). X-coordinates were calculated as averaged distances of corresponding CDF curves from a boundary (a positive ΔCDF distance corresponds to a protein predicted to be ordered by CDF analysis, while a negative ΔCDF distance corresponds to a protein predicted to be disordered by CDF analysis, see plot A). Y-coordinates were obtained as distances from spots corresponding to proteins to the boundary. Positive and negative ΔCH distances correspond to proteins predicted by the CH-plot to be disordered or ordered, respectively (see plot B). (D) CH-CDF plot for ribosomal proteins. Quadrants contain differently disordered proteins: the light red quadrant contains extended IDPs (predicted to be disordered by CDF and CH-plot analysis), the light pink quadrant contains native molten globules and/or hybrid proteins containing ordered and disordered regions (predicted to be disordered by CDF and ordered by CH-plot), the light blue quadrant contains globular proteins (predicted to be ordered by both CDF and CH-plot analyses), whereas the light cyan quadrant contains proteins predicted to be ordered by CDF and disordered by CH-plot.

Figure 6. Quantitative ensemble description of an IDP. (A) Cartoon representation of four conformational states of the α -MoRE of the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (N_{TAIL}), as derived by NMR spectroscopy through RDCs measurements. Modified from ref.⁶² (B) Secondary chemical shift differences values for the $^{13}C\alpha$ atoms of N_{TAIL} alone (blue bars) and in complex with the α -helical C-terminal domain of the phosphoprotein (P_{XD} ; red bars) with respect to a random coil chemical shift standard. In the free form, the values for the region encompassing the residues 90-110 (red bars) are shifted downfield (to larger ppm values) indicating a transiently populated α -helix in this region that is stabilized in the presence of P_{XD} (blue bars). Modified from ref.⁵⁸ (C) Comparison of experimental (blue) and back-calculated (red) DH-NH RDCs from the model of N_{TAIL} shown in (A). Adapted with permission from ref.⁶² Copyright 2011 National Academy of Sciences.

Figure 7. NMR PRE measurements in α -synuclein. Difference between the measured peak intensities (black bars) and the paramagnetic effect expected for a random coil (red line) suggests long-range N- to C-terminal tertiary structures in agreement with the observed broadening of the peak intensities. Adapted with permission from ref.³³² Copyright 2007 Elsevier.

Figure 8. NMR-based model allowing positioning of the disordered N_{TAIL} region of the measles virus nucleoprotein within the viral nucleocapsid. Front (A) and top (B) views of 13 N_{TAIL} conformers sampling an ensemble of conformations that point out from the surface of the viral nucleocapsid. Adapted with permission from ref.⁶² Copyright 2011 National Academy of Sciences.

Figure 9. The protein quartet model of protein function. Function can arise from four different conformations of the polypeptide chain, or from transitions between any of the states. Reprinted with permission from ref.¹³ Copyright 2002 John Wiley & Sons, Inc.

Figure 10. Mechanisms to prevent IDPs from degradation by default. IDPs can be stabilized by intermolecular interactions (A), interactions with gatekeepers (B), nanny proteins (C), DNA binding-sites (D), "decoy" DNA binding-sites (E), intramolecular interactions (F), local folding at the N-terminal region (G) or interactions with ribosome or ribosome-associated proteins (H). Reprinted with permission from ref.¹¹⁷ Copyright 2011 John Wiley & Sons, Inc.

Figure 11. Continuum of disorder. Functional disordered segments can be as small as only a few amino acids residues, or they can occupy rather long regions or ends. Different levels of order and disorder. From left to right, no disorder; disordered N- and C-termini; disordered linker; disordered loop; disordered domain; disordered protein with some residual structure; wholly disordered, mostly collapsed protein; wholly disordered, extended protein. Corresponding disordered regions are shown in red. Adapted with permission from ref.¹¹⁹ Copyright 2005 John Wiley & Sons, Inc.

Figure 12. Conformational behavior of proteins. Schematic representation of the physical principles of molecule separation by size-exclusion chromatography. The porous column matrix is shown by gray spheres. Small and large molecules are shown as pink and yellow spheres respectively. A vertical arrow on the left indicate the migration direction through the SEC matrix. Large molecules migrate faster than small molecules. At the right, relative hydrodynamic volumes occupied by a long polypeptide chain in four different conformations: RC: random coil, PMG: premolten globule, MG: molten globule and folded. Spheres show an increase in the Stokes radius when a protein possessing the same length is progressively unfolded.

Figure 13. Spectroscopic features determined by CD. (A) CD spectra representing different protein conformations are shown (α -helix, β -sheet and RC). RC-like conformations could be distinguished from secondary structures by a large negative ellipticity at 290 nm and an ellipticity close to 0 at

222nm. A negative ellipticity at 222 nm reveals the presence of transient secondary structures. Indeed, analysis of far-UV CD spectra in terms of double wavelength plot, $[\Theta]_{222}$ versus $[\Theta]_{200}$, (B), allows IDPs to be subdivided in coil-like (green diamonds) and PMG-like subclasses (blue circles).¹³

Figure 14. Distinctive properties of IDPs: diversity and functional role.

Figure 15. Examples of structurally divergent MoREs. MoREs (red ribbons) and partners (green surfaces) are shown. (A) An α -MoRF, Proteinase Inhibitor IA3, bound to Proteinase A (PDB entry 1DP5). (B) A β -MoRF, viral protein pVlc, bound to Human Adenovirus 2 Proteinase (PDB entry 1AVP). (C) An i-MoRF, Amphiphysin, bound to α -adaptin C (PDB entry 1KY7). (D) A complex-MoRF, β -amyloid precursor protein (β APP), bound to the PTB domain of the neuron specific protein X11 (PDB entry 1X11). Partner interfaces (gray surfaces) are also indicated. Reprinted with permission from ref.¹⁷ Copyright 2008 Springer.

Figure 16. Adaptability of IDPs to their partners. A chameleon sequence (the C-terminal region of p53) is shown in complex with four different partners and exhibiting four different structural forms. Reprinted with permission from ref.²¹³ Copyright 2008 Springer.

Figure 17. Schematic representation of coupled and decoupled affinity. For the binding of two globular proteins, affinity and specificity are expected to be proportional (Upper). When an induced folding occurs in case of IDPs (as for instance α -MoREs), affinity and specificity are decoupled due to the entropic cost of the folding.²⁰

Figure 18. Proposed mechanism by Espinoza-Fonseca for the synergistic model. Left: once the IDP is close enough to its target via “fly-casting”, conformer selection comes into play in that the target chooses a specific conformational state of the preformed region (oval) among the populated ones. Right: folding coupled to binding constitutes the final step where the MoRE is stabilized by the partner in the fully bound complex. Reprinted with permission from ref.²²⁶ Copyright 2009 Elsevier.

Figure 19. Fuzzy complexes. Static (A) and dynamic (B-D) disorder in protein-protein interactions are represented with the arrow indicating increasing disorder. The binding partner is rendered as a solid grey surface whereas the colored ribbon corresponds to the IDP in complex possessing residual disorder (dotted line). In the static model, there is more than one stable conformation whereas in the dynamic model parts or the entirety of the IDP remain disordered in the bound form. Reprinted with permission from ref.²³³ Copyright 2008 Elsevier Trends Journals.

Figure 20. Abundance of intrinsic disorder in disease-associated proteins. Percentages of disease-associated proteins with ≥ 30 to ≥ 100 consecutive residues predicted to be disordered. Corresponding data for signaling and ordered proteins (i.e. PDB_S25) are shown for the comparison. Analyzed sets of disease-related proteins included 1786, 487, 689, and 285 proteins for cancer, CVD, neurodegenerative disease and diabetes, respectively. Data were taken from ref.³²³

Figure 21. Folding, nonfolding or misfolding of a polypeptide chain. A polypeptide chain can either gain a well-defined 3-D structure, or can stay non-folded in the three different states (i.e. RC, PMG and MG), or also can misfold to form amyloid-like fibrils.³³³

References

- (1) Fisher, E. *Ber. Dt. Chem. Ges.* 1894, 27, 2985.
- (2) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* 1958, 44, 98.
- (3) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* 1958, 181, 662.
- (4) Berman, H. M. *Acta Crystallogr. A* 2008, 64, 88.
- (5) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* 2000, 28, 235.
- (6) Li, X.; Obradovic, Z.; Brown, C. J.; Garner, E. C.; Dunker, A. K. *Genome Inform Ser Workshop Genome Inform* 2000, 11, 172.
- (7) Chouard, T. *Nature* 2011, 471, 151.
- (8) Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guillot, S.; Dunker, A. K., *Pacific Symposium on Biocomputing 98*, Altman, R. B., Dunker, A. K., Hunter, L., Klein, T. E., Eds., World Scientific Pub Co: London, 1998, pp 437-48.
- (9) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* 1999, 293, 321.
- (10) 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.; Han, K.-H.; Jones, D. T.; Longhi, S.; Metallo, S. J.; Nishikawa, K.; Nussinov, R.; Obradovic, Z.; Pappu, R. V.; Rost, B.; Selenko, P.; Subramaniam, V.; Sussman, J. L.; Tompa, P.; Uversky, V. N. *Intrinsically Disordered Proteins* 2013, 1, e24157.
- (11) Tompa, P. *Trends Biochem. Sci.* 2002, 27, 527.
- (12) Uversky, V. N. *Eur. J. Biochem.* 2002, 269, 2.
- (13) Uversky, V. N. *Protein Sci.* 2002, 11, 739.
- (14) Uversky, V. N. *J. Biomed. Biotechnol.* 2010, 2010, 568068.
- (15) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* 2000, 41, 415.
- (16) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graph. Model.* 2001, 19, 26.
- (17) Dunker, A. K.; Oldfield, C. J.; Meng, J.; Romero, P.; Yang, J. Y.; Chen, J. W.; Vacic, V.; Obradovic, Z.; Uversky, V. N. *BMC Genomics* 2008, 9 Suppl 2, S1.
- (18) Campen, A.; Williams, R. M.; Brown, C. J.; Meng, J.; Uversky, V. N.; Dunker, A. K. *Protein Pept. Lett.* 2008, 15, 956.
- (19) Williams, R. M.; Obradovic, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Young, J.; Takayama, S.; Brown, C. J.; Dunker, A. K. *Pac. Symp. Biocomput.* 2001, 89.
- (20) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K. *Biochemistry* 2005, 44, 1989.
- (21) Csizmok, V.; Szollosi, E.; Friedrich, P.; Tompa, P. *Mol. Cell. Proteomics* 2006, 5, 265.
- (22) Romero, P.; Obradovic, Z.; Dunker, A. K. *FEBS Lett.* 1999, 462, 363.
- (23) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* 2001, 42, 38.
- (24) Lise, S.; Jones, D. T. *Proteins* 2005, 58, 144.
- (25) Longhi, S. *Methods Molecular Biology* 2010, 609, 307.
- (26) Ferron, F.; Longhi, S.; Canard, B.; Karlin, D. *Proteins* 2006, 65, 1.
- (27) Lieutaud, P.; Ferron, F.; Habchi, J.; Canard, B.; Longhi, S., *Advances in Protein and Peptide Sciences*, Dunn, B., Ed., Bentham Science Publishers (eBook), 2013, pp 441-492.
- (28) Bourhis, J. M.; Canard, B.; Longhi, S. *Curr. Protein Pept. Sci.* 2007, 8, 135.

- (29) Uversky, V. N.; Radivojac, P.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K. *Methods Mol. Biol.* 2007, 408, 69.
- (30) Prilusky, J.; Felder, C. E.; Zeev-Ben-Mordehai, T.; Rydberg, E. H.; Man, O.; Beckmann, J. S.; Silman, I.; Sussman, J. L. *Bioinformatics* 2005, 21, 3435.
- (31) Coeytaux, K.; Poupon, A. *Bioinformatics* 2005, 21, 1891.
- (32) Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J. *Nucleic Acids Res.* 2003, 31, 3701.
- (33) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* 2007, 35, D786.
- (34) Fukuchi, S.; Sakamoto, S.; Nobe, Y.; Murakami, S. D.; Amemiya, T.; Hosoda, K.; Koike, R.; Hiroaki, H.; Ota, M. *Nucleic Acids Res.* 2012, 40, D507.
- (35) Di Domenico, T.; Walsh, I.; Martin, A. J.; Tosatto, S. C. *Bioinformatics* 2012, 28, 2080.
- (36) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. *Structure (Camb)* 2003, 11, 1453.
- (37) Ward, J. J.; McGuffin, L. J.; Bryson, K.; Buxton, B. F.; Jones, D. T. *Bioinformatics* 2004, 20, 2138.
- (38) Yang, Z. R.; Thomson, R.; McNeil, P.; Esnouf, R. M. *Bioinformatics* 2005, 21, 3369.
- (39) Garbuzynskiy, S. O.; Lobanov, M. Y.; Galzitskaya, O. V. *Protein Sci.* 2004, 13, 2871.
- (40) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. *Bioinformatics* 2005, 21, 3433.
- (41) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. *J. Mol. Biol.* 2005, 347, 827.
- (42) Tompa, P., *Structure and Function of Intrinsically Disordered Proteins*, CRC Press, Taylor & Francis Group: Boca Raton, Florida, 2010.
- (43) Callebaut, I.; Labesse, G.; Durand, P.; Poupon, A.; Canard, L.; Chomilier, J.; Henrissat, B.; Mornon, J. P. *Cell. Mol. Life Sci.* 1997, 53, 621.
- (44) Xue, B.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. *FEBS Lett.* 2009, 583, 1469.
- (45) Mohan, A.; Sullivan, W. J., Jr.; Radivojac, P.; Dunker, A. K.; Uversky, V. N. *Mol. Biosyst.* 2008, 4, 328.
- (46) Huang, F.; Oldfield, C.; Meng, J.; Hsu, W. L.; Xue, B.; Uversky, V. N.; Romero, P.; Dunker, A. K. *Pac. Symp. Biocomput.* 2012, 128.
- (47) Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N. *Biochim. Biophys. Acta* 2010, 1804, 996.
- (48) Lieutaud, P.; Canard, B.; Longhi, S. *BMC Genomics* 2008, 9, S25.
- (49) Blocquel, D.; Habchi, J.; Gruet, A.; Blangy, S.; Longhi, S. *Mol. Biosyst.* 2012, 8, 392.
- (50) Kosol, S.; Contreras-Martos, S.; Cedeno, C.; Tompa, P. *Molecules* 2013, 18, 10802.
- (51) Dyson, H. J.; Wright, P. E. *Chem. Rev.* 2004, 104, 3607.
- (52) Bermel, W.; Bertini, I.; Chill, J.; Felli, I. C.; Haba, N.; Kumar, M. V. V.; Pierattelli, R. *ChemBioChem* 2012, 13, 2425.
- (53) Receveur-Bréchet, V.; Bourhis, J. M.; Uversky, V. N.; Canard, B.; Longhi, S. *Proteins* 2006, 62, 24.
- (54) Eliezer, D. *Curr. Opin. Struct. Biol.* 2009, 19, 23.
- (55) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* 2008, 128, 052204.
- (56) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. *Biochemistry* 2012, 51, 2224.
- (57) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* 2010, 132, 1270.

- (58) Gely, S.; Lowry, D. F.; Bernard, C.; Ringkjober-Jensen, M.; Blackledge, M.; Costanzo, S.; Darbon, H.; Daughdrill, G. W.; Longhi, S. *J. Mol. Recognit.* 2010, 23, 435.
- (59) Jensen, M. R.; Markwick, P. R.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernado, P.; Blackledge, M. *Structure* 2009, 17, 1169.
- (60) Dyson, H. J.; Wright, P. E. *Adv. Protein Chem.* 2002, 62, 311.
- (61) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* 2010, 132, 8407.
- (62) Jensen, M. R.; Communie, G.; Ribeiro, E. A., Jr.; Martinez, N.; Desfosses, A.; Salmon, L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; Ruigrok, R. W.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* 2011, 108, 9839.
- (63) Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* 2009, 131, 18314.
- (64) Fisher, C. K.; Stultz, C. M. *Curr. Opin. Struct. Biol.* 2011, 21, 426.
- (65) Belle, V.; Rouger, S.; Costanzo, S.; Longhi, S.; Fournel, A., *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation*, Uversky, V. N., Longhi, S., Eds., John Wiley and Sons: Hoboken, New Jersey 2010.
- (66) Habchi, J.; Martinho, M.; Gruet, A.; Guigliarelli, B.; Longhi, S.; Belle, V. *Methods Mol. Biol.* 2012, 895, 361.
- (67) Drescher, M. *Top. Curr. Chem.* 2012, 321, 91.
- (68) Lorenzi, M.; Puppo, C.; Lebrun, R.; Lignon, S.; Roubaud, V.; Martinho, M.; Mileo, E.; Tordo, P.; Marque, S. R.; Gontero, B.; Guigliarelli, B.; Belle, V. *Angew. Chem. Int. Ed. Engl.* 2011, 50, 9108.
- (69) Mileo, E.; Etienne, E.; Martinho, M.; Lebrun, R.; Roubaud, V.; Tordo, P.; Gontero, B.; Guigliarelli, B.; Marque, S. R.; Belle, V. *Bioconjug. Chem.* 2013, 24, 1110.
- (70) Wang, J.; Verkhivker, G. M. *Phys. Rev. Lett.* 2003, 90, 188101.
- (71) Chu, X.; Gan, L.; Wang, E.; Wang, J. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, E2342.
- (72) Wang, J.; Xu, L.; Wang, E. *Biophys. J.* 2007, 92, L109.
- (73) Wang, Y.; Chu, X.; Longhi, S.; Roche, P.; Wang, E.; Wang, J. *Proc. Natl. Acad. Sci. U.S.A.* 2013, in press.
- (74) Bernado, P.; Svergun, D., *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation*, Uversky, V. N., Longhi, S., Eds., John Wiley and Sons: Hoboken, New Jersey USA, 2010, pp 451-476.
- (75) Ando, T.; Kodera, N. *Methods Mol. Biol.* 2012, 896, 57.
- (76) Miyagi, A.; Tsunaka, Y.; Uchihashi, T.; Mayanagi, K.; Hirose, S.; Morikawa, K.; Ando, T. *Chemphyschem* 2008, 9, 1859.
- (77) Sandal, M.; Valle, F.; Tessari, I.; Mammi, S.; Bergantino, E.; Musiani, F.; Brucale, M.; Bubacco, L.; Samori, B. *PLoS Biol.* 2008, 6, e6.
- (78) Ferreon, A. C.; Moran, C. R.; Gambin, Y.; Deniz, A. A. *Methods Enzymol.* 2010, 472, 179.
- (79) Ferreon, A. C.; Ferreon, J. C.; Wright, P. E.; Deniz, A. A. *Nature* 2013, 498, 390.
- (80) Milles, S.; Koehler, C.; Gambin, Y.; Deniz, A. A.; Lemke, E. A. *Mol. Biosyst.* 2012, 8, 2531.
- (81) Ando, T.; Uchihashi, T.; Kodera, N. *Annu. Rev. Biophys.* 2013, 42, 393.
- (82) Dunker, A. K.; Obradovic, Z. *Nat. Biotechnol.* 2001, 19, 805.
- (83) Greenblatt, J.; Li, J. *J. Biol. Chem.* 1982, 257, 362.
- (84) Lisse, T.; Bartels, D.; Kalbitzer, H. R.; Jaenicke, R. *Biol. Chem.* 1996, 377, 555.
- (85) Kriwacki, R. W.; Wu, J.; Siuzdak, G.; Wright, P. E. *J. Am. Chem. Soc.* 1996, 118, 5320.

- (86) Kriwacki, R. W.; Wu, J.; Tennant, L.; Wright, P. E.; Siuzdak, G. *J. Chromatogr. A* 1997, 777, 23.
- (87) Fontana, A.; Polverino de Laureto, P.; Spolaore, B.; Frare, E.; Zambonin, M., *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation*, Uversky, V., Longhi, S., Eds., John Wiley and Sons: Hoboken, New Jersey, 2010, pp 569-626.
- (88) Hernandez, M. A.; Avila, J.; Andreu, J. M. *Eur. J. Biochem.* 1986, 154, 41.
- (89) Weinreb, P. H.; Zhen, W.; Poon, A. W.; Conway, K. A.; Lansbury, P. T., Jr. *Biochemistry* 1996, 35, 13709.
- (90) Etoh, Y.; Simon, M.; Green, H. *Biochem. Biophys. Res. Commun.* 1986, 136, 51.
- (91) Lynch, W. P.; Riseman, V. M.; Bretscher, A. *J. Biol. Chem.* 1987, 262, 7429.
- (92) Livernois, A. M.; Hnatchuk, D. J.; Findlater, E. E.; Graether, S. P. *Anal. Biochem.* 2009, 392, 70.
- (93) Kalthoff, C. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2003, 786, 247.
- (94) Cortese, M. S.; Baird, J. P.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* 2005, 4, 1610.
- (95) Uversky, V. N.; Longhi, S., *Instrumental analysis of intrinsically disordered proteins*, John Wiley & Sons Hoboken, 2010.
- (96) Ellis, R. J. *Trends Biochem. Sci.* 2001, 26, 597.
- (97) Minton, A. P. *Biophys. J.* 2005, 88, 971.
- (98) Bolen, D. W.; Baskakov, I. V. *J. Mol. Biol.* 2001, 310, 955.
- (99) Munishkina, L. A.; Cooper, E. M.; Uversky, V. N.; Fink, A. L. *J. Mol. Recognit.* 2004, 17, 456.
- (100) Breydo, L.; Reddy, K. D.; Piai, A.; Felli, I. C.; Pierattelli, R.; Uversky, V. N. *Biochim. Biophys. Acta* 2014, 1844, 346.
- (101) Szasz, C. S.; Alexa, A.; Toth, K.; Rakacs, M.; Langowski, J.; Tompa, P. *Biochemistry* 2011, 50, 5834.
- (102) Asher, G.; Tsvetkov, P.; Kahana, C.; Shaul, Y. *Genes Dev.* 2005, 19, 316.
- (103) Sheaff, R. J.; Singer, J. D.; Swanger, J.; Smitherman, M.; Roberts, J. M.; Clurman, B. E. *Mol. Cell* 2000, 5, 403.
- (104) Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; Shirakawa, M. *Nature* 2009, 458, 106.
- (105) Bodart, J. F.; Wieruszkeski, J. M.; Amniai, L.; Leroy, A.; Landrieu, I.; Rousseau-Lescuyer, A.; Vilain, J. P.; Lippens, G. *J. Magn. Reson.* 2008, 192, 252.
- (106) Binolfi, A.; Theillet, F. X.; Selenko, P. *Biochem. Soc. Trans.* 2012, 40, 950.
- (107) Waudby, C. A.; Camilloni, C.; Fitzpatrick, A. W.; Cabrita, L. D.; Dobson, C. M.; Vendruscolo, M.; Christodoulou, J. *PLoS One* 2013, 8, e72286.
- (108) McNulty, B. C.; Young, G. B.; Pielak, G. J. *J. Mol. Biol.* 2006, 355, 893.
- (109) Belle, A.; Tanay, A.; Bitincka, L.; Shamir, R.; O'Shea, E. K. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 13004.
- (110) Tompa, P.; Prilusky, J.; Silman, I.; Sussman, J. L. *Proteins* 2008, 71, 903.
- (111) Prakash, S.; Tian, L.; Ratliff, K. S.; Lehotzky, R. E.; Matouschek, A. *Nat. Struct. Mol. Biol.* 2004, 11, 830.
- (112) Eralles, J.; Coffino, P. *Biochim. Biophys. Acta* 2014, 1843, 216.
- (113) Hegyi, H.; Tompa, P. *PLoS Comput. Biol.* 2008, 4, e1000017.
- (114) Gsponer, J.; Futschik, M. E.; Teichmann, S. A.; Babu, M. M. *Science* 2008, 322, 1365.
- (115) Vavouri, T.; Semple, J. I.; Garcia-Verdugo, R.; Lehner, B. *Cell* 2009, 138, 198.

- (116) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. *Curr. Opin. Struct. Biol.* 2011, 21, 432.
- (117) Suskiewicz, M. J.; Sussman, J. L.; Silman, I.; Shaul, Y. *Protein Sci.* 2011, 20, 1285.
- (118) Simister, P. C.; Schaper, F.; O'Reilly, N.; McGowan, S.; Feller, S. M. *PLoS Biol.* 2011, 9, e1000591.
- (119) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *J. Mol. Recognit.* 2005, 18, 343.
- (120) Uversky, V. N. *Biochim. Biophys. Acta* 2013, 1834, 932.
- (121) Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K. *Proteins* 2003, 53 Suppl 6, 566.
- (122) Le Gall, T.; Romero, P. R.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *J. Biomol. Struct. Dyn.* 2007, 24, 325.
- (123) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* 2005, 6, 197.
- (124) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* 2010, 107, 8183.
- (125) Das, R. K.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, 13392.
- (126) Tcherkasskaya, O.; Uversky, V. N. *Protein Pept. Lett.* 2003, 10, 239.
- (127) Marsh, J. A.; Forman-Kay, J. D. *Biophys. J.* 2010, 98, 2383.
- (128) Meng, W.; Lyle, N.; Luan, B.; Raleigh, D. P.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, 2123.
- (129) Uversky, V. N.; Dunker, A. K. *Biochim. Biophys. Acta* 2010, 1804, 1231.
- (130) Uversky, V. N.; Dunker, A. K. *Anal. Chem.* 2012, 84, 2096.
- (131) Fasman, G. D., *Circular dichroism and conformational analysis of biomolecules*, Plenum Press: New York, 1996.
- (132) Li, P.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q. X.; Nixon, B. T.; Rosen, M. K. *Nature* 2012, 483, 336.
- (133) Tompa, P. *Intrinsically Disordered Proteins* 2013, 1, e24068.
- (134) Bernado, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. J. *Am. Chem. Soc.* 2005, 127, 17968.
- (135) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* 2007, 129, 5656.
- (136) Schneider, R.; Huang, J. R.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon, L.; Jensen, M. R.; Blackledge, M. *Mol. Biosyst.* 2012, 8, 58.
- (137) Huang, J. R.; Grzesiek, S. *J. Am. Chem. Soc.* 2010, 132, 694.
- (138) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J. R.; Jensen, M. R.; Segard, S.; Bernado, P.; Charavay, C.; Blackledge, M. *Bioinformatics* 2012, 28, 1463.
- (139) Bernado, P.; Svergun, D. I. *Mol. Biosyst.* 2012, 8, 151.
- (140) Feldman, H. J.; Hogue, C. W. *Proteins* 2000, 39, 112.
- (141) Feldman, H. J.; Hogue, C. W. *Proteins* 2002, 46, 8.
- (142) Eyal, E.; Najmanovich, R.; McConkey, B. J.; Edelman, M.; Sobolev, V. J. *Comput. Chem.* 2004, 25, 712.
- (143) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. *Protein Sci.* 2003, 12, 2001.
- (144) Svergun, D. I.; Barabero, C.; Koch, M. H. *J Appl Cryst* 1995, 28, 768.
- (145) Shen, Y.; Bax, A. *J. Biomol. NMR* 2007, 38, 289.
- (146) Daughdrill, G. W.; Kashtanov, S.; Stancik, A.; Hill, S. E.; Helms, G.; Muschol, M.; Receveur-Brechot, V.; Ytreberg, F. M. *Mol. Biosyst.* 2012, 8, 308.

- (147) Schwalbe, M.; Ozenne, V.; Bibow, S.; Jaremko, M.; Jaremko, L.; Gajda, M.; Jensen, M. R.; Biernat, J.; Becker, S.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *Structure* 2013.
- (148) Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I. C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; Sussman, J.; Svergun, D. I.; Uversky, V. N.; Vendruscolo, M.; Wishart, D.; Wright, P. E.; Tompa, P. *Nucleic Acids Res.* 2014, 42, D326.
- (149) Tompa, P. *Curr. Opin. Struct. Biol.* 2011, 21, 419.
- (150) Tran, H. T.; Mao, A.; Pappu, R. V. *J. Am. Chem. Soc.* 2008, 130, 7380.
- (151) Szollosi, E.; Bokor, M.; Bodor, A.; Perczel, A.; Klement, E.; Medzihradsky, K. F.; Tompa, K.; Tompa, P. *J. Proteome Res.* 2008, 7, 2291.
- (152) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rügger, S.; Reymond, L.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A* 2010, 107, 14609.
- (153) Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Spät, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* 2012, 109, 17800.
- (154) Brocca, S.; Samalikova, M.; Uversky, V. N.; Lotti, M.; Vanoni, M.; Alberghina, L.; Grandori, R. *Proteins* 2009, 76, 731.
- (155) Li, J.; Uversky, V. N.; Fink, A. L. *Biochemistry* 2001, 40, 11604.
- (156) Uversky, V. N.; Li, J.; Souillac, P.; Jakes, R.; Goedert, M.; Fink, A. L. *J. Biol. Chem.* 2002, 277, 25.
- (157) Guo, J. T.; Jaromczyk, J. W.; Xu, Y. *Proteins* 2007, 67, 548.
- (158) Zhang, Y.; Stec, B.; Godzik, A. *Structure* 2007, 15, 1141.
- (159) Hazy, E.; Tompa, P. *Chemphyschem* 2009, 10, 1415.
- (160) Uversky, V. N. *Biochim. Biophys. Acta* 2011, 1814, 693.
- (161) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* 2002, 41, 6573.
- (162) Mukhopadhyay, R.; Hoh, J. H. *FEBS Lett.* 2001, 505, 374.
- (163) Patel, S. S.; Belmont, B. J.; Sante, J. M.; Rexach, M. F. *Cell* 2007, 129, 83.
- (164) Kriwacki, R. W.; Hengst, L.; Tennant, L.; Reed, S. I.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A* 1996, 93, 11504.
- (165) Russo, A. A.; Jeffrey, P. D.; Patten, A. K.; Massague, J.; Pavletich, N. P. *Nature* 1996, 382, 325.
- (166) Daughdrill, G. W.; Chadsey, M. S.; Karlinsey, J. E.; Hughes, K. T.; Dahlquist, F. W. *Nat. Struct. Biol.* 1997, 4, 285.
- (167) Peti, W.; Nairn, A. C.; Page, R. *FEBS J.* 2013, 280, 596.
- (168) Peti, W.; Nairn, A. C.; Page, R. *Curr. Phys. Chem.* 2012, 2, 107.
- (169) Tompa, P.; Szasz, C.; Buday, L. *Trends Biochem. Sci.* 2005, 30, 484.
- (170) Lu, Y.; Bennick, A. *Arch. Oral Biol.* 1998, 43, 717.
- (171) Kalmar, L.; Homola, D.; Varga, G.; Tompa, P. *Bone* 2012, 51, 528.
- (172) Toth-Petroczy, A.; Oldfield, C. J.; Simon, I.; Takagi, Y.; Dunker, A. K.; Uversky, V. N.; Fuxreiter, M. *PLoS Comput. Biol.* 2008, 4, e1000243.
- (173) Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *Prog. Biophys. Mol. Biol.* 2008, 98, 85.
- (174) Balazs, A.; Csizmok, V.; Buday, L.; Rakacs, M.; Kiss, R.; Bokor, M.; Udupa, R.; Tompa, K.; Tompa, P. *FEBS J.* 2009, 276, 3744.
- (175) Dosztanyi, Z.; Chen, J.; Dunker, A. K.; Simon, I.; Tompa, P. *J. Proteome Res.* 2006, 5, 2985.

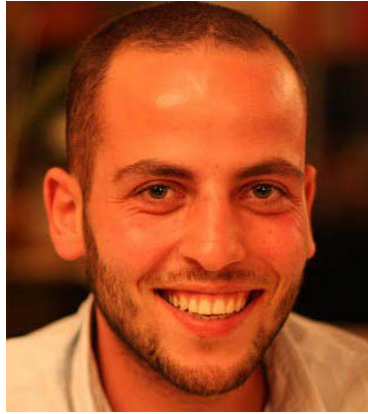
- (176) Haynes, C.; Oldfield, C. J.; Ji, F.; Klitgord, N.; Cusick, M. E.; Radivojac, P.; Uversky, V. N.; Vidal, M.; Iakoucheva, L. M. *PLoS Comput. Biol.* 2006, 2, e100.
- (177) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. *FEBS J.* 2005, 272, 5129.
- (178) Barbar, E. *Biochemistry* 2008, 47, 503.
- (179) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* 2004, 32, 1037.
- (180) Radivojac, P.; Vacic, V.; Haynes, C.; Cocklin, R. R.; Mohan, A.; Heyen, J. W.; Goebel, M. G.; Iakoucheva, L. M. *Proteins* 2010, 78, 365.
- (181) Tompa, P.; Csermely, P. *FASEB J.* 2004, 18, 1169.
- (182) Kim, T. D.; Paik, S. R.; Yang, C. H. *Biochemistry* 2002, 41, 13782.
- (183) Bhattacharyya, J.; Das, K. P. *J. Biol. Chem.* 1999, 274, 15505.
- (184) Chakrabortee, S.; Boschetti, C.; Walton, L. J.; Sarkar, S.; Rubinsztein, D. C.; Tunnacliffe, A. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 18073.
- (185) Kovacs, D.; Kalmar, E.; Torok, Z.; Tompa, P. *Plant Physiol.* 2008, 147, 381.
- (186) Andley, U. P.; Mathur, S.; Griest, T. A.; Petrash, J. M. *J. Biol. Chem.* 1996, 271, 31973.
- (187) Pasta, S. Y.; Raman, B.; Ramakrishna, T.; Rao Ch, M. *J. Biol. Chem.* 2002, 277, 45821.
- (188) Santner, A. A.; Croy, C. H.; Vasanwala, F. H.; Uversky, V. N.; Van, Y. Y.; Dunker, A. K. *Biochemistry* 2012, 51, 7250.
- (189) Reichmann, D.; Xu, Y.; Cremers, C. M.; Ilbert, M.; Mittelman, R.; Fitzgerald, M. C.; Jakob, U. *Cell* 2012, 148, 947.
- (190) Neduva, V.; Linding, R.; Su-Angrand, I.; Stark, A.; Masi, F. D.; Gibson, T. J.; Lewis, J.; Serrano, L.; Russell, R. B. *PLoS Biol.* 2005, 3, e405.
- (191) Fuxreiter, M.; Tompa, P.; Simon, I. *Bioinformatics* 2007, 23, 950.
- (192) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Romero, P.; Uversky, V. N.; Dunker, A. K. *Biochemistry* 2005, 44, 12454.
- (193) Neduva, V.; Russell, R. B. *FEBS Lett.* 2005, 579, 3342.
- (194) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J. Mol. Biol.* 2004, 338, 1015.
- (195) Lee, S. H.; Kim, D. H.; Han, J. J.; Cha, E. J.; Lim, J. E.; Cho, Y. J.; Lee, C.; Han, K. H. *Curr. Protein Pept. Sci.* 2012, 13, 34.
- (196) Fuxreiter, M. *Mol. Biosyst.* 2012, 8, 168.
- (197) Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.; Uversky, V. N. *J. Mol. Biol.* 2006, 362, 1043.
- (198) Vacic, V.; Oldfield, C. J.; Mohan, A.; Radivojac, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* 2007, 6, 2351.
- (199) Puntervoll, P.; Linding, R.; Gemund, C.; Chabanis-Davidson, S.; Mattingsdal, M.; Cameron, S.; Martin, D. M.; Ausiello, G.; Brannetti, B.; Costantini, A.; Ferre, F.; Maselli, V.; Via, A.; Cesareni, G.; Diella, F.; Superti-Furga, G.; Wyrwicz, L.; Ramu, C.; McGuigan, C.; Gudavalli, R.; Letunic, I.; Bork, P.; Rychlewski, L.; Kuster, B.; Helmer-Citterich, M.; Hunter, W. N.; Aasland, R.; Gibson, T. J. *Nucleic Acids Res.* 2003, 31, 3625.
- (200) Davey, N. E.; Shields, D. C.; Edwards, R. J. *Nucleic Acids Res.* 2006, 34, 3546.
- (201) Fuxreiter, M.; Tompa, P., *Fuzziness: structural disorder in protein complexes*, Springer Science + Business Media, LLC Landes Bioscience: New York, USA, 2012.
- (202) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* 2002, 12, 54.

- (203) Tompa, P.; Fuxreiter, M.; Oldfield, C. J.; Simon, I.; Dunker, A. K.; Uversky, V. N. *Bioessays* 2009, 31, 328.
- (204) Chen, J. W.; Romero, P.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* 2006, 5, 888.
- (205) Chen, J. W.; Romero, P.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* 2006, 5, 879.
- (206) Xue, B.; Brown, C. J.; Dunker, A. K.; Uversky, V. N. *Biochim. Biophys. Acta* 2013, 1834, 725.
- (207) Dosztanyi, Z.; Meszaros, B.; Simon, I. *Bioinformatics* 2009, 25, 2745.
- (208) Cilia, E.; Pancsa, R.; Tompa, P.; Lenaerts, T.; Vranken, W. F. *Nat. Commun.* 2013, 4, 2741.
- (209) Meszaros, B.; Simon, I.; Dosztanyi, Z. *PLoS Comput. Biol.* 2009, 5, e1000376.
- (210) Hammes, G. G.; Chang, Y. C.; Oas, T. G. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106, 13737.
- (211) Wang, Y.; Tang, C.; Wang, E.; Wang, J. *PLoS Comput. Biol.* 2012, 8, e1002471.
- (212) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* 2009, 19, 31.
- (213) Oldfield, C. J.; Meng, J.; Yang, J. Y.; Yang, M. Q.; Uversky, V. N.; Dunker, A. K. *BMC Genomics* 2008, 9 Suppl 1, S1.
- (214) Pancsa, R.; Fuxreiter, M. *IUBMB Life* 2012, 64, 513.
- (215) Tsai, C. D.; Ma, B.; Kumar, S.; Wolfson, H.; Nussinov, R. *Crit. Rev. Biochem. Mol. Biol.* 2001, 36, 399.
- (216) Tsai, C. J.; Ma, B.; Sham, Y. Y.; Kumar, S.; Nussinov, R. *Proteins* 2001, 44, 418.
- (217) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 17327.
- (218) Fersht, A. R.; Sato, S. *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 7976.
- (219) Kiefhaber, T.; Bachmann, A.; Jensen, K. S. *Curr. Opin. Struct. Biol.* 2012, 22, 21.
- (220) Rogers, J. M.; Steward, A.; Clarke, J. *J. Am. Chem. Soc.* 2013, 135, 1415.
- (221) Shammass, S. L.; Travis, A. J.; Clarke, J. *J. Phys. Chem. B* 2013.
- (222) Giri, R.; Morrone, A.; Toto, A.; Brunori, M.; Gianni, S. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, 14942.
- (223) Huang, Y.; Liu, Z. *J. Mol. Biol.* 2009, 393, 1143.
- (224) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* 2000, 97, 8868.
- (225) Csizmok, V.; Bokor, M.; Banki, P.; Klement, E.; Medzihradzsky, K. F.; Friedrich, P.; Tompa, K.; Tompa, P. *Biochemistry* 2005, 44, 3955.
- (226) Espinoza-Fonseca, L. M. *Biochem. Biophys. Res. Commun.* 2009, 382, 479.
- (227) Radhakrishnan, I.; Perez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* 1997, 91, 741.
- (228) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* 2007, 447, 1021.
- (229) Savvides, S. N.; Raghunathan, S.; Futterer, K.; Kozlov, A. G.; Lohman, T. M.; Waksman, G. *Protein Sci.* 2004, 13, 1942.
- (230) Sigalov, A. B.; Zhuravleva, A. V.; Orekhov, V. Y. *Biochimie* 2007, 89, 419.
- (231) Sigalov, A. B.; Kim, W. M.; Saline, M.; Stern, L. J. *Biochemistry* 2008, 47, 12942.
- (232) Communie, G.; Habchi, J.; Yabukarski, F.; Blocquel, D.; Schneider, R.; Tarbouriech, N.; Papageorgiou, N.; Ruigrok, R. W.; Jamin, M.; Jensen, M. R.; Longhi, S.; Blackledge, M. *PLoS Pathog.* 2013, 9, e1003631.
- (233) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* 2008, 33, 2.
- (234) Meszaros, B.; Tompa, P.; Simon, I.; Dosztanyi, Z. *J. Mol. Biol.* 2007, 372, 549.
- (235) Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* 1999, 285, 2177.
- (236) Shammass, S. L.; Rogers, J. M.; Hill, S. A.; Clarke, J. *Biophys. J.* 2012, 103, 2203.
- (237) Huang, Y.; Liu, Z. *Chemistry (Easton)* 2013, 19, 4462.

- (238) Wong, E. T.; Na, D.; Gsponer, J. *PLoS Comput. Biol.* 2013, 9, e1003192.
- (239) Brown, C. J.; Johnson, A. K.; Dunker, A. K.; Daughdrill, G. W. *Curr. Opin. Struct. Biol.* 2011, 21, 441.
- (240) Davey, N. E.; Cowan, J. L.; Shields, D. C.; Gibson, T. J.; Coldwell, M. J.; Edwards, R. J. *Nucleic Acids Res.* 2012, 40, 10628.
- (241) Espinoza-Fonseca, L. M. *Mol. Biosyst.* 2012, 8, 237.
- (242) Fuxreiter, M.; Tompa, P. *Trends Biochem. Sci.* 2009, 34, 3.
- (243) Polverini, E.; Rangaraj, G.; Libich, D. S.; Boggs, J. M.; Harauz, G. *Biochemistry* 2008, 47, 267.
- (244) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr. *Angew. Chem. Int. Ed. Engl.* 2005, 44, 7342.
- (245) Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G. *Nat. Methods* 2007, 4, 798.
- (246) Yang, X. J. *Oncogene* 2005, 24, 1653.
- (247) Faradonbeh, M. Z.; Gharechahi, J.; Mollamohammadi, S.; Pakzad, M.; Taei, A.; Rassouli, H.; Baharvand, H.; Salekdeh, G. H. *Mol. Biosyst.* 2012, 8, 1833.
- (248) Mersfelder, E. L.; Parthun, M. R. *Nucleic Acids Res.* 2006, 34, 2653.
- (249) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.; Uversky, V. N. *J. Proteome Res.* 2007, 6, 1917.
- (250) Dunker, A. K.; Brown, C. J.; Obradovic, Z. *Adv. Protein Chem.* 2002, 62, 25.
- (251) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Obradovic, Z. *J. Proteome Res.* 2007, 6, 1882.
- (252) Uversky, V. N. *Curr. Pharm. Des.* 2013, 19, 4191.
- (253) Sambrook, J. *Nature* 1977, 268, 101.
- (254) Black, D. L. *Annu. Rev. Biochem* 2003, 72, 291.
- (255) Boue, S.; Letunic, I.; Bork, P. *Bioessays* 2003, 25, 1031.
- (256) Ast, G. *Nat. Rev. Genet.* 2004, 5, 773.
- (257) Kelemen, O.; Convertini, P.; Zhang, Z.; Wen, Y.; Shen, M.; Falaleeva, M.; Stamm, S. *Gene* 2013, 514, 1.
- (258) Stamm, S.; Ben-Ari, S.; Rafalska, I.; Tang, Y.; Zhang, Z.; Toiber, D.; Thanaraj, T. A.; Soreq, H. *Gene* 2005, 344, 1.
- (259) Brett, D.; Hanke, J.; Lehmann, G.; Haase, S.; Delbruck, S.; Krueger, S.; Reich, J.; Bork, P. *FEBS Lett.* 2000, 474, 83.
- (260) Johnson, J. M.; Castle, J.; Garrett-Engele, P.; Kan, Z.; Loerch, P. M.; Armour, C. D.; Santos, R.; Schadt, E. E.; Stoughton, R.; Shoemaker, D. D. *Science* 2003, 302, 2141.
- (261) Graveley, B. R. *Trends Genet.* 2001, 17, 100.
- (262) Minneman, K. P. *Mol. Interv.* 2001, 1, 108.
- (263) Thai, T. H.; Kearney, J. F. *J. Immunol.* 2004, 173, 4009.
- (264) Thai, T. H.; Kearney, J. F. *Adv. Immunol.* 2005, 86, 113.
- (265) Scheper, W.; Zwart, R.; Baas, F. *Neurogenetics* 2004, 5, 223.
- (266) Romero, P. R.; Zaidi, S.; Fang, Y. Y.; Uversky, V. N.; Radivojac, P.; Oldfield, C. J.; Cortese, M. S.; Sickmeier, M.; LeGall, T.; Obradovic, Z.; Dunker, A. K. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 8390.
- (267) Buljan, M.; Chalancon, G.; Dunker, A. K.; Bateman, A.; Balaji, S.; Fuxreiter, M.; Babu, M. M. *Curr. Opin. Struct. Biol.* 2013, 23, 443.
- (268) Liu, S.; Cheng, C. *Wiley Interdiscip. Rev. RNA* 2013, 4, 547.
- (269) Skotheim, R. I.; Nees, M. *Int. J. Biochem. Cell Biol.* 2007, 39, 1432.
- (270) Fu, R. H.; Liu, S. P.; Huang, S. J.; Chen, H. J.; Chen, P. R.; Lin, Y. H.; Ho, Y. C.; Chang, W. L.; Tsai, C. H.; Shyu, W. C.; Lin, S. Z. *Cell Transplant.* 2013, 22, 653.

- (271) Niblock, M.; Gallo, J. M. *Biochem. Soc. Trans.* 2012, 40, 677.
- (272) Mills, J. D.; Janitz, M. *Neurobiol. Aging* 2012, 33, 1012.e11.
- (273) Lara-Pezzi, E.; Gomez-Salinerio, J.; Gatto, A.; Garcia-Pavia, P. *J. Cardiovasc. Transl. Res.* 2013, 6, 945.
- (274) Venables, J. P.; Burn, J. *Nucleic Acids Res.* 2006, 34, e103.
- (275) Hastings, M. L.; Resta, N.; Traum, D.; Stella, A.; Guanti, G.; Krainer, A. R. *Nat. Struct. Mol. Biol.* 2005, 12, 54.
- (276) Orban, T. I.; Olah, E. *Mol. Pathol.* 2003, 56, 191.
- (277) Uversky, V. N. *Front. Biosci. (Landmark Ed)* 2009, 14, 5188.
- (278) Koshland, D. E., Jr.; Nemethy, G.; Filmer, D. *Biochemistry* 1966, 5, 365.
- (279) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* 1965, 12, 88.
- (280) Hilser, V. J.; Thompson, E. B. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 8311.
- (281) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. *Annu. Rev. Biophys.* 2012, 41, 585.
- (282) Garcia-Pino, A.; Balasubramanian, S.; Wyns, L.; Gazit, E.; De Greve, H.; Magnuson, R. D.; Charlier, D.; van Nuland, N. A.; Loris, R. *Cell* 2010, 142, 101.
- (283) Motlagh, H. N.; Hilser, V. J. *Proc. Natl. Acad. Sci. U.S.A.* 2012, 109, 4134.
- (284) Follis, A. V.; Chipuk, J. E.; Fisher, J. C.; Yun, M. K.; Grace, C. R.; Nourse, A.; Baran, K.; Ou, L.; Min, L.; White, S. W.; Green, D. R.; Kriwacki, R. W. *Nat. Chem. Biol.* 2013, 9, 163.
- (285) Gruet, A.; Dosnon, M.; Vassena, A.; Lombard, V.; Gerlier, D.; Bignon, C.; Longhi, S. *J. Mol. Biol.* 2013, 425, 3495.
- (286) Tompa, P. *Bioessays* 2003, 25, 847.
- (287) Kato, M.; Han, T. W.; Xie, S.; Shi, K.; Du, X.; Wu, L. C.; Mirzaei, H.; Goldsmith, E. J.; Longgood, J.; Pei, J.; Grishin, N. V.; Frantz, D. E.; Schneider, J. W.; Chen, S.; Li, L.; Sawaya, M. R.; Eisenberg, D.; Tycko, R.; McKnight, S. L. *Cell* 2012, 149, 753.
- (288) Beck, R.; Deek, J.; Safinya, C. R. *Biochem. Soc. Trans.* 2012, 40, 1027.
- (289) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* 2004, 337, 635.
- (290) Schad, E.; Tompa, P.; Hegyi, H. *Genome Biol.* 2011, 12, R120.
- (291) Tompa, P.; Dosztanyi, Z.; Simon, I. *J. Proteome Res.* 2006, 5, 1996.
- (292) Burra, P. V.; Kalmar, L.; Tompa, P. *PLoS One* 2010, 5, e12069.
- (293) Pancsa, R.; Tompa, P. *PLoS One* 2012, 7, e34687.
- (294) Davey, N. E.; Trave, G.; Gibson, T. J. *Trends Biochem. Sci.* 2011.
- (295) Tokuriki, N.; Oldfield, C. J.; Uversky, V. N.; Berezovsky, I. N.; Tawfik, D. S. *Trends Biochem. Sci.* 2009, 34, 53.
- (296) Jordan, I. K.; Sutter, B. A.; McClure, M. A. *Mol. Biol. Evol.* 2000, 17, 75.
- (297) Narechania, A.; Terai, M.; Burk, R. D. *J. Gen. Virol.* 2005, 86, 1307.
- (298) Rancurel, C.; Khosravi, M.; Dunker, K. A.; Romero, P. R.; Karlin, D. *J. Virol.* 2009, 83, 10719.
- (299) Bourhis, J. M.; Canard, B.; Longhi, S. *Virology* 2006, 344, 94.
- (300) Longhi, S. *Protein Pept. Lett.* 2010, 17, 930.
- (301) Longhi, S.; Oglesbee, M. *Protein Pept. Lett.* 2010, 17, 961.
- (302) Habchi, J.; Longhi, S. *Mol. Biosyst.* 2012, 8, 69.
- (303) Karlin, D.; Longhi, S.; Receveur, V.; Canard, B. *Virology* 2002, 296, 251.
- (304) Longhi, S.; Receveur-Brechot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. *J. Biol. Chem.* 2003, 278, 18638.
- (305) Karlin, D.; Ferron, F.; Canard, B.; Longhi, S. *J. Gen. Virol.* 2003, 84, 3239.

- (306) Xue, B.; Williams, R. W.; Oldfield, C. J.; Goh, G. K.; Dunker, A. K.; Uversky, V. N. *Protein Pept. Lett.* 2010, 17, 932.
- (307) Uversky, V. N.; Longhi, S., *Flexible viruses: structural disorder in viral proteins*, Uversky, V. N., Longhi, S.; John Wiley and Sons: Hoboken, 2012.
- (308) Alves, C.; Cunha, C. *Future Virol.* 2012, 7, 1183.
- (309) Xue, B.; Dunker, A. K.; Uversky, V. N. *J. Biomol. Struct. Dyn.* 2012, 30, 137.
- (310) Galea, C. A.; High, A. A.; Obenauer, J. C.; Mishra, A.; Park, C. G.; Punta, M.; Schlessinger, A.; Ma, J.; Rost, B.; Slaughter, C. A.; Kriwacki, R. W. *J. Proteome Res.* 2009, 8, 211.
- (311) Tompa, P. *Trends Biochem. Sci.* 2012, 37, 509.
- (312) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. *Genome Inform. Ser. Workshop Genome Inform.* 2000, 11, 161.
- (313) Galea, C. A.; Wang, Y.; Sivakolundu, S. G.; Kriwacki, R. W. *Biochemistry* 2008, 47, 7598.
- (314) Huber, A. H.; Stewart, D. B.; Laurents, D. V.; Nelson, W. J.; Weis, W. I. *J. Biol. Chem.* 2001, 276, 12301.
- (315) Paunola, E.; Mattila, P. K.; Lappalainen, P. *FEBS Lett.* 2002, 513, 92.
- (316) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. *J. Mol. Biol.* 2002, 323, 573.
- (317) Brown, C. J.; Takayama, S.; Campen, A. M.; Vise, P.; Marshall, T. W.; Oldfield, C. J.; Williams, C. J.; Dunker, A. K. *J. Mol. Evol.* 2002, 55, 104.
- (318) Bellay, J.; Han, S.; Michaut, M.; Kim, T.; Costanzo, M.; Andrews, B. J.; Boone, C.; Bader, G. D.; Myers, C. L.; Kim, P. M. *Genome Biol.* 2011, 12, R14.
- (319) Tokuriki, N.; Tawfik, D. S. *Science* 2009, 324, 203.
- (320) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. *PLoS Comput. Biol.* 2008, 4, e1000002.
- (321) Olson, K. E.; Narayanaswami, P.; Vise, P. D.; Lowry, D. F.; Wold, M. S.; Daughdrill, G. W. *J. Biomol. Struct. Dyn.* 2005, 23, 113.
- (322) Chen, S. C.; Chen, F. C.; Li, W. H. *Mol. Biol. Evol.* 2010, 27, 2548.
- (323) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* 2008, 37, 215.
- (324) Uversky, V. N. *Curr. Alzheimer Res.* 2008, 5, 260.
- (325) Ribeiro-Silva, C.; Gilberto, S.; Gomes, R. A.; Mateus, L.; Monteiro, E.; Barroso, E.; Coelho, A. V.; da Costa, G. A.; Freire, A. P.; Cordeiro, C. *Amyloid* 2011, 18, 191.
- (326) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* 2006, 75, 333.
- (327) Uversky, V. N.; Oldfield, C. J.; Midic, U.; Xie, H.; Xue, B.; Vucetic, S.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K. *BMC Genomics* 2009, 10 Suppl 1, S7.
- (328) Cheng, Y.; Legall, T.; Oldfield, C. J.; Mueller, J. P.; Van, Y. Y.; Romero, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *Trends Biotechnol.* 2006, 24, 435.
- (329) Vassilev, L. T. *Cell Cycle* 2004, 3, 419.
- (330) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. A. *Science* 2004, 303, 844.
- (331) Metallo, S. J. *Curr. Opin. Chem. Biol.* 2010, 14, 481.
- (332) Sung, Y. H.; Eliezer, D. *J. Mol. Biol.* 2007, 372, 689.
- (333) Uversky, V. N. *Cell. Mol. Life Sci.* 2003, 60, 1852.

Authors

Johnny Habchi. After obtaining the bachelor degree from the “Lebanese University of Beirut” in 2008, Johnny HABCHI achieved his Masters studies in “Structural Biology” in 2009 at the “Université de la Méditerranée” (Aix-Marseille II). He prepared his PhD within the “Structural disorder and Molecular Recognition” team at the laboratory Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille (FR) under the supervision of Dr. Sonia LONGHI. He got his PhD in biophysics from the Aix-Marseille University in 2012. Since October 2012 he is a post-doctoral fellow in Michele Vendruscolo’s group within the Department of Chemistry, University of Cambridge (UK). So far, he has authored 11 papers with most of them focusing on the characterization of the intrinsically disordered regions within the replicative complex of Henipaviruses and the induced folding that these regions undergo in the presence of their partners.



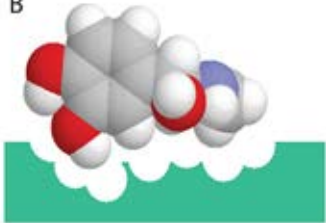
Peter Tompa. Peter (1959, Budapest) is married and has two daughters. He graduated in organic chemistry (1983, ELTE University, Budapest), and obtained his PhD in 1991 (ELTE University, Budapest) based on research carried out in the Institute of Enzymology, Hungarian Academy of Sciences (HAS), Budapest, on the interaction of glycolytic enzymes. He did his postdoc in the same institute under the supervision of Prof. Peter Friedrich on the structure-function characterization of the calcium-activated protease calpain (1992-2001), and became an independent PI in 2002. He started research on intrinsically disordered proteins (IDPs) in 2000, and played an active and decisive role in the rise of this field by suggesting a functional classification of IDPs, observing the role of structural disorder in chaperone function and developing basic concepts of the unusual modes of function of IDPs (moonlighting, fuzziness, disordered domains, supertertiary structure of proteins). He served on the Molecular Biology Evaluation Board of the Hungarian national grant agency OTKA (2005-2008), was chair of the scientific committee Biochemistry and Molecular Biology of HAS (2008-2010), and also deputy director of the Institute of Enzymology in the period 2007-2009. Currently he is the director of VIB (Flanders Institute of Biotechnology) Department of Structural Biology, Brussels (2011-), and also a professor at VUB (Free University Brussels, 2012-). He also holds the position of professor of biochemistry and protein sciences in the Institute of Enzymology, Budapest (2006-). He has research groups both in Budapest and Brussels. He was promoter of 10 doctorates, published 121 papers, 10 book chapters, and the first monograph of the field "Structure and function of intrinsically disordered proteins" (2009) by Taylor and Francis, Inc (CRC Press). He received about 5500 citations and his H-index is 32.

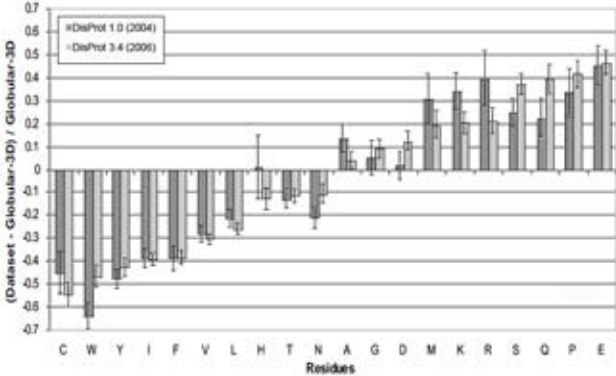


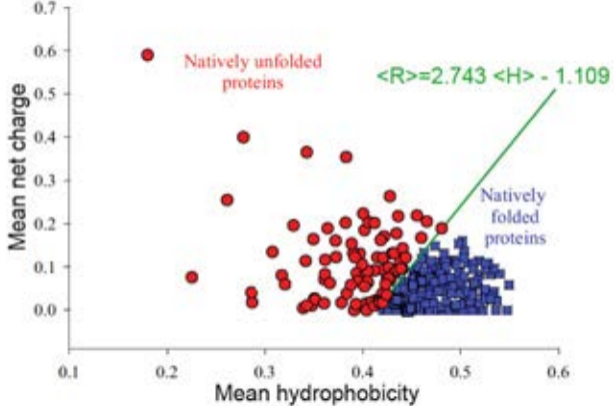
Sonia Longhi is Director of Research at the Center for the National Scientific Research (CNRS), heading the "Structural Disorder and Molecular Recognition" group within the laboratory Architecture et Fonction des Macromolécules Biologiques (AFMB). She obtained a PhD in molecular biology from the University of Milan in 1993. She got a HDR in structural virology from the University of Aix-Marseille I in 2003. Her scientific focus is on intrinsically disordered proteins (IDPs) and the mechanistic and functional aspects of the interactions they establish with partners. She has authored more than 80 scientific publications in international peer review journal, edited a book on measles virus nucleoprotein and co-edited with Prof. Vladimir Uversky a book entitled "Instrumental analysis of intrinsically disordered proteins" and one entitled "Flexible viruses - Structural disorder in viral proteins".

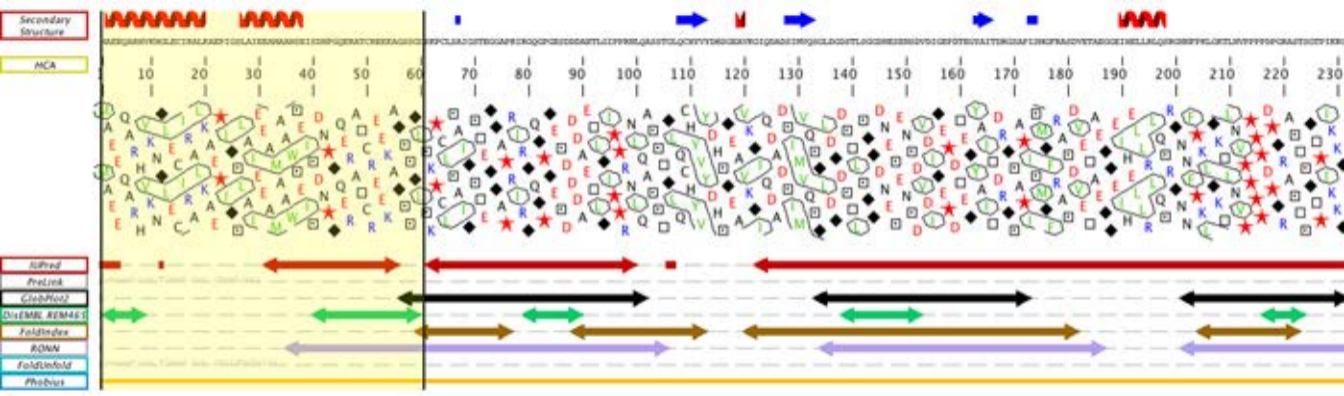


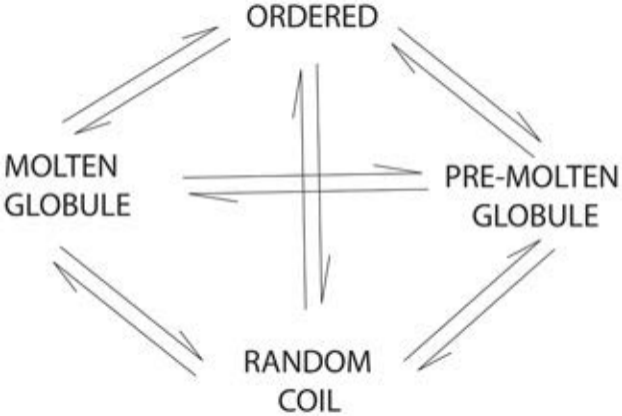
Dr. Uversky obtained his Ph.D. in biophysics from Moscow Institute of Physics and Technology (1991) and D.Sc. in biophysics from Institute of Experimental and Theoretical Biophysics, Russian Academy of Sciences (1998). He spent early career working on protein folding at Institute of Protein Research and the Institute for Biological Instrumentation (Russian Academy of Sciences). In 1998, he moved to the University of California Santa Cruz to work on protein folding, misfolding, and protein intrinsic disorder. In 2004, he moved to the Center for Computational Biology and Bioinformatics at the Indiana University Purdue University Indianapolis to work on the intrinsically disordered proteins. Since 2010, he is with the Department of Molecular Biology at the University of South Florida.

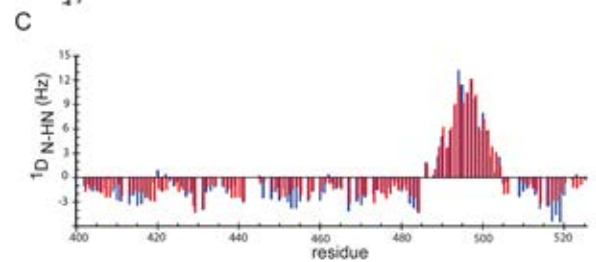
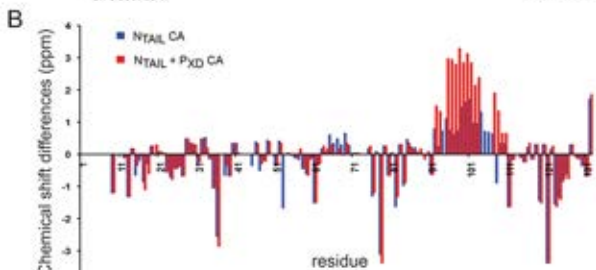
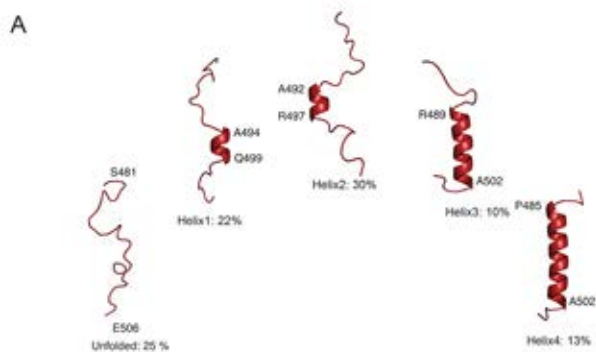
A**B**



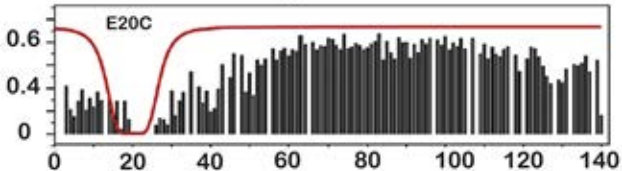


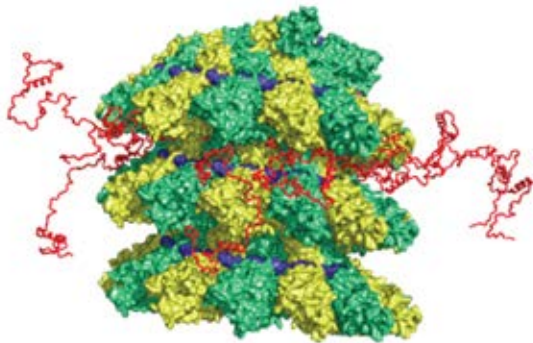
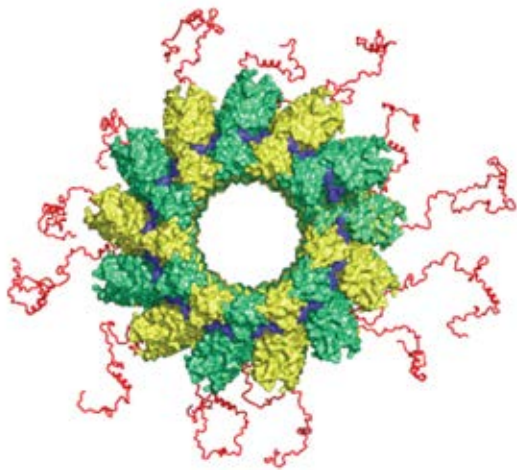


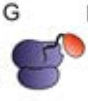
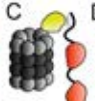




α -synuclein

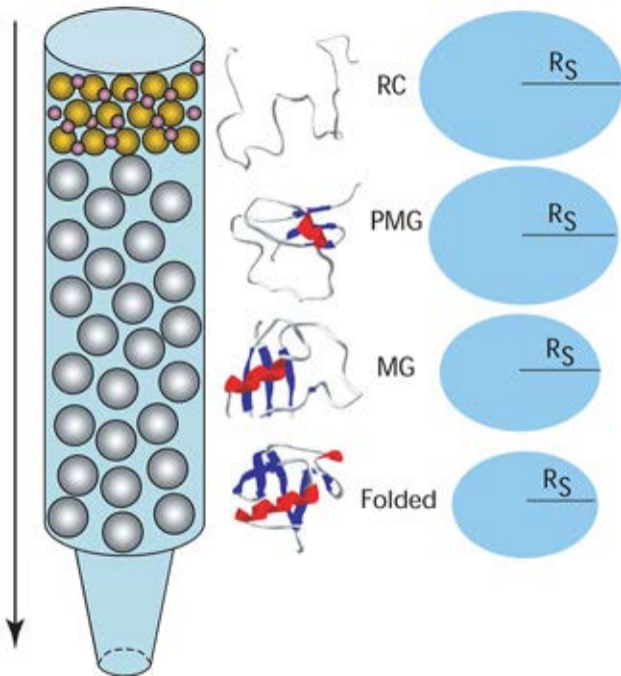


A**B**

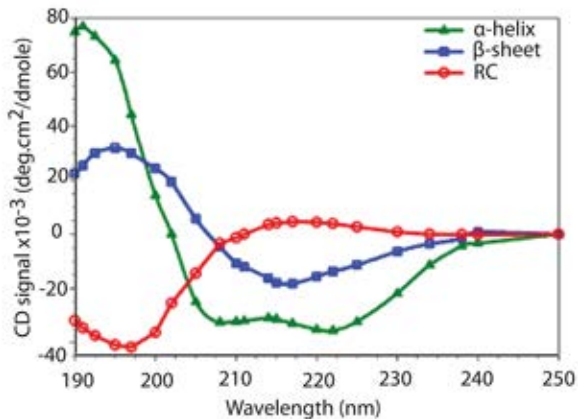




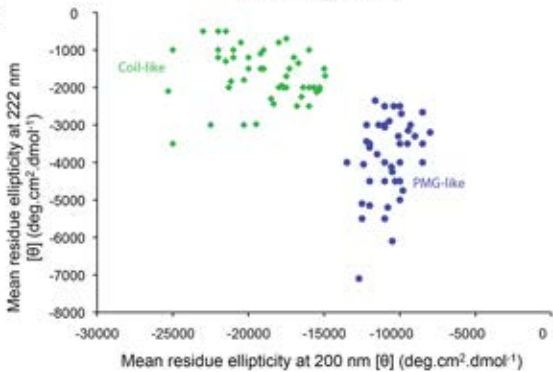
Disorder

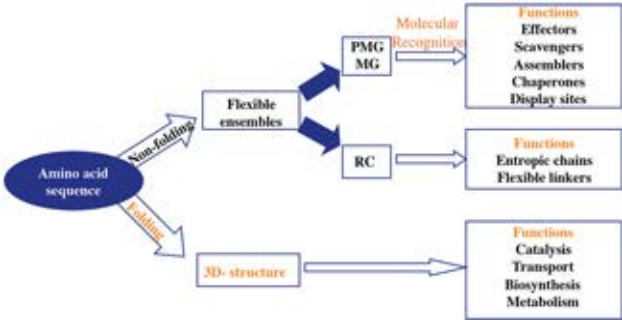


A

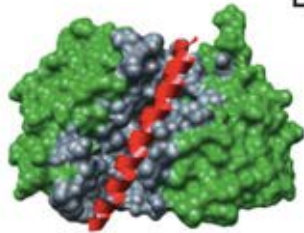


B

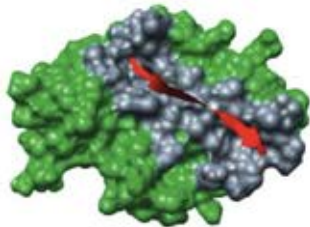




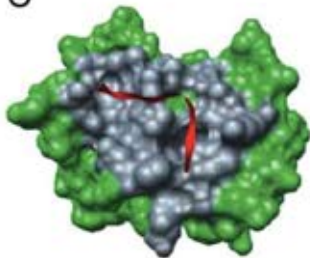
A



B

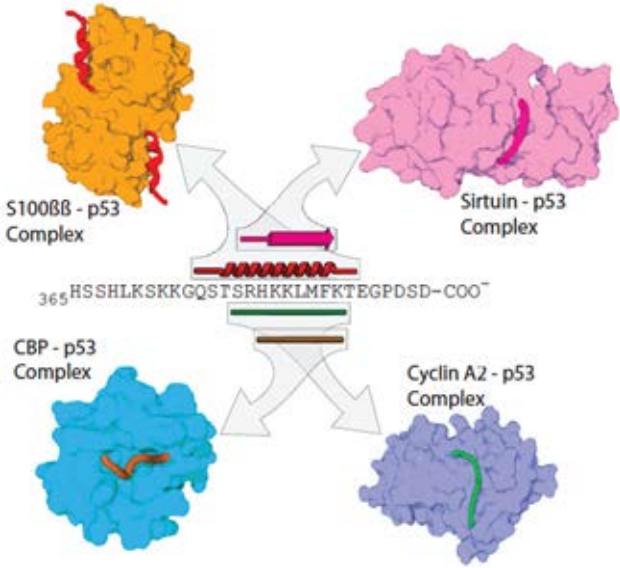


C

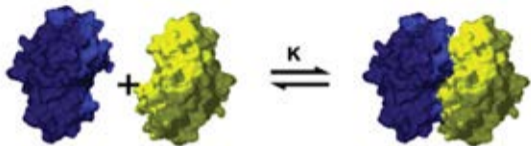


D

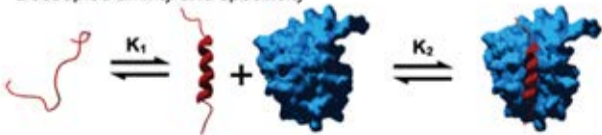




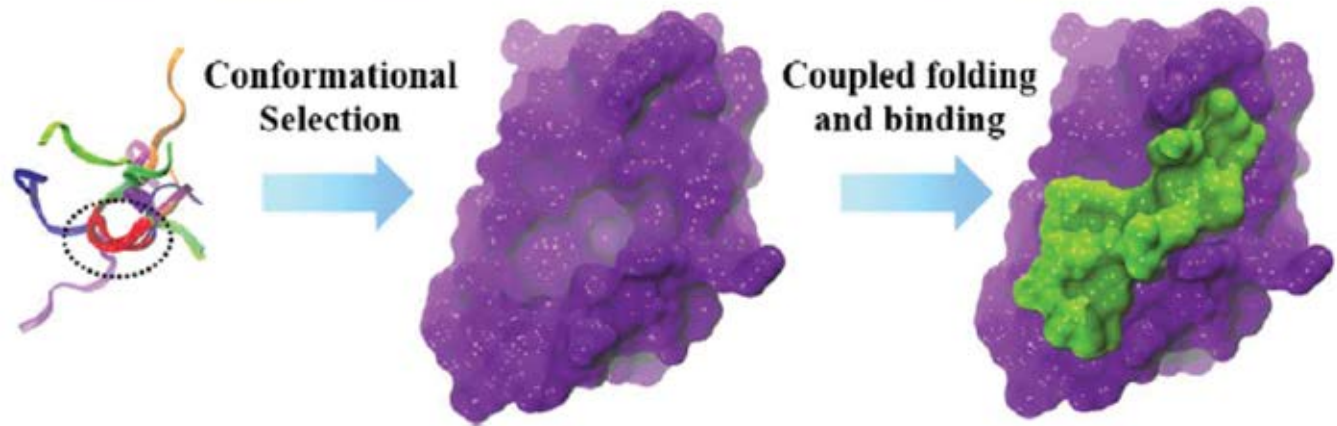
Coupled affinity and specificity



Decoupled affinity and specificity

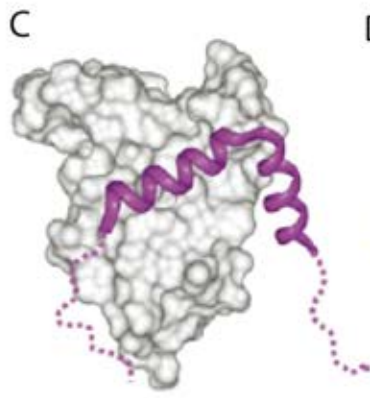
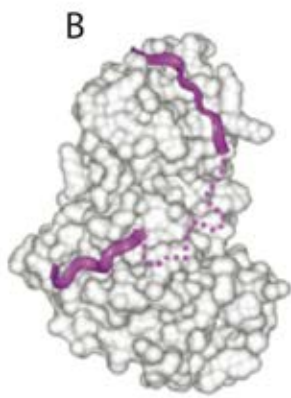
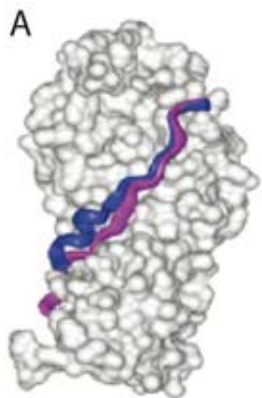


Synergistic model of IDP binding



Static

Dynamic



Disorder

