



HAL
open science

Introducing Protein Intrinsic Disorder

Johnny Habchi, Peter Tompa, Sonia Longhi, Vladimir N Uversky

► **To cite this version:**

Johnny Habchi, Peter Tompa, Sonia Longhi, Vladimir N Uversky. Introducing Protein Intrinsic Disorder. *Chemical Reviews*, 2014, 114 (13), pp.6561-6588. 10.1021/cr400514h . hal-04688543

HAL Id: hal-04688543

<https://hal.science/hal-04688543v1>

Submitted on 5 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Introducing Protein Intrinsic Disorder

Journal:	<i>Chemical Reviews</i>
Manuscript ID:	cr-2013-00514h.R1
Manuscript Type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Habchi, Johnny; CNRS and Aix-Marseille University, AFMB, UMR 7257 Tompa, Peter; VIB Department of Structural Biology, Longhi, Sonia; CNRS and Aix-Marseille University, AFMB, UMR 7257 Uversky, Vladimir; University of South Florida, Molecular Medicine

SCHOLARONE™
Manuscripts

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	3
2. Historical overview: Role of bioinformatics in establishing the IDP field.....	4
3. Computational means for IDPs analysis	6
3.1. Sequence peculiarities of IDPs	6
3.2. Computational tools for disorder prediction	7
4. Multiparametric approaches for assessing intrinsic disorder: Experimental view on conformational ensembles and induced folding transitions.....	12

1		2
2		
3	4.1. NMR	12
4		
5	4.2. Further quantitative techniques for characterizing structural disorder	17
6		
7		
8	4.3. Complementary methods for studying protein disorder	19
9		
10		
11	4.4. Analyzing disorder in a living cell.....	21
12		
13	4.4.1. In vivo existence of protein disorder	21
14		
15		
16	4.4.2. Survival of IDPs in the cell	23
17		
18		
19	5. Structural and conformational characteristics of IDPs.....	25
20		
21		
22	5.1. IDP structure from the polymeric physics viewpoint	25
23		
24		
25	5.2. Evaluating residual structure in IDPs	26
26		
27		
28	5.3. IDPs as conformational ensembles.....	28
29		
30		
31	5.4. Electrostatic and charge effects of compaction of the structural ensemble of IDPs...	30
32		
33	5.5. IDPs, their environment and functional misfolding.....	31
34		
35		
36	6. Intrinsic disorder and protein function.....	32
37		
38		
39	6.1. Functional diversity of IDPs through the prism of structural plasticity	33
40		
41		
42	6.2. Short recognition motifs in the interactions of IDPs. Interaction networks, one-to-	
43	many and many-to-one binding	36
44		
45		
46	6.3. Variability of binding modes attainable by disordered proteins: folding before or after	
47	binding	40
48		
49		
50	6.4. Dynamic or fuzzy complexes	44
51		
52		
53	6.5. IDPs and posttranslational modifications	45
54		
55		
56		
57	6.6. Alternative splicing in IDP function and dysfunction.....	47
58		
59		
60		

1		
2		
3	6.7. Intrinsic disorder and allosteric coupling in proteins	49
4		
5	6.8. Multi-valent interactions and phase-separation of IDPs	50
6		
7		
8	7. Cellular biology of IDPs. An overview	51
9		
10		
11	7.1. Abundance of intrinsic disorder in various proteomes	51
12		
13		
14	7.2. Involvement of structural disorder in biological processes	53
15		
16		
17	8. Distinctive features of IDP evolution	54
18		
19		
20	9. Intrinsic disorder and human diseases	56
21		
22	9.1. Correlation of disorder and diseases	56
23		
24		
25	9.2. IDPs as drug targets	57
26		
27		
28	Abbreviations	59
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

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

1
2
3 structure for functionality. The explanatory power of these 3-D structures continued to reinforce the
4
5 static view of protein structure that remained unquestioned. In contrast to this view, already in 1958,
6
7 Koshland suggested the “induced-fit” model based on the observations that some enzymes could act
8
9 on differently shaped substrates and hence a degree of flexibility is inevitable in function.²
10

11
12
13 Furthermore, questions such as “what does the missing electron density in most of the deposited
14
15 structures in the PDB correspond to?”, “why some proteins are highly sensitive *in vitro* to
16
17 proteolysis?”, and “why some proteins possess a particular behavior during the purification
18
19 process?” further opened eyes towards flexibility and focused attention on proteins basically
20
21 distinct from well-known globular proteins. A combined answer to these questions ruled out flexible
22
23 proteins as artifacts and enlightened the “dark side” of Structural Biology, that of disordered proteins
24
25 (i.e. proteins that lack 3-D structure). This review provides a detailed description of those proteins
26
27 that broke the protein rules by their inherent flexibility and peculiar features, now generally termed
28
29 “intrinsically disordered” proteins (IDPs) or regions (IDRs). This review is intended as a general
30
31 introduction to the series of thematic reviews in this special issue of Chemical Reviews on IDPs. Our
32
33 key message is that IDPs possess no well-defined 3-D structure but rather adopt an ensemble of
34
35 conformations in solution, yet they are functional. A literature survey is provided by summarizing the
36
37 main aspects of IDPs that led to an exponential increase of interest in these proteins. Moreover,
38
39 through the different parts of this review, biochemical and biophysical approaches that are
40
41 frequently used to assess intrinsic disorder are detailed.
42
43
44
45
46
47

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

49
50
51
52

53 It is now half a century since the first crystal structure of a protein (e.g. the atomic structure of
54
55 myoglobin determined in 1958 by John Kendrew³) was published leading to the “birth” of Structural
56
57 Biology. Soon after, the field grew rapidly with a dozen of atomic structures of proteins being
58
59
60

1
2
3 published by the early 1970's, leading to the establishment of the predecessor of PDB, the
4
5 Brookhaven Data Bank.^{4,5} A well-folded, albeit dynamic, structure was thought to be the hallmark of
6
7 protein function. This view was also built on the success of the previous half a century of the lock-
8
9 and-key view explaining the specificity of enzymes (Fig. 1A) and complementarity of antibody to
10
11 antigen structure. The classical structure-function paradigm is centered on the idea that protein
12
13 function depends on a well-defined three-dimensional (3-D) structure.
14
15

16
17
18 Indeed, enzymes have a rather well-defined binding pocket for the formation of an enzyme-substrate
19
20 (ES) complex (Fig. 1A), which ensures a tight fit between the binding pocket and the substrate (Fig.
21
22 1B). The fact that conditions causing denaturation of proteins (treatment by acid, alkali or urea) lead
23
24 to the loss of enzyme activity lent strong support to the protein structure-function paradigm.
25
26 Consequently, the notion that many proteins or regions of proteins could be not ordered, but
27
28 intrinsically disordered was unacceptable for a long time. Indeed, computational studies carried out
29
30 as early as in 2000 predict that ordered proteins and domains cover only about half of the sequence
31
32 space in various proteomes,⁶ which has led to "breaking the protein rules".⁷
33
34
35

36
37 The re-adaptation of the structure-function paradigm started as early as 1998 when Romero et al.
38
39 predicted that more than 15,000 proteins in the Swiss Protein (SwissProt) database contain
40
41 intrinsically disordered regions (IDRs) of at least 40 consecutive amino acids.⁸ Soon after, Wright and
42
43 Dyson unveiled in 1999 that a large proportion of gene sequences are likely to be unfolded in
44
45 solution, under native, functional conditions.⁹
46
47
48

49
50 Through the last years, the terminology seem to have settled on "intrinsically disordered",¹⁰ for
51
52 proteins/regions that exist as dynamic ensembles, within which atom positions and backbone
53
54 Ramachandran angles exhibit extreme temporal fluctuations without specific equilibrium values.^{8,11-16}
55
56
57
58
59
60

1
2
3 Their detailed characterization suggests that our traditional views of protein structure and function
4
5 need to be re-assessed.
6
7

8 9 **3. Computational means for IDPs analysis**

10 11 12 13 14 15 **3.1. Sequence peculiarities of IDPs**

16
17
18
19 The frequency of amino acids in disordered proteins significantly differs from that of ordered
20 proteins.^{11,12,16} Amino acids frequencies plotted as a function of the flexibility index of residues show
21 a distinctive pattern¹⁷ (Fig. 2): IDPs are often enriched in charged and structure-breaking residues
22 (Pro and Gly) and in Ala. This collection of amino acids has been called “disorder promoting amino
23 acids”. On the other hand, these proteins are depleted in hydrophobic and aromatic residues and
24 also possess fewer Cys and Asn residues, collectively termed “order promoting amino acids”.^{18,19} It is
25 generally thought that the absence of structure is encoded by the amino acid sequence. The ensuing
26 low mean hydrophobicity and a high net charge represents an important prerequisite for the
27 absence of a compact structure in proteins under physiological conditions,¹⁵ because it provides less
28 driving force for protein compaction while also contributing to charge-charge repulsion. This simple
29 principle can be applied to predict IDPs by plotting the absolute net charge as a function of the mean
30 normalized hydrophobicity, a plot denoted as charge-hydrophobicity (CH) plot or Uversky plot (Fig. 3).¹⁵
31 In this plot, IDPs cluster in the high net charge - low net hydrophobicity half of the plane (Fig. 3), and
32 the distance from the separating line (the formula of the linear function is: $\langle R \rangle = 2.743 \langle H \rangle - 1.109$)
33 may carry information on the extent and type of disorder for the whole chain (Fig. 3).^{20,21}
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52
53
54 Protein disorder is also related to low sequence complexity (i.e. repetitiveness and limited diversity
55 in amino acid composition).^{22,23} Accordingly, protein disorder abounds in certain types of charged
56
57
58
59
60

1
2
3 and Pro-rich repetitive regions.²⁴ 25% of all amino acids in the SwissProt database are found in such
4
5 regions, and 34% of all proteins have at least one such segment.¹¹ It is to be noted though that some
6
7 special cases of low-complexity sequences are found in proteins with a certain amino acid
8
9 periodicity, yet ordered such as coiled-coils (collagen, for example). This biased composition together
10
11 with these peculiar sequence features enabled the development of reliable computational tools for
12
13 the analysis of structural disorder.
14
15

19 **3.2. Computational tools for disorder prediction**

20
21
22
23 The recognition of disordered regions from sequence facilitates the functional annotation of proteins
24
25 and is instrumental in delineating protein domains amenable to crystallization.^{25,26} To this end, many
26
27 disorder predictors have been developed.²⁵⁻²⁷ (cf. also the table in the Wikipedia article found at
28
29 http://en.wikipedia.org/wiki/Intrinsically_unstructured_proteins). The reliability of disorder
30
31 prediction benefits from the use of several methods based on different concepts, different
32
33 physicochemical parameters, or different implementations.²⁵⁻²⁹ In principle, this problem can be
34
35 approached from three distinct directions:
36
37

38
39
40 1) Simple amino acid propensities reflecting some basic physical or sequence features. Such
41
42 propensity-based predictors rely on simple statistics of amino acid propensity, the physical/chemical
43
44 features of amino acids or on a preliminary concept on the physical background of disorder. The CH-
45
46 plot¹⁵ with its derivative FoldIndex,³⁰ PreLink,³¹ and GlobPlot,³² to name few, fit into this category.
47
48

49
50
51 2) Machine-learning algorithms (MLAs) are trained on datasets of disordered regions. Presently three
52
53 databases of experimentally characterized disordered proteins are available: the Database of Protein
54
55 Disorder, DisProt (<http://www.disprot.org/>),³³ IDEAL ([http://www.ideal.force.cs.is.nagoya-](http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/)
56
57 [u.ac.jp/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
58
59
60

1
2
3 publicly available database, with about 1500 different regions. PONDR,²³ DisEMBL,³⁶ DISOPRED2,³⁷
4
5 and RONN,³⁸ to name few, fall in this category of predictors since they have been trained on disorder
6
7 data coming from either short regions of missing density from the PDB or from nuclear magnetic
8
9 resonance (NMR) or from the DisProt database.³³ Regions of missing electron density in the PDB are
10
11 generally short, as long regions prevent crystallization.²⁵ As such, short disorder is over-represented
12
13 in the database of disordered regions, and hence these predictors tend to perform better in
14
15 predicting short disorder than long disorder.
16
17

18
19
20 3) The tendency of amino acids to make or avoid contacts with each other. The underlying idea is
21
22 that IDPs are unfolded because they cannot make sufficient inter-residue contacts to overcome the
23
24 large decrease in configurational entropy during folding. A simple statistical analysis of residue
25
26 contact numbers is performed by FoldUnfold³⁹ that calculates the expected average number of
27
28 contacts per residue from the amino acid sequence alone. IUPred^{40,41} evaluates the energy resulting
29
30 from inter-residues interactions, based on their inter-residues contacts in globular proteins.^{25,42} Since
31
32 these predictors are not trained on data sets of disordered regions they avoid the shortcomings and
33
34 biases associated with these disordered datasets. Therefore, they are expected to perform better
35
36 than the former methods on disordered proteins presently under-represented in training datasets
37
38 (i.e., fully or mostly disordered proteins).
39
40
41
42

43
44 Another non-automated method that is very useful for unveiling unstructured regions is hydrophobic
45
46 clusters analysis (HCA).⁴³ HCA provides a two-dimensional helical representation of protein
47
48 sequences in which hydrophobic clusters are plotted along the sequence.⁴³ As such, HCA is not *stricto*
49
50 *sensu* a predictor, nevertheless, disordered regions are recognizable as they are depleted (or even
51
52 devoid) in hydrophobic clusters (an HCA plot is presented in Fig. 4). The power of HCA analysis relies
53
54 also on its ability to highlight coiled-coils, regions with a biased composition and regions with
55
56
57
58
59
60

1
2
3 potential for induced folding (i.e. a disorder-to-order transition upon binding to a partner or ligand)
4
5 (see also section 6.3).
6
7

8
9 Predictors can also be classified based on the binary nature of the prediction. Examples of binary
10 predictors are the CH-plot and the cumulative distribution function (CDF) analysis.^{20,44} The CDF
11 analysis summarizes the per-residue predictions by plotting predicted disorder scores against their
12 cumulative frequency, which allows ordered and disordered proteins to be distinguished based on
13 the distribution of prediction scores.^{20,44}
14
15
16
17
18
19

20
21
22 The CH-CDF plot (see Fig. 5) is an analytical tool combining the outputs of two binary predictors, the
23 CH-plot and the CDF plot.^{45,46} The vertical distance on the CH-plot from the location of the protein to
24 the boundary line is a scale of disorder (or structure) tendency of the protein (Fig. 5A). This distance
25 is referred to as the CH-distance. In CDF-plots, curves of ordered proteins tend to stay on the upper
26 left half, whereas curves of disordered proteins tend to locate at the lower right half of the plot. An
27 approximately diagonal boundary line separating the two groups can be identified and the average
28 distance of the CDF curves from this boundary (CDF-distance) is a measure of the disorder (order)
29 status of a given protein (Fig. 5B). The CH-CDF plot is based on both the CH-distance and the CDF-
30 distance (Fig. 5C).^{45,46} It provides very useful information on the general disorder status of a given
31 protein. After setting up boundaries at CH=0 and CDF=0, the entire CH-CDF plot can be split into four
32 quadrants. Starting from the upper right quadrant, by taking the clockwise sequence, the four
33 quadrants are named Q1 (upper right), Q2 (lower right), Q3 (lower left), and Q4 (upper left). Proteins
34 in Q1 are structured by CDF, but disordered by CH; proteins in Q2 are predicted to be structured by
35 both CDF and CH; proteins in Q3 are disordered by CDF but structured by CH; and proteins in Q4 are
36 predicted to be disordered by both methods. The location of a given protein in this CH-CDF plot gives
37 information about its overall physical and structural characteristics. Fig. 5D shows the application of
38 this tool for the analysis of ribosomal proteins.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 In general, low sequence complexity, low predicted secondary structure content and high sequence
6 variability constitute an indicator of flexibility of proteins or protein domains.²⁶ It should be stressed
7 that it is difficult and maybe impractical to establish the “best” predictor at the moment. Some
8 predictors perform better on short disordered regions (i.e. DISOPRED2 and PreLink) while other
9 predictors (IUPred for instance) perform well in predicting long disordered segments and finally
10 some predictors, such as PONDR, GlobPlot and FoldIndex, have been trained on both short and long
11 disorder and provide a balanced performance.²⁵ Therefore, to avoid pitfalls, different predictors
12 should be combined, as performed by metapredictors that seek a consensus of the scores of
13 different predictors relying on different principles (PONDR-FIT for instance).⁴⁷ Alternatively, one
14 could simply use publicly available servers allowing multiple predictions, such as “MeDor”
15 (MEtserver of DisORder), a server that does not provide a consensus prediction but is intended to
16 speed up the analysis of protein disorder thanks to the simultaneous submission and retrieval of
17 various disorder predictions, the prediction of secondary structures and a HCA plot (see Fig. 4).⁴⁸ (For
18 a general procedure to predict structural disorder, see ref.²⁵; for a detailed description of
19 metapredictors, see ref.²⁷)
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 An extreme extension of the combined use of different predictors is the combination of *in silico* and
41 experimental approaches with the ultimate goal of inferring as many structural information as
42 possible while limiting the experimental characterization to relatively low-demanding experiments.
43 An illustration of such an approach can be found in ref.⁴⁹ where a spectroscopic and computational
44 analysis were combined. In that study, the authors plotted the ratio between the Θ_{222} and Θ_{200}
45 ($\Theta_{222}/\Theta_{200}$) of a set of IDPs under study, along with the $\Theta_{222}/\Theta_{200}$ ratio of a set of well-characterized
46 random coil-like and premolten globule-like proteins.¹³ The authors then set an arbitrary threshold of
47 the $\Theta_{222}/\Theta_{200}$ ratio that allows discrimination between random coil-like IDPs and IDPs adopting a
48 premolten-like conformation. Then, they generated a plot in which the distance of each IDP under
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 study from this threshold was plotted as a function of its CH-distance in the CH plot. This analysis was
4
5 intended to combine, and hence extend, two previous methods^{13,15} so as to allow random coil-like
6
7 forms to be readily and easily distinguished from premolten globule-like forms among proteins
8
9 predicted to be intrinsically disordered by the hydrophathy/charge method. In the resulting plot,
10
11 increasingly negative CH distances designate proteins with increasing disorder, while increasingly
12
13 positive $\Theta_{222}/\Theta_{200}$ distances designate IDPs becoming progressively more collapsed, as a
14
15 consequence of an increased content in regular secondary structure. Thus, the left bottom quadrant
16
17 is expected to correspond to IDPs adopting a random coil-like conformation, while the right bottom
18
19 quadrant is supposed to designate IDPs adopting a premolten globule-like conformation.
20
21
22
23

24
25 Although the number of known disordered sequences based on the analysis of amino acid sequence
26
27 parts missing in the crystal structure of proteins and analysis of properties of already characterized
28
29 IDPs and their sequences is not too large, it is not too small either. In fact, the available data are
30
31 sufficient to create reliable computational tools (i.e., the existing data are enough to generate both
32
33 training and testing sets needed for the development of predictors). These tools are able to rather
34
35 accurately find sequence features similar to those used in training. The progressively increasing
36
37 number of disorder predictors is an obvious reflection of these facts. The situation is further
38
39 complicated by the fact that the structural spectrum of intrinsic disorder is very broad (in fact, it is
40
41 much broader than that of ordered proteins). This means that existing computational tools and their
42
43 various combinations are doomed to be sensitive to some specific (and rather limited) features and
44
45 would not be able to cover the whole spectrum of disorder. We consider the realm of computational
46
47 tools for disorder prediction to be similar to the multiparametric experimental approach for
48
49 structural characterization of proteins. In that approach, although each technique is able to see only
50
51 a part of picture, the whole picture still can be restored by looking to the structure at different angles
52
53 (i.e., using different techniques sensitive to the different levels of protein structure). Obviously, it
54
55 would be more appropriate to say that the generalized criteria for disorder are still evolving and that
56
57
58
59
60

1
2
3 only the availability of a reasonably large number of experimentally characterized IDPs will provide a
4
5 solid basis for justified protein disorder prediction. On the other hand, this does not mean that these
6
7 predictors should not be developed before such large number of experimentally characterized IDPs
8
9 will be collected. On the contrary, the existing experimentally validated criteria of intrinsic disorder
10
11 can be and have to be used for the creation of corresponding computational tools. These algorithms
12
13 represent a very important addition to the researcher's arsenal and open possibility for large scale
14
15 analysis. One should remember that the interest to protein intrinsic disorder and appreciation of its
16
17 importance for protein function was generated via the recognition that IDPs are not rare exceptions,
18
19 but a new rule. Furthermore, this recognition was based on the bioinformatics and computational
20
21 analyses, where various predictors of protein intrinsic disorder were used.
22
23
24
25

26 **4. Multiparametric approaches for assessing intrinsic disorder: Experimental** 27 28 29 **view on conformational ensembles and induced folding transitions** 30 31 32

33
34 Structural disorder can be detected and analyzed by many (bio)physical techniques, some indirect,
35
36 others more quantitative in providing structural data. Among them, one of the most powerful in
37
38 providing quantitative information is NMR.
39
40
41
42

43 **4.1. NMR** 44 45 46 47

48 NMR is widely recognized as the predominant, quantitative technique in studies of IDPs. This section
49
50 will provide a description of the most broadly used NMR experiments and of the most recent NMR
51
52 advancements that have enabled a transition from a qualitative to a quantitative description of
53
54 IDPs.⁵⁰
55
56
57
58
59
60

1
2
3 A global characterization of IDPs can be inferred by either simply recording a one-dimensional
4 spectrum (1-D) and deriving the information from the low spread of the proton resonance
5 frequencies due to the presence of all the residues in the same environment, or by performing a
6 wide-line NMR relaxation experiment that provides information about the presence of a large
7 hydrate layer in agreement with a disordered, open and extended state. Additionally, using pulsed-
8 field gradient NMR, one can also measure the diffusion coefficient of proteins from which
9 hydrodynamic parameters can be determined.⁴² Although these latter are less direct indicators of
10 disorder as compared to chemical shifts, they provide useful hints because IDPs possess peculiar
11 hydrodynamic parameters as compared to structure proteins. Heteronuclear single quantum
12 coherence (HSQC) is one of the most frequent experiments in the IDP literature. It represents the
13 starting point of resonance assignment (i.e. the process of identifying which resonance belongs to
14 which residue of the protein), which is essential for a meaningful interpretation of more advanced
15 NMR experiments. Indeed, high-field NMR spectrometers together with uniformly and specifically
16 labeled proteins have overcome the problem of proton signal overlap of IDPs. Resonance
17 assignments can be made using the dispersion of ¹⁵N and ¹³C nuclei, which are more sensitive to local
18 amino acid sequence.⁵¹ Recently, methods exclusively based on ¹³C detection have been reported,
19 which afford an independent strategy to simplify crowded spectra as well as to perform sequence-
20 specific assignment.⁵² HSQC experiments are also useful for detecting interactions with unlabeled
21 binding-proteins, because a change in the chemical environment and in relaxation due to the
22 interaction make NMR parameters of residues directly involved shift or even disappear from the
23 spectrum.⁵³ Once a resonance assignment has been achieved, a variety of NMR parameters can be
24 determined to characterize the structural and dynamic behavior at the residue level, thus obtaining
25 sequence-specific structural information (Fig. 6).
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52
53
54 In order to determine local secondary structural preferences, many methods have been described in
55 the literature such as chemical shifts (which is the most powerful tool for achieving this goal),
56
57
58
59
60

1
2
3 coupling constants (which are directly related to bond torsion angles) and short range nuclear
4
5 Overhauser effect (NOEs) (which can be used to corroborate secondary structure propensities
6
7 detected through chemical shifts).⁵⁴
8
9

10
11 The primary observable in NMR studies of unfolded and partly folded proteins is the chemical shifts
12
13 that are sensitive to polypeptide backbone torsion angles. Deviation of chemical shifts from random
14
15 coil values (termed secondary chemical shift, SCS, or chemical shift index, CSI) for $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and
16
17 ^{13}CO are sensitive to local secondary structure (i.e. population and location) and hence provide
18
19 important insights into the structures populated in the conformational ensemble in IDPs.^{51,55,56} One
20
21 potential problem associated with this kind of approach concerns incorrect frequency referencing,
22
23 which can result in systematic errors in the secondary shifts. In order to address this problem, the $\text{C}\alpha$
24
25 and $\text{C}\beta$ chemical shifts (that shift in opposing directions for α -helical segments) can be used
26
27 simultaneously to estimate the level of secondary structure in disordered proteins.⁵⁷ In Fig. 6B, an
28
29 example taken from ⁵⁸ is provided showing the $^{13}\text{C}\alpha$ chemical shifts of the intrinsically disordered C-
30
31 terminal domain of the measles virus nucleoprotein (N_{TAIL}) both in its free and bound (i.e. in complex
32
33 with the C-terminal domain of the phosphoprotein, XD) states. This enabled to reveal that the
34
35 protein has a propensity to form an α -helix even in the free state.
36
37
38
39

40
41 More detailed information about local conformational sampling can be derived from inter proton
42
43 NOEs. However, quantitative interpretation of NOEs is complicated by the strong sensitivity of the
44
45 interaction on the range of dynamic timescales commonly encountered in unfolded proteins.⁵⁹ In
46
47 addition, long range NOEs characteristic of tertiary interactions are usually not observed in IDPs due
48
49 to fluctuations and heterogeneity of the structure.⁴² Another interesting parameter that one could
50
51 infer from HSQC-based experiments is relaxation rates (such as ^{15}N relaxation experiments) that
52
53 reflect backbone and side-chains dynamics and could provide information about local disorder.
54
55 Analysis of the relaxation data is particularly informative on local structural preferences, such as
56
57
58
59
60

1
2
3 hydrophobic clusters, secondary structure elements and transient long-range contacts.⁵⁴ In fact,
4
5 variations in backbone mobility can be correlated with intrinsic properties of the amino acids in the
6
7 sequence. Clusters of small amino acids such as glycine and alanine show increased backbone
8
9 mobility, compared to the average. In contrast, local hydrophobic interactions that persist in urea
10
11 cause some restrictions of backbone motions, hence supporting the fact that a RC never exists, not
12
13 even under the harshest denaturing conditions (i.e., in solutions containing 8M urea or 6 M
14
15 guanidinium chloride).⁶⁰
16
17
18
19

20 The above-listed traditional NMR parameters only provide information about the local dynamics and
21
22 structural state. This situation has changed by the advent of methods based on residual dipolar
23
24 couplings (RDCs) and paramagnetic relaxation enhancements (PREs), which yield a high number of
25
26 well-defined quantitative parameters reporting on long-range interactions (PRE) and global structure
27
28 (PRE, RDC).^{54,59,61,62}
29
30
31
32

33 PRE has been most successful in detecting long-range contacts in disordered protein ensembles. The
34
35 method relies on the introduction of a paramagnetic nitroxide spin label at a specific position within
36
37 the amino acid sequence.⁵⁴ HSQC spectra are then recorded with the spin-label in the paramagnetic
38
39 (oxidized) and diamagnetic (reduced) states, and differences in one of several parameters, such as
40
41 line-width, relaxation rate, or intensity are determined. Limitations of this approach include the need
42
43 of a number of cysteine-carrying mutants of the protein, and possible interference of non-native
44
45 side-chain and/or the radical with native long-range contacts. Nevertheless, these measurements are
46
47 extremely powerful, because they provide unambiguous evidence of the presence of fluctuating
48
49 tertiary structure that can be very difficult to identify by other techniques.⁵⁹ An example taken from
50
51 ref.⁵⁴ illustrating how a long-range distance could be inferred from PRE measurements is provided in
52
53 Fig. 7. Moreover, many examples in the literature have used PRE distance information to constrain
54
55
56
57
58
59
60

1
2
3 MD simulations and finally arrive at a reasonably confined distribution of structural states in the
4
5 ensemble of IDPs.^{61,63}
6
7

8
9 As compared to chemical shifts and NOEs, RDCs present a powerful tool that provides a far more
10
11 extensive and quantitative description of local, and possibly, global order in the unfolded state.⁵⁵
12
13 Indeed, RDCs measured between pairs of nuclei in partially aligned proteins, are very sensitive
14
15 probes of time and ensemble-averaged conformational equilibria exchanging on timescales up to the
16
17 millisecond and can therefore be used to characterize both the structure and the dynamics of
18
19 unfolded proteins.⁵⁹ An example taken from ref.⁶² showing how RDCs have been used to probe the
20
21 level and nature of residual structure in the C-terminal region of the MeV nucleoprotein (N_{TAIL}) is
22
23 provided in Fig. 8A.
24
25
26
27

28
29 The unique power of NMR in characterizing IDPs is that the combination of different approaches
30
31 enables to describe the real structure of IDPs; i.e., their structural ensemble, either alone or in the
32
33 bound form.⁶⁴ Indeed, recently the disordered C-terminal domain of the measles virus nucleoprotein
34
35 has been studied by a combination of NMR spectroscopy, SANS (small angle neutron scattering) and
36
37 electron microscopy (EM), providing an *in situ* ensemble model describing the conformational
38
39 behavior of the disordered region at the surface of the nucleocapsid (see Fig. 8)⁶² (For further
40
41 description, see section 5.3)
42
43
44
45

46
47 In all, NMR provides quantitative residue-level information on structure and dynamics of IDPs, which
48
49 enables to develop structural ensembles, the current best representation of the structure of IDPs. It
50
51 should not be forgotten, however, that NMR has some inherent limitations, due to which it is best
52
53 complemented by other quantitative biophysical techniques. Examples of such limitations, are
54
55 averaging of fast exchange events, low signal/noise ratio, difficulty in defining ensemble shape and
56
57 poor ability to provide information on global crossing between meta-stable sub-states, not to
58
59
60

1
2
3 mention the fact that NMR is highly demanding in terms of protein amounts, that isotopic labeling is
4
5 rather expensive and that some proteins are very poorly expressed in minimal media.
6
7
8
9

10 11 12 **4.2. Further quantitative techniques for characterizing structural disorder** 13 14 15

16
17 The limitations of NMR call upon complementing and/or combining its results with some other
18
19 biophysical techniques which provide additional insights into the structural details of IDPs, and also
20
21 enable approximations not yet amenable for NMR. Among these, we can cite for instance site-
22
23 directed spin-labeling EPR spectroscopy,⁶⁵⁻⁶⁷ an approach powerful in documenting and mapping
24
25 structural transitions in IDPs, which is currently being expanded thanks to the introduction of a new
26
27 generation of spin labels targeting residues other than cysteine.^{68,69} Concomitantly to experimental
28
29 techniques, *in silico* approaches have also been developed that enable physical quantification of
30
31 folding and binding energy landscapes of IDPs at atomistic level. Among them, we cite those
32
33 established by the group of Jin Wang who developed an atomic hybrid model for IDPs by integrating
34
35 local physics-based interactions and non-local structure-based interactions.⁷⁰⁻⁷² In their recent
36
37 approach, they have quantitatively uncovered the intrinsic energy landscape (not the free energy
38
39 landscape often discussed in the field) of folding and binding. To this end, they have explored the
40
41 underlying density of states, which is a statistical distribution in microcanonical ensemble extracted
42
43 from the conventional canonical ensemble binding–folding dynamics using a structure-based model
44
45 both with and without considering energetic roughness.⁷¹ To visualize the landscape, they chose the
46
47 energy and the structural similarity of generated conformers to native structure defined as the
48
49 fraction of native contacts as reaction coordinates. The authors showed that the energy landscape of
50
51 folding and binding is funneled towards the native state. The energy landscape topography
52
53 representing the degree of the funnelness is quantified by the dimensionless ratio of the slope and
54
55 the roughness modularized by the entropy (size) of the funnel. The authors showed that this
56
57
58
59
60

1
2
3 landscape topography ratio determines the thermodynamics, kinetics, and the association
4
5 mechanism of the binding–folding dynamics. Using this newly developed IDP model in combination
6
7 with physics-based simulations (with solvent), Wang *et al.* recently described the unbound
8
9 ensembles of the α -MoRE (residues 484-504) located within the intrinsically disordered C-terminal
10
11 domain of the measles virus (MeV) nucleoprotein (N_{TAIL}) and deciphered the mechanism by which the
12
13 α -MoRE recognizes the X domain (XD) of the MeV phosphoprotein.⁷³
14
15

16
17
18 Among experimental techniques providing quantitative information, SAXS and single-molecule
19
20 approaches, which allow the observation of transient intermediates, have enabled a better
21
22 understanding of the behavior of IDPs. In particular, SAXS studies, combined with ensemble
23
24 optimization methods (EOM), have led to low-resolution description of IDPs as realistic
25
26 conformational ensembles (see ref.⁷⁴ and references therein cited).
27
28

29
30
31 In-bulk methodologies to studying IDPs, however, have two main intrinsic limitations. The first
32
33 limitation is connected to the limited capability to single out the monomeric state of the investigated
34
35 IDPs. In the experimental conditions, the monomeric state of the protein under study might easily
36
37 co-exist along with soluble oligomers that can form quickly in solution at the concentrations required
38
39 by the sample preparations. Single-molecule methods operating at low concentrations or with
40
41 immobilized molecules allow the properties of the monomeric species to be monitored without any
42
43 interference by aggregation already started. The second limitation of the traditional in-bulk
44
45 methodologies is connected to their time and ensemble averaging. They can observe only average
46
47 properties of 10^{14} – 10^{17} molecules at a time. Even in the case in which no oligomers are present in the
48
49 sample, the intrinsic ensemble averaging of in-bulk methodologies might prevent to single out poorly
50
51 populated states. Single-molecule (SM) approaches, such as high-speed atomic force microscopy (HS-
52
53 AFM),^{75,76} AFM-based force spectroscopy (FS),⁷⁷ and fluorescence resonance energy transfer
54
55 (FRET),⁷⁸⁻⁸⁰ therefore represent a powerful tool to investigate the structure and the dynamics of IDPs.
56
57
58
59
60

1
2
3 While SM-FRET reports changes in the mean distance between two residues (steady-state FRET) and
4 intramolecular distance distributions (time-resolved FRET) enabling to study conformational
5 equilibria in timescales shorter than a few milliseconds, AFM-based SM-FS is particularly sensitive to
6 the formation of secondary structures, and probes timescales from milliseconds to seconds, much
7 longer than those of SM-FRET.
8
9
10
11
12

13
14
15 HS-AFM allows direct observation of structural dynamics and dynamic processes of biological
16 molecules, with a sub-second to sub-100 ms temporal resolution.^{75,81} Various dynamic protein
17 processes have been successfully visualized by this approach. HS-AFM is not only applicable to well-
18 structured proteins but also to IDPs. The group of Toshio Ando was the first to use HS-AFM to directly
19 visualize IDPs attached on a mica surface.⁷⁶
20
21
22
23
24
25
26
27

28
29 Altogether, conformational and spectroscopic studies carried out thus far have unveiled that IDPs do
30 not represent a homogeneous structural class, and range from fully extended (random-coil like) to
31 compact (molten-globule like) states. To accommodate all these states in a functional framework,
32 Keith Dunker elaborated the Protein Trinity Hypothesis,⁸² which posits that a native protein can be in
33 one of three states, the ordered state, the collapsed-disordered (molten globule, MG) state and the
34 extended-disordered state (random coil, RC), and that function can arise from any of the three states
35 or from transitions between them.⁸² This model was subsequently expanded to include the pre-
36 molten globule (PMG) state, which corresponds to an intermediate state between the RC and the
37 MG (Fig. 9)¹³ (see also section 5.1).
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 **4.3. Complementary methods for studying protein disorder**

53
54
55
56
57
58
59
60

1
2
3 Certain low-tech approaches have provided ample early insight into the phenomenon of structural
4 disorder, and paved the way for the general acceptance of this concept. Even today they may
5 provide important insight and/or fast information on the structural state of a protein or its region.
6
7
8

9
10
11 IDPs can be recognized by their peculiar biochemical behavior. Indeed, due to their amino acid
12 sequence bias and the lack of a tightly packed hydrophobic core, the presence of disordered proteins
13 is often apparent during the purification process. Owing to their unique amino acid composition,
14 IDPs frequently bind less sodium dodecyl sulfate (SDS) molecules than globular proteins leading to an
15 aberrant mobility in SDS polyacrylamide gel electrophoresis (SDS-PAGE). Their apparent molecular
16 mass determined by this technique is usually 1.2-1.8-fold higher than expected from the amino acid
17 sequence measured by mass spectrometry.¹¹ Although this abnormal behavior has been ascribed in
18 the past to enrichment in acidic residues, recent findings suggest that other additional parameters,
19 such as extension in solution, can be responsible for the abnormal electrophoretic migration of IDPs
20 in SDS-PAGE.⁴⁹
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 An abnormal mobility is also observed in gel filtration (GF) or size-exclusion chromatography (SEC)
36 experiments, in which IDPs have unusually high apparent molecular mass.¹²
37
38
39
40

41 Furthermore, flexible regions are known to be highly susceptible to proteolysis. Thus, limited
42 proteolysis can be used to distinguish ordered and disordered proteins, as an increased proteolytic
43 degradation *in vitro* argues for increased flexibility. Early observations on the curiously protease
44 sensitive behavior of proteins were made on bacteriophage λ N,⁸³ dsp16⁸⁴ and p21,^{85,86} as outlined
45 in general in refs.^{11,16,87}
46
47
48
49
50
51
52
53
54

55 A further special behavior of IDPs is their insensitivity to high temperatures and their stability
56 towards acidic treatment, as observed early for MAP2,⁸⁸ α -synuclein,⁸⁹ involucrin,⁹⁰ and caldesmon,⁹¹
57
58
59
60

1
2
3 for example. These conditions usually causing the precipitation and/or denaturation of globular
4
5 proteins,⁵³ as often exploited in the purification of IDPs.⁹²⁻⁹⁴ These peculiarities of IDPs provide the
6
7 first line of indirect evidence of their unusual structural state. Direct structural information, however,
8
9 can be obtained by a wide variety of experimental approaches (reviewed in ref.⁹⁵).

10 11 12 13 14 15 16 17 **4.4. Analyzing disorder in a living cell**

18 19 20 21 22 **4.4.1. In vivo existence of protein disorder**

23
24
25
26
27 A major challenge in the field of IDPs is to assess to what extent *in vitro* observations on the
28
29 structural state and function of IDPs can be extrapolated to the living cell. Indeed, concentrations of
30
31 proteins and other macromolecules reach 300-400 mg/ml in the cell, giving rise to a crowding
32
33 effect^{96,97} that could affect the structural state of IDPs. Crowding might force IDPs to assume
34
35 compact or even folded states caused by excluded volume effects, making this issue critical with
36
37 respect to their physiological structural state and function. The presence of their binding partners
38
39 might suggest that IDPs are partner-bound (and folded) most of the time. There are two direct
40
41 approaches to address these issues: 1) mimicking crowding conditions in the test tube and 2)
42
43 characterizing IDPs directly in the living cell.
44
45

46
47
48 An ideal *in vitro* crowding agent combines three different effects: an excluded volume effect, an
49
50 increase in viscosity and an influence on the solvation/hydration of proteins. To mimic these,
51
52 accepted approaches are to apply high molecular-weight (Mw) polymers, such as Dextran and Ficoll
53
54 70, or small molecule osmolytes, such as sucrose, trifluoroethanol (TFE) or trimethylamine N-oxide
55
56 (TMAO). The high-Mw polymers primarily exert excluded volume effect, whereas osmolytes primarily
57
58
59
60

1
2
3 act upon solvation of the protein backbone.⁹⁸ Osmolytes tend to stabilize function-related secondary
4
5 structure of IDPs, as usually demonstrated by CD spectroscopy.⁵³ Crowding by Dextran and Ficoll-70
6
7 does elicit some compaction but not folding of IDPs,^{42,99,100} as demonstrated on different IDPs under
8
9 different conditions and using a variety of techniques.¹⁰¹ It appears that the formation of local
10
11 secondary structural elements of IDPs is promoted by crowding and thus preformed structural
12
13 elements (PSEs) may receive even more credit than suggested by *in vitro* studies.⁴²
14
15
16
17

18 To obtain direct information from the cell, two approaches could be used: assessing proteasomal
19
20 degradation and in-cell NMR studies. In the first approach, the ability of the 20S proteasome to
21
22 degrade unfolded proteins is exploited. This degradation “by default” (i.e. ubiquitin-independent
23
24 degradation) provides evidence on the physiological structural status of proteins. It was used to
25
26 characterize *in vivo* several IDPs including p53,¹⁰² and p21^{Cip1},¹⁰³ for which the default degradation by
27
28 the proteasome argues for their disordered state in the cell.
29
30
31
32

33 The application of in-cell NMR requires the deposition of labeled proteins within the cell either by
34
35 over-expressing under labeling conditions, by covalently linking cell-penetrating peptides to the
36
37 protein, or by microinjecting it into the cytoplasm of the cell.^{42,104} An example is the tau protein that
38
39 has been studied by in-cell NMR in *Xenopus* oocytes: while its microtubule-binding region became
40
41 ordered, its long projection domain remained largely disordered.¹⁰⁵ Very recently, the group of
42
43 Michele Vendruscolo has reported a study where in-cell NMR spectroscopy was used to observe
44
45 directly the structure and dynamics of α -synuclein within *E. coli* cells. In agreement with previous
46
47 studies by the group of Philipp Selenko,¹⁰⁶ the results indicate that, at least within the bacterial
48
49 cytosol, α -synuclein populates a highly dynamic state that, despite the highly crowded environment,
50
51 and it has the same characteristics as the disordered monomeric form observed in aqueous
52
53 solution.¹⁰⁷ α -synuclein was observed to be largely disordered in the periplasmic space of *E. coli*.¹⁰⁸
54
55
56
57
58
59
60

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

1
2
3 control the availability of IDPs, and that IDPs do undergo a faster turnover as compared to globular
4
5 proteins. Furthermore, the residence time (i.e. abundance, rate of synthesis, and half-lives of mRNA
6
7 and proteins) of IDPs, in different organisms including human cells, was found rather low due to
8
9 increased transcript clearance and proteolytic dégradation.¹¹⁴ The need of tight regulation of IDPs is
10
11 probably explained by their noted “dosage sensitivity” which denotes genes that are harmful when
12
13 overexpressed. Intrinsic protein disorder was found to be an important determinant for proteins
14
15 encoded by these genes, and explained by their inclination to be involved in promiscuous molecular
16
17 interactions.¹¹⁵ Regulation of dosage-sensitive genes occurs at the transcriptional, RNA, and protein
18
19 levels.¹¹⁶

20
21
22
23
24 Recent studies by Suskiewicz and colleagues proposed several possible further mechanisms by which
25
26 IDPs could be regulated *in vivo*,¹¹⁷ where IDPs would be protected from degradation by default (see
27
28 Fig. 10) through interactions with partners. In fact, forming a functional complex *in vivo* provides a
29
30 way to avoid degradation either by promoting order or by masking IDRs (Fig. 10A). Apparently, some
31
32 binding partners interact with IDPs principally to prevent their proteolysis. Those partners could be
33
34 proteasome gatekeepers (Fig. 10B) interacting with the 20S proteasome or nannies (Fig. 10C) that
35
36 are usually more specific for particular IDPs. Moreover, IDPs can bind to DNA (Fig. 10D), (for instance
37
38 IDPs functioning as transcription factors) or also to “decoy” binding sites (Fig. 10E) that can compete
39
40 with functional regions for binding to transcription factors. On the other hand, IDPs make
41
42 interactions with themselves also (Fig. 10F). In order to avoid proteolysis initiated by a disordered N-
43
44 terminus, an IDP could be stabilized by either a folded N-terminal domain (Fig. 10G), or by the
45
46 interaction with ribosomal proteins (Fig. 10H) until a longer sequence is synthesized that could be
47
48 stabilized by either a suitably nanny (Fig. 10C) or by a self-organization into a loopy structure (Fig.
49
50 10F).^{117,118} These proposed mechanisms actually relieve chaperones from the duty of guarding IDPs.
51
52
53
54
55
56
57
58
59
60

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

1
2
3 positioning of the secondary structure elements, and a compact state close to that of ordered
4
5 globular proteins.¹³ The considerable increase in the affinity for hydrophobic fluorescence probes
6
7 (ANS, for instance), the accessibility to proteases and the increase in the hydrodynamic radius which
8
9 corresponds to a volume increase of approximately 50% with respect to that the native globular
10
11 state, are diagnostic markers of the MG state.¹³ Recent analyses of IDPs by the group of Pappu
12
13 indicated sequence polarity and distribution of charged residues as determinants of IDP compaction:
14
15 indeed polar IDPs were found to favor collapsed ensembles in water despite the absence of
16
17 hydrophobic groups.^{124,125}
18
19

20
21
22 Extended disordered proteins also should not always be considered random coils. For practical
23
24 reasons, they can be divided into two separate sub-families,¹³ PMG and RC conformations. The PMG
25
26 state can effectively interact with the hydrophobic fluorescent probe ANS, though more weakly than
27
28 a MG, which suggests that some hydrophobic clusters are formed in this state. Therefore, the PMG
29
30 probably represents a “squeezed” and partially ordered form of the coil with some residual
31
32 secondary structure.¹²⁶ The RC, or rather, RC-like, state has hydrodynamic dimensions typical of
33
34 considerably unfolded polypeptide chains that are devoid of any ordered secondary structure.¹³
35
36
37

38
39 In general, PMGs are somewhat enriched in hydrophobic residues and have regions corresponding to
40
41 residual secondary structure, which can be pointed out by HCA analysis indicating regions with a
42
43 potential for induced folding (see section 3.2). It should be noted however that the content in regular
44
45 secondary structure is not a major determinant of protein compaction.^{49,127,128}
46
47
48

51 **5.2. Evaluating residual structure in IDPs**

52
53
54

55
56 The hydrodynamic volume of a single-domain globular protein in the MG, PMG and the unfolded
57
58 state, typically increases 1.5, 3 and 12 times, respectively, in comparison with that of the native
59
60

1
2
3 state.^{129,130} Hence, several techniques such as GF, small angle X-ray scattering (SAXS), analytical
4
5 ultracentrifugation and dynamic light scattering (DLS), to name a few, may help in determining
6
7 whether a protein is collapsed or extended due to the difference in its hydrodynamic properties (Fig.
8
9 12).¹³ An empirical formula connecting the Stokes radii (R_s) to the apparent molecular masses
10
11 (appMM) of proteins in different conformational states deduced from GF analysis enable to
12
13 distinguish between the various protein conformations (Fig. 12).^{12,13,95} Residual structure in IDPs can
14
15 also be unveiled by measuring the increase in hydrodynamic properties using DLS or SEC upon
16
17 heating or chemical denaturation.⁹⁵
18
19

20
21
22 The residual structure in IDPs can also be unveiled by spectroscopic techniques such as circular
23
24 dichroism (CD).¹³¹ The unfolded polypeptide chain is characterized by an intensive minimum in the
25
26 vicinity of 190 and 200nm and an ellipticity close to zero in the vicinity of 222 nm (Fig. 13A).⁵³
27
28 Consequently, the estimation of the secondary structure content allows MGs, with stable secondary
29
30 structures, to be distinguished from RCs or PMGs (Fig. 13A). To discriminate between the extended
31
32 states (i.e. PMG or RC) a “double wavelength” plot plotting the mean residue ellipticity at 222 nm as
33
34 a function of the mean residue ellipticity at 200 nm, was developed by Uversky (Fig. 13B).^{13,95}
35
36 Discrimination can also be improved by applying denaturants and/or heating the protein, and by the
37
38 simultaneous use of CD and hydrodynamic techniques or by the combined use of near-UV CD.¹³
39
40
41
42

43
44 Interactions between multivalent IDPs may lead to the formation of higher-order structural
45
46 organization, as demonstrated by sharp liquid-liquid-demixing phase separations, generating
47
48 micrometre-sized liquid droplets in aqueous solution. This macroscopic transition corresponds to a
49
50 molecular transition between small complexes and large, dynamic supramolecular polymers, with
51
52 the valency of the interacting species being directly related to the concentrations needed for phase
53
54 transition. Importantly, phase transitions can be related with changes in protein activity and can be
55
56 governed by posttranslational modifications (PTMs), as shown in the case of the actin-regulatory
57
58
59
60

1
2
3 neural Wiskott-Aldrich syndrome protein (N-WASP).¹³² The widespread occurrence of multivalent
4
5 systems suggests that similar phase transitions may be used to spatially organize and biochemically
6
7 regulate information in biological processes.¹³³
8
9

10 11 12 **5.3. IDPs as conformational ensembles** 13 14 15

16
17 The full structural description of IDPs is only feasible by ensemble approaches, primarily based on
18
19 NMR observables, such as chemical shifts (CSs), Residual Dipolar Couplings (RDCs), ¹⁵N R₂ relaxation
20
21 rates, Paramagnetic Relaxation Enhancement (PRE) distance restraints, J-couplings, pulsed field
22
23 gradient (PFG)-derived Rh values, and ¹H-¹⁵N heteronuclear Nuclear Overhauser Effects (hetNOEs),
24
25 complemented by SAXS, which all together yield detailed information on short- and long-range
26
27 contacts and backbone flexibility.¹³⁴⁻¹³⁶ By all this information combined, one can describe the
28
29 structure of an IDP as an ensemble of conformations. To this end, two broad approaches have been
30
31 developed: either through restrained molecular dynamics simulations, or the generation of a large
32
33 number of conformations and selection of a subset that fits the data.⁶⁴
34
35
36
37

38
39 In molecular dynamics simulations, the conformational space accessible to a given protein is sampled
40
41 by penalizing deviations between the experimental data and the corresponding values back-
42
43 calculated from the structures sampled during the simulations.⁶³ It should be noted that there is no
44
45 guarantee that the generated structural ensemble would be consistent with experimental data not
46
47 used as restraints. This is a general consequence that we cannot measure sufficient number of
48
49 restraints at the moment.^{64,137}
50
51
52

53
54 The second approach starts with generating a very large pool of conformations, which may be
55
56 completely random or constrained by experimental or theoretical data (e.g. Ramachandran Ψ/Φ
57
58 angles taken from coil regions of PDB structures). Programs such as Flexible-Meccano,¹³⁸ Ensemble
59
60

1
2
3 Optimization Method,^{135,139} and TRaDES^{140,141} are currently used for this purpose. If not provided by
4
5 the initial algorithm, side chains can be modeled in with an algorithm such as SCCOMP¹⁴² or
6
7 SCRWL.¹⁴³ Expected experimental observables are then back-calculated for each conformation in the
8
9 pool (with CRY SOL for SAXS,¹⁴⁴ or ShiftX or SPARTA for NMR¹⁴⁵).

10
11
12
13 From the pool, a subset of structures (on the order of 50-100) is selected through step-wise
14
15 optimization until the best fit is achieved. The ensemble describes the conformational state of the
16
17 IDP, to the limit of the amount and quality of experimental data. Obviously, one of the most
18
19 important issues in IDP research is to define the number (or range) of conformers for such ensembles
20
21 to be relevant. Although this can only be addressed by performing the corresponding computational
22
23 experiments, we expect the size of the ensemble to be roughly proportional to the extent of disorder
24
25 in a given protein. It is also possible that the size of the ensemble can depend on the particular type
26
27 of IDP function. Recent analyses revealed that in most cases 50-100 structures in the ensemble
28
29 suffice in describing experimental observations⁵⁹, and even in large (such as α -synuclein and tau) and
30
31 highly extended (p53TAD) disordered proteins, their conformational ensemble can be reliably
32
33 described by about 200-400 members.^{146,147}

34
35
36
37 Ensembles can be deposited in the newly developed protein ensemble database (pE-DB,
38
39 <http://pedb.vib.be>), which hosts primary experimental data, algorithms used for calculation and
40
41 ensembles for IDPs.¹⁴⁸ Because ensemble are not unequivocally defined by structural data available,
42
43 one protein can have several ensembles, and it is the major aspiration of the IDP field to find out
44
45 how relevant ensembles are, i.e. if the function of IDPs can be interpreted by the ensembles. If we
46
47 achieve this goal, it will signal the dawn of a new era in unstructural biology.¹⁴⁹
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 collapse into compact states.¹²⁵ The conformational preferences of strong polyampholytes, on the
4
5 other hand, are determined by both the fraction of charged residues and the sequence distribution
6
7 of oppositely charged residues. In the case of well-mixed sequences, electrostatic repulsions and
8
9 attractions within the chain are counterbalanced, leading to a preference to a self-avoiding random
10
11 walk or generic Flory-type random coil conformational state. When oppositely charged residues are
12
13 segregated within the sequence, hairpin-like conformations emerge because long-range electrostatic
14
15 attractions are preferred. It was also observed that the naturally occurring polyampholytes are
16
17 usually well mixed, with a preference for random-coil ensembles. These inferences have been further
18
19 born out by FRET experiments, which show that IDPs exhibit expansion at low ionic strength that
20
21 correlates with their net charge.¹⁵² Charge-balanced polypeptides exhibit an additional collapse at
22
23 low ionic strength conditions, yet they show low internal friction, which may have a significant
24
25 impact on reconfiguration time.¹⁵³ Reconfiguration time is particularly relevant for the kinetics of
26
27 conformational reorganization, which is directly linked with the dynamics of the interactions of IDPs
28
29 with cellular binding partners.
30
31
32
33
34
35
36
37

38 **5.5. IDPs, their environment and functional misfolding**

39
40
41
42

43 Noteworthy, many examples in the literature pointed out that the structure of IDPs is very sensitive
44
45 to their environment, often displaying an inside-out behavior, i.e. a partial folding of IDPs with
46
47 increasing temperature or increasing/decreasing pH.^{13,154-156} Indeed, this behavior makes perfect
48
49 sense: increasing the temperature can increase the strength of the hydrophobic interactions leading
50
51 to a stronger hydrophobic driving force for partial folding, whereas variations in pH minimize the
52
53 large net charge present at neutral pH and consequently decrease the charge/charge intramolecular
54
55 repulsion and thus shift the conformational ensemble towards more compact states.^{13,53}
56
57
58
59
60

1
2
3 Intriguingly, some regions cross the disorder/order boundary or assume different structures as a
4
5 result of slight environmental changes. This “ambiguous” behavior has been addressed via two
6
7 concepts, Chameleon Sequences¹⁵⁷ and Dual-Personality (DP) sequences.¹⁵⁸ While chameleon
8
9 sequences fold in a template-dependent manner and could change conformation within the same
10
11 protein as a result of point mutations, ligand binding or a change in pH, DP fragments are close to the
12
13 order/disorder boundary. Owing to DP fragments, identical proteins solved in different conditions
14
15 and functional states can be captured in both conformational states: ordered and disordered. DP
16
17 sequences possess an amino acid composition intermediary between ordered globular and
18
19 disordered proteins. DP fragments are thus defined as a distinct structural category.⁴² When IDPs
20
21 appear to have the capacity to adapt to the structure of distinct partners and to carry out different,
22
23 occasionally opposing functions, the concept is denoted as “moonlighting”.¹⁵⁹
24
25
26
27

28
29 Finally, how does Nature prevent highly promiscuous IDPs from undergoing unwanted interactions
30
31 with non-native partners? Interesting findings by Uversky¹⁶⁰ suggest that IDPs escape unwanted
32
33 interactions via functional misfolding (i.e. via the formation of non-native intramolecular
34
35 interactions). In fact, since binding of IDPs to partners is frequently driven by hydrophobic residues
36
37 organized in patches, the occurrence of PSEs with a non-polar continuous surface will surely promote
38
39 the mutual attraction of these regions within the IDP and the formation of a structure where these
40
41 elements are oriented to each other according to their interacting surfaces. In other words, the
42
43 formation of native-like functionally relevant secondary structure elements containing sites for
44
45 specific binding of native partners has to be accompanied by partial compaction of a polypeptide
46
47 chain due to non-native intramolecular interactions (i.e. misfolding), thus native binding sites will no
48
49 longer be exposed to interact with non-native partners.
50
51
52
53
54

55 **6. Intrinsic disorder and protein function**

56
57
58
59
60

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

1
2
3 with inhibitor 2 (I2) binding to, and inhibiting protein phosphatase 1.^{167,168} Effector functions of IDPs
4
5 are probably the most straightforward to interpret in terms of the classical structure-function
6
7 paradigm because the resulting complexes are often found in the PDB. When the effector has both
8
9 activities, activator and inhibitor, sometimes with the same partner, the term “moonlighting”¹⁶⁹ is
10
11 used (see also section 5.5).
12
13

14
15
16 The third class is that of scavengers, in which the IDP stores and/or neutralizes small ligands. In fact,
17
18 the open and extended structure of IDPs is particularly suited to bind small ligands, such as ions and
19
20 organic compounds, either for disposal or for sequestration and later release upon the need of the
21
22 organism.⁴² For example, salivary proline-rich glycoproteins, which constitute about two-thirds of the
23
24 total protein in saliva, are scavengers that neutralize polyphenolic plant compounds (i.e. tannins).¹⁷⁰
25
26 Biomineralization proteins storing and delivering calcium/phosphate to the site of the buildup of the
27
28 mineral phase in bone and teeth also exemplify this mode of action.¹⁷¹
29
30
31

32
33 The fourth class constitutes assemblers, which assemble, stabilize and regulate large multiprotein
34
35 complexes such as ribosome, cytoskeleton and chromatin.¹¹ In agreement, both prediction and
36
37 structural data show statistically significant correlations between disorder and either the number of
38
39 protein subunits of complexes^{42,172} or established scaffold function.^{173,174} In addition, high disorder
40
41 also appears in hub proteins.^{175,176} In interactomes, a few proteins possess a large number of
42
43 connections (i.e. hubs), whereas most others have very few, often only one, connection. This
44
45 arrangement suggests a functional specialization in which hubs are preferentially involved in
46
47 organizing the network. Interestingly, a range of bioinformatics studies suggests that hub proteins
48
49 have an elevated level of disorder (reviewed in ref.¹⁷⁷). For example, Elisar Barbar and colleagues¹⁷⁸
50
51 describe how dynein light chain (LC8) interacts with multiple partners, many of which are not
52
53 connected with dynein or microtubule-based transport, the function that was first attributed to this
54
55 protein.¹⁷⁸
56
57
58
59
60

1
2
3
4
5 A special case of molecular recognition is exhibited by the fifth class, display sites, which mediate
6 regulatory PTMs, such as phosphorylation or limited proteolysis. PTM of proteins has three structural
7 requirements: an appropriate local sequence, structural exposure, and flexibility of the site so that it
8 can be productively accommodated by the active site of the modifying enzyme; these requirements
9 are met by structural disorder. Indeed, computational and experimental evidence support the
10 abundance of PTM such as phosphorylation¹⁷⁹ and ubiquitination¹⁸⁰ sites in disordered regions of
11 proteins. In addition, a recent study suggests that IDPs in yeast are more likely substrates of a large
12 number of kinases.¹¹⁴ The power of DisPhos (disorder-enhanced phosphorylation) in predicting
13 phosphorylation sites¹⁷⁹ revealed that phosphorylation sites primarily occur in regulatory, cancer-
14 associated and cytoskeletal proteins, as opposed to proteins involved in degradation, biosynthesis,
15 and metabolism, a feature associated with protein disorder from early on.¹⁷⁹
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31 A recent addition to the list of functional modes of IDPs is chaperone function. A bioinformatics
32 analysis showed that a high proportion of residues in RNA (54.2%) and protein (36.7%) chaperones
33 fall into locally disordered regions.¹⁸¹ In addition, certain chaperone proteins are disordered along
34 their entire length, e.g. α -synuclein,¹⁸² β -casein,¹⁸³ or LEA proteins,^{184,185} or chaperone activity is
35 localized to disordered segments, such as in the small heat-shock protein (Hsp) α -crystallin.^{186,187}
36 Disordered regions in chaperones can bind different partners (i.e. promiscuity), enable an enhanced
37 speed of interactions and uncouple specificity from binding strength, which makes sense as
38 chaperones need to bind a wide range of unrelated misfolded substrates. Moreover, the highly
39 hydrophilic character of IDRs affords a solubilizing effect, preventing their aggregation^{42,181} as
40 demonstrated by fusing disordered segments to a range of aggregation-prone proteins in expression
41 studies.¹⁸⁸ To accommodate all different aspects of the mechanism of action of disordered
42 chaperones, an “entropy transfer” model has been suggested, in which transient binding,
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 solubilization and entropic exclusion elements are blended with transient interaction with the
4
5 misfolded region of the substrate (client).¹⁸¹
6
7

8
9 The role of disorder in chaperone action has been most convincingly demonstrated in the recent
10
11 example of the redox-regulated chaperone Hsp33.¹⁸⁹ This chaperone, which protects organisms
12
13 against oxidative stress, is inactive in its fully folded state, but it undergoes oxidative unfolding within
14
15 its redox-sensor domain upon stress. The IDR thus generated is used to discriminate between
16
17 unfolded and folded partners. Upon returning to normal conditions, Hsp33 returns to the folded
18
19 state and transfers bound clients to ATP-dependent foldases, i.e. its activity depends on internal
20
21 order-to-disorder transitions.
22
23
24
25
26

27 **6.2. Short recognition motifs in the interactions of IDPs. Interaction networks, one-** 28 29 30 **to-many and many-to-one binding** 31 32

33
34 As also apparent from the foregoing classification, IDPs often function via molecular recognition,
35
36 when they bind partner molecules in and induced folding process. Their mode of binding is thought
37
38 to confer many advantages, and is often mediated by short recognition elements (motifs).^{190,191}
39
40
41
42

43
44 Different concepts regarding short recognition elements have been extensively discussed in the
45
46 literature. Indeed, depending on whether the idea is approached from a structural point of view or
47
48 defined at the sequence level, a short motif could be denoted as “Molecular Recognition Element”
49
50 (MoRE) / “Molecular Recognition Feature” (MoRF)¹⁹² or “Linear Motif” (LM), respectively (LM is also
51
52 denoted as “Eukaryotic Linear Motif” (ELM) or “Short Linear Motif” (SLiM)).¹⁹³
53
54
55
56
57
58
59
60

1
2
3 Moreover, Fuxreiter et al. in 2004 elaborated the concept of preformed structural elements (PSEs)
4 correlating the probable structural preferences of IDPs in the unbound state.¹⁹⁴ The structure of the
5 recognition element of an IDP in complex with its partner is often known, and comparison of
6 structures adopted in the bound form and that observed (or predicted) in the free form suggested
7 that IDPs have rather strong intrinsic preferences for the conformation they attain when bound to
8 their partners, with these preferences being highest for α -helical conformations. This may be
9 interpreted in terms of partial pre-formation of their recognition segments in the free state.¹⁹⁴
10 Indeed, often a similar structure in the unbound and bound states is observed when the IDP is
11 characterized by NMR, which formed the basis of the related concept of Pre-structured Motifs
12 (PreSMos).¹⁹⁵
13
14
15
16
17
18
19
20
21
22
23
24
25
26

27 In a similar vein, MoREs/MoRFs has been established as short regions of (predicted) disorder that
28 became structured and thus deposited into PDB, in complex with their partner. MoREs/MoRFs can
29 adopt α -helix, β -strand or irregular secondary structure conformations (leading to α -, β - and I-
30 MoREs, respectively) or a combination of different secondary structural forms (complex-MoREs) (Fig.
31 15).¹⁹² Although an exact relationship has not yet been systematically tested,¹⁹⁶ it is appealing to
32 assume that MoREs/MoRFs could correspond to PSEs. Repartition of secondary structures, which
33 suggests significant differences between structures of IDPs in the bound state and those of globular
34 proteins, supports this similarity. Indeed, secondary structure statistics on 26 IDRs in complex with
35 their partners and on a reference dataset of globular proteins revealed that helices are almost
36 equally populated in both datasets. On the other hand, β -structures in IDPs are about 50% less
37 frequent than in globular proteins. The most significant difference is the increased level of coil
38 conformation in bound IDPs. The analysis of the distribution of secondary structures in
39 MoREs/MoRFs in the PDB is in complete agreement with these findings: such short binding elements
40 have 27% of their residues in α -helical conformation, 12% in β -strands and approximately 48% in
41 irregular conformations (13% missing from the atomic coordinates).^{20,42,197,198}
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 Molecular recognition by short recognition elements (motifs) can - and has been historically -
6 approached from the completely different direction of short sequence patterns determining
7 functional interactions and enzymatic modification, i.e. LMs, ELMs, and SLiMs.^{190,199,200} These motifs
8 are linear in the sense that 3-D organization is not required to bring distant segments of the molecule
9 together to build up the recognizable unit.¹⁹³ Disorder prediction of LMs and their flanking regions
10 for the experimentally characterized examples of the ELM database (a database gathering Eukaryotic
11 linear motifs) (<http://elm.eu.org>)¹⁹⁹ suggests that LMs and their flanking regions are segments of
12 intrinsic disorder within a more ordered environment.¹⁹¹ Their inherent flexibility probably allows
13 them to adopt various conformations and to bind to multiple partners. LMs, however, are enriched
14 in hydrophobic residues (Trp, Leu, Cys and Tyr), charged Arg and Asp, and Pro residues, and they are
15 depleted in Gly and Ala. This could be explained by the fact that their restrictive sites (RS) directly
16 involved in recognition are either hydrophobic and rigid, or charged and flexible, whereas at the non-
17 restrictive sites (NRS), excessive flexibility, even exceeding average IDPs, can be observed.^{42,190,191}
18 Overall, the unique amino acid composition suggests a mixed nature of LMs, with a few specificity-
19 determining residues strongly favoring order grafted on a completely disordered carrier sequence
20 flanking the region critical for interaction.⁴²
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 Although not fully substantiated,²⁰¹ the three concepts of short recognition motifs (PSEs, LMs,
42 MoREs) can probably be considered as manifestations of the same underlying principle of binding of
43 an ordered partner by a short segment within a disordered region, which undergoes a disorder-to-
44 order transition or induced folding upon binding.²⁰² All the recognition motifs are short, on the order
45 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
46 MoREs.^{20,159} Moreover, they possess an amino acid composition that resembles that of IDPs with
47 some notable deviations (i.e. specificity-determinant residues favoring order) and they can even in
48 some cases be already preformed in the free state (PSEs, PreSMOs).^{191,193-195}
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 It is to be noted, however, that binding regions within IDPs might correspond more to a domain than
6 a motif, with their length exceeding 20-30 residues.²⁰³ Indeed, these regions possess typical
7 characteristics of domains: 1) they are structurally and functionally independent of the remainder of
8 the protein molecule, 2) they can be recognized by homology due to evolutionary conservation of
9 sequences,²⁰⁴⁻²⁰⁶ and 3) they possess at least one specific function. In fact, about 14% of Pfam
10 domains (i.e. conserved evolutionary domains gathered in the Pfam database) selected from the
11 SwissProt database possess more than 50% of predicted disorder and 4% are fully disordered.
12
13
14
15
16
17
18
19
20
21

22 Due to their functional importance, predicting recognition elements (motifs) from sequence is of
23 great interest. Although the information content of short sites is very limited, due to their
24 enrichment in hydrophobic residues, indirect techniques have some success in delineating them. As
25 mentioned, for example, the HCA plot⁴³ (see section 3.2), can indeed unveil such binding sites.²⁵ The
26 PONDR VL-XT,²³ ANCHOR (<http://anchor.enzim.hu>)²⁰⁷ and DynaMine²⁰⁸ disorder predictors are
27 sensitive to local tendency of ordering, and are thus informative in highlighting potential induced
28 folding regions. In the case of PONDR VL-XT, it was shown that there is a correlation between
29 downward spikes in disorder scores and binding regions within IDRs.¹⁹² On the other hand, ANCHOR
30 relies on the pairwise energy estimation approach (e.g. the approach of IUPred, see section 3.1)
31 which identifies segments that are in disordered regions and cannot form enough favorable intra-
32 chain interactions to fold on their own and are likely to gain stabilizing energy by interacting with a
33 globular protein partner.²⁰⁹ Although motif discovery from sequence alone is fraught with extremely
34 high false positive rates, inclusion of additional data, such as enrichment in functionally related
35 proteins, can significantly increase its confidence, as demonstrated in the SLIMDisc approach (for an
36 example see ref.²⁰⁰). Furthermore, an ELM server (<http://elm.eu.org/>) has been developed in view of
37 investigating short functional sites in modular Eukaryotic proteins.¹⁹⁹ The recently developed
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 dynamics predictor, DynaMine,²⁰⁸ shows that recognition motifs have characteristic patterns of local
4
5 dynamics, which could be used for improving the accuracy of motif prediction.
6
7
8
9

10 **6.3. Variability of binding modes attainable by disordered proteins: folding before** 11 12 **or after binding** 13

14
15
16
17 It is currently accepted that the crux of the binding of IDPs is binding-induced folding, which can
18 occur between an IDP and a globular partner, but also between two IDPs, with the resulting
19 complexes often being amenable to structural studies.²⁰² It is of significant interest whether folding
20 occurs before, after or concomitantly to binding since experimental evidence seems to support all
21 these varieties. NMR, molecular dynamics (MD) and site-directed mutagenesis are approaches to
22 study the folding and binding events.⁴² Although conformational selection requires that the IDP
23 populates at least partly the bound state in the free form, which therefore presages the final bound
24 conformation,^{194,195} it is important to emphasize that the pre-existence of folded structures in IDPs
25 does not necessarily imply a conformational selection mechanism.^{73,210,211} To further complicate the
26 scenario, many IDPs show a high degree of adaptability with respect to the partner (Fig. 16), i.e. they
27 can fold into different structures on binding to different partners. This plasticity raises questions
28 about the role of PSEs in the recognition process, and suggests that the conformation sampled in
29 solution will not necessarily be favored upon binding to a partner;^{212,213} i.e., there seems to be no
30 general mechanism for the binding-induced folding of IDPs.⁴² Altogether, and in spite of recent
31 breakthroughs, our understanding of the molecular features that encode the binding efficiency of
32 IDPs is far from being elucidated and still matter of controversies (see ref.²¹⁴ for a review on this
33 topic).
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Generally, one should note that a major functional consequence of induced folding is a decoupling
4
5 between specificity and affinity (i.e. binding strength) to enable weak and reversible interactions (Fig.
6
7 17).¹⁹² In contrast with interactions involving globular proteins, where specificity and affinity are
8
9 often coupled,¹⁹² when binding involves disorder-to-order transitions, entropy becomes a main
10
11 factor in the free energy of binding, and the large decrease in conformational entropy uncouples
12
13 binding strength from specificity, which makes highly specific interactions reversible.^{159,192} To limit
14
15 unfavorable effects on binding and ensure kinetically and energetically “facile” interactions between
16
17 IDPs and their physiological partner(s),¹⁹⁴ the “conformational selection” model was proposed, in
18
19 which pre-existent native-like conformations dominate the binding process.^{215,216} It is however to be
20
21 expected that some degree of conformational adaptation will be required even when a preformed
22
23 structural element in an IDP binds to its target.²¹² On the other hand, unlike “conformational
24
25 selection”, much experimental evidence in the literature proposes that folding of IDPs can only take
26
27 place upon binding and hence contacts (usually hydrophobic) with the binding-partner constitute the
28
29 major driving force for the folding.^{202,212} In fact, recently, Phi-value analysis developed to study the
30
31 mechanism of folding of globular proteins^{217,218} has been implemented in the IDP field. By this
32
33 technique, residue-specific structural information is inferred by comparing the kinetics of the
34
35 reaction (folding and/or binding) of the wild- type protein with a series of conservative single
36
37 mutants, probing the presence of native-like structure in the transition state. Detailed kinetic and
38
39 thermodynamic analysis showed that IDPs may bind both by conformational selection and induction
40
41 of folding in the bound state.²¹⁹⁻²²²

42
43
44
45
46
47
48 Primary binding via an extended, disordered state, is the key element of the so-called “fly-casting”
49
50 binding mechanism.^{223,224} It was assumed that IDPs can have an enhanced capture radius for a
51
52 specific binding site, which confers an effective enhanced ability to search for a partner (also termed
53
54 protein fishing).²²⁴ This mechanism is closely related to another concept of recognition by short-
55
56 motifs, that of “Primary Contact Sites” (PCS), which derives from the observation that IDPs can often
57
58
59
60

1
2
3 attain the bound state very rapidly, suggesting that certain regions within their fluctuating structural
4 ensemble might be exposed for initiating a productive interaction with the partner.^{42,225}
5
6
7

8
9 It appears that the three mechanisms “conformational selection”, “induced folding” and “fly-casting”
10 can be merged into a single “synergistic model”, in which all these elements synergistically
11 participate. A brief description of such a consensus mechanism from ref.²²⁶ is provided (Fig. 18): the
12 first step can be described as the non-specific reeling of the IDP by its target molecule via a “fly-
13 casting” mechanism. Once the IDP is close enough to its target, specific encounter is facilitated by
14 PSEs, and “conformational selection” comes into play. Full formation of the complex, largely
15 dominated by the folding coupled to binding, constitutes the third and last step of the consensus
16 mechanism. The driving force of binding is provided by both the favorable intermolecular interaction
17 energy and restriction of conformational entropy by binding conformationally restricted PSEs.²²⁶ An
18 excellent example studied by NMR and MD experiments is the pKID and CBP complex
19 formation.^{227,228}
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 It is of note, however, that in some complexes the binding does not induce folding. In such “random
36 complexes”, the IDP interacts with its partner via transient contacts, and it behaves similar to its free
37 form. This mechanism seems to operate in the C-terminal domain of the full length *E. coli* SSB that is
38 disordered even when bound to DNA,²²⁹ and in T-cell receptor cytoplasmic tails forming homo-
39 oligomers without assuming a folded state.^{230,231} A similarly high flexibility in the bound form was
40 also observed in the complex between the C-terminal disordered N_{TAIL} domain of the nucleoprotein
41 and the X domain (XD) of the phosphoprotein from Hendra virus, where the resonance behavior
42 suggests α -helical fraying of the MoRE of N_{TAIL} on the surface of XD.²³² This binding via a disordered
43 state has been termed “fuzziness”^{159,233} (see section 6.4).
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The chemical nature of the interface of IDP complexes also constitutes an area of interest. Indeed, it
4
5 has been shown that the chemical and physical features of interfaces of IDPs are distinct from those
6
7 of globular proteins in four major aspects.^{198,234}
8
9

10
11 1) Geometric particularities. IDPs use a large proportion of their surface for interaction with their
12
13 partner (50% for IDPs whereas globular proteins use 5%-15% only).²³⁴ In terms of absolute values,
14
15 however, the interfaces of IDPs are slightly smaller than those of ordered complexes ($1141 \pm 110 \text{ \AA}^2$
16
17 for IDPs²³⁴ versus $1600 \pm 400 \text{ \AA}^2$ for globular proteins,²³⁵) which stems from distinct interaction
18
19 topologies, i.e. from the fact that in 70% of the cases the binding surface of the IDP represents a
20
21 sequentially continuous segment, and it never contains more than three separate segments. Ordered
22
23 proteins, on the other hand, hardly ever use a single segment for binding to their partner, and their
24
25 segmentation number may occasionally even exceed 10. It has been inferred that these differences
26
27 might also be relevant for the weaker interactions of IDPs that might also entail less specificity.^{236,237}
28
29
30
31
32

33 2) *Chemical nature of interfaces: buried versus exposed residues.* The ratio of buried-to-exposed area
34
35 of IDPs is much smaller for both polar and hydrophobic residues, which suggests that IDPs keep even
36
37 their very few hydrophobic residues exposed for contact with the partner. In addition, interfaces of
38
39 complexes involving IDPs were found to be enriched in hydrophobic residues, with these latter
40
41 contributing significantly to the stability of the interaction interface.²³⁴ In a recent study, although
42
43 enrichment in hydrophobic residues was confirmed, polar interactions were found to play a larger
44
45 role in these complexes than in structured protein complexes, with interfaces being more
46
47 complementary with respect to electrostatics than interfaces of globular proteins.²³⁸
48
49
50
51
52

53 3) *Interaction energies at the interface.* Ordered proteins tend to establish more stabilizing
54
55 interactions within their polypeptide chains, whereas IDPs derive more stabilization from the
56
57 interaction with the partner.²³⁴ These features are also in line with the noted chemical nature of the
58
59
60

1
2
3 interface, the higher ratio of hydrophobic amino acids and tighter fit, which compensate for the
4
5 larger unfavorable decrease in entropy that occurs in folding induced upon binding.
6
7

8
9 4) *Conservation of interfaces*. The regions that become ordered upon complex formation in IDPs are
10 the most conserved,²³⁴ in contrast with the general tendency of evolutionary variability of IDPs.²³⁹ In
11 fact, interface positions show the most pronounced tendency to be conserved, in complete line with
12 their functional importance, which helps their identification from sequence as linear motifs.²⁴⁰
13
14
15

16
17
18 All these findings are in line with the exposure of hydrophobic amino acids in ELMs¹⁹¹ and MoRFs¹⁹⁸
19 and their tight interactions due to a better adaptation to the structure of the partner enabled by
20 their induced folding. Recently, it has been confirmed that aromatic-aromatic intermolecular
21 interactions in complexes implying IDPs occur frequently at the interface of the complexes, which
22 provide both specificity and stability to the complexes and are enthalpically favorable.²⁴¹
23
24
25
26
27
28
29
30
31
32

33 34 **6.4. Dynamic or fuzzy complexes**

35
36
37 In spite of the insight provided by the folding coupled to binding concept, ordering of IDPs is very
38 often not complete even in the presence of the partner, i.e. the protein remains partially disordered
39 in the bound state (Fig. 19). This led Tompa and Fuxreiter to coin the term “fuzziness”,^{159,233} which
40 represents an extension of the concept of disorder in the bound state.^{159,242}
41
42
43
44
45
46
47

48
49 Apparently, this phenomenon presents a widespread novel paradigm of protein structure and
50 function as supported by a dozen of IDP-complex examples gathered and described in detail,^{201,233}
51 pointing to the abundance of fuzziness in IDP-complexes. Fuzziness may mean only static structural
52 promiscuity, when the IDP has more than one stable bound states, or it may encompass
53 dynamic/disordered parts of the bound IDP. A linker region connecting two binding regions, or a
54
55
56
57
58
59
60

1
2
3 flanking region might remain disordered, sometimes even the entire IDP fails to acquire a stable
4
5 structure, as observed in the case of the cytoplasmic domain of T-cell receptors.^{230,231}
6
7

8
9 Many functional advantages can result from “fuzziness”, including interactions with alternative
10
11 partners and simultaneous interactions with different partners.⁴² Indeed, the residual plasticity often
12
13 allows adaptation of the same motif to different partners, or a variable arrangement of the
14
15 recognition motifs, which can mediate interactions with alternative partners (i.e. promiscuity).
16
17 Moreover, the disordered tails in complexes serve for partner fishing with nonspecific, transient
18
19 contacts, the first step in the above-proposed model for IDPs interactions.²²⁴ Fuzzy parts of the
20
21 complexes can harbor regulatory PTM sites or they can mediate interactions with additional
22
23 partners. They can even directly or indirectly interfere with (promote or inhibit) binding of the part
24
25 that undergoes folding transition. An illustrative example is provided by a proline-rich segment of the
26
27 myelin basic protein (MBP) that can bind to several SH3 domain-containing proteins.²⁴³ In addition,
28
29 fuzziness provides a way to reduce entropic penalty thereby affording enhanced affinity.
30
31
32
33
34
35

36 **6.5. IDPs and posttranslational modifications**

37
38
39

40
41 PTMs of proteins are reversible or irreversible chemical changes of a polypeptide chain that occur
42
43 after translation. PTMs range from the enzymatic cleavage of peptide bonds to the covalent
44
45 additions of particular chemical groups, lipids, carbohydrates or even entire proteins to amino acid
46
47 side chains. PTMs extend the range of amino acid structures and properties, and diversify structures
48
49 and functions of proteins.²⁴⁴ As many as 300 post-translational modifications of proteins are known
50
51 to occur physiologically,²⁴⁵ and because of the various PTMs, the actual number of chemically
52
53 different amino acids in proteins is believed to increase from 20 to more than 140.²⁴⁴ In general,
54
55 proteomes are significantly more complex than one can expect from the analysis of the encoding
56
57
58
59
60

1
2
3 genomes mostly due to two mechanisms, PTMs and alternative splicing of their mRNAs (see
4
5 below).²⁴⁴
6
7

8
9 Some PTMs (e.g., phosphorylation, acetylation or palmitoylation) are readily reversible by the action
10 of specific deconjugating enzymes. The interplay between modifying and demodifying enzymes
11 allows for rapid and economical control of protein function. A similar control by protein degradation
12 and *de novo* synthesis would take much longer time and cost much more energy. PTM can occur at
13 any stage of the protein's life. Some proteins are modified shortly after their translation is completed
14 and prior to the final steps of their folding. These early PTMs might affect the protein folding
15 efficiency, protein conformational stability and even determine the fate of the nascent protein via
16 directing it to distinct cellular compartments. Other proteins are modified after their folding and
17 localization are completed. Here, PTMs can activate or inactivate catalytic activity or otherwise
18 influence the biological activity of a protein.²⁴⁴
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33 Some proteins require multiple different posttranslational modifications for their function. For such
34 multi-PTM proteins, modified sites in proteins can not only mediate individual functions, but can also
35 function together to fine-tune molecular interactions and to modulate overall protein activity and
36 stability.^{38,246} One illustrative example of such proteins is provided by a family of nuclear IDPs,
37 histones, which require methylation, acetylation, phosphorylation, ubiquitylation, ADP-ribosylation,
38 and SUMOylation at different stages of their action, with different modifications affecting histone–
39 DNA interactions and also histone–histone interfaces, thus providing the capacity to disrupt intra-
40 nucleosomal interactions and to alter nucleosome stability.²⁴⁷ Although the N-terminal domains of
41 the core histones are known to contain an extraordinary number of sites that can be subjected to
42 PTM, over 30 histone modifications have been recently identified in the core domains too.²⁴⁸
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Based on the peculiarities of the conformational state of the site where the PTM would take place,
4
5 PTMs can be grouped into two major classes, namely modifications mostly affecting ordered proteins
6
7 and modifications occurring in disordered proteins or disordered regions.^{154,249} Among disorder-
8
9 targeting PTMs are phosphorylation, acetylation, acylation, adenylation, ADP ribosylation, amidation,
10
11 carboxylation, formylation, glycosylation, methylation, sulfation, prenylation, ubiquitination, and
12
13 Ubl-conjugation (i.e., covalent attachment of ubiquitin-like proteins, including SUMO, ISG15, Nedd8,
14
15 and Atg8).¹⁵⁴ Many sites of protein phosphorylation are located within intrinsically disordered
16
17 regions,^{161,179,250} and several other types of PTMs, such as acetylation, protease digestion,
18
19 ubiquitination, fatty acid acylation, and methylation, also abound in regions of intrinsic
20
21 disorder.^{161,180,250,251} It has been suggested that sites of the enzyme-catalyzed PTMs in eukaryotic cells
22
23 exhibit a preference for intrinsically disordered regions.^{129,252} In fact, the modifying enzyme binds and
24
25 modifies similar sites in a wide variety of proteins, which is facilitated by disorder of flanking
26
27 regions.^{129,252}
28
29
30
31
32
33

34 **6.6. Alternative splicing in IDP function and dysfunction**

35
36
37

38
39 Alternative splicing (AS) is a process by which two or more mature mRNAs are produced from a
40
41 single precursor pre-mRNA by the inclusion and omission of different segments.^{253,254} AS is commonly
42
43 observed mostly in multicellular eukaryotes.^{255,256} For humans and other mammals, multiple proteins
44
45 are often produced from a single gene, and 40 – 60% the genes yield proteins *via* AS.²⁵⁷⁻²⁶⁰ It was
46
47 hypothesized that AS very likely provides an important mechanism for enhancing protein diversity in
48
49 multicellular eukaryotes,²⁶¹ playing a crucial role in regulation of protein-protein interactions, ligand
50
51 binding, and enzymatic activity,²⁶²⁻²⁶⁵ and yielding protein isoforms with different, even opposing,
52
53 functions.
54
55
56
57
58
59
60

1
2
3 The existence of a strong prevalence of intrinsic disorder in protein regions affected by alternative
4 splicing of their corresponding pre-mRNAs has been noticed long ago.²⁶⁶ Such commonness of
5 intrinsic disorder in protein regions affected by alternative splicing is used by nature to avoid
6 structural complications. In fact, removal of a piece of sequence from a structured protein can lead
7 to protein missfolding, causing protein aggregation and loss of function. Such structural catastrophe
8 can be avoided when AS maps to intrinsically disordered regions.²⁶⁶ On the other hand, since
9 molecular recognition is often achieved via relatively short amino acid segments located within
10 disordered regions, modification of such functions can be easily accomplished by AS affecting the
11 corresponding disordered regions. In this way, diverse functional profiles can be generated for the
12 transcribed alternatively spliced products of the same gene. Furthermore, since regulatory and
13 signaling elements in disordered regions can be comprised of just a few more or less continuous
14 amino acids, and since a high density of functionally important segments can be located within
15 disordered regions, functionality of IDPs can be completely rewired via AS.^{266,267}

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33 AS is known to occur in almost all human genes, and therefore alterations of this process are
34 intimately connected to the pathogenesis of various human diseases, ranging from cancer,^{268,269} to
35 neurodegenerative²⁷⁰⁻²⁷² and cardiovascular diseases.²⁷³ Many of the pathology-related proteins
36 affected by pathology-related AS are intrinsically disordered. For example, aberrant splicing events
37 are described for a number of familial cancer syndromes and affect the tumor suppressor genes such
38 as *BRCA1*, *APC*, *p53*, *FHIT*, and *LKB1*,^{269,274-276} many of which encode intrinsically disordered proteins.
39 One of the major players in Alzheimer's disease pathology, microtubule-associated protein tau, is a
40 typical IDP,²⁷⁷ the function and dysfunction of which is controlled via multiple AS events affecting its
41 gene *MAPT*.²⁷¹ Similarly, six genes, including *PARK2*, *SNCAIP*, *LRRK2*, *SNCA*, *SRRM2*, and *MAPT*, are
42 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

1
2
3 Finally, in a slightly different though related mechanism, the discovery of regulatory sites that can
4 enhance or dampen the interaction with a partner while being located outside the primary binding
5 site, provides an example illustrating the extreme allostery (i.e. the long-range nature of the effects
6 of amino acid substitutions) that typifies IDPs.²⁸⁵
7
8
9
10

11 12 13 14 **6.8. Multi-valent interactions and phase-separation of IDPs** 15 16

17
18
19 IDPs are often repetitive in nature,²⁸⁶ and are thus capable of mutual multi-valent binding which
20 results in their separation from solution state in a second-order phase transition generating a
21 hydrogel-like state. This phase transition might represent a novel and special aspect of the
22 functionality of IDPs.¹³³ The original observation was made by Li and co-workers,¹³² who studied the
23 interactions of IDPs with tandem copies of SH3 domains and SH3 interaction motifs, and showed that
24 the proteins engaged in multi-valent interactions producing sharp sol-gel phase transition and phase
25 separation. Not only is phase separation a physical phenomenon, but it also causes an increase in
26 actin-polymerization activity in the system made up of NCK–nephrin–N-WASP proteins. Further, the
27 sol-gel transition is controlled by the level and stoichiometry of the proteins and post-translational
28 modification, such as phosphorylation.
29
30

31
32 A similar physical phenomenon is induced by low-complexity regions in RNA-binding proteins, which
33 can form cellular RNA granules.²⁸⁷ This mechanism may be important in the physical regulation of
34 mRNA localization and translation, through reversible phase transitions with a highly dynamic
35 hydrogel state. In addition, hydrogels formed by one disordered RNA-binding protein can bind IDRs
36 of other members of the protein family, which, through the formation of heterogenous systems,
37 might drive regulatory cross-talk between distinct sub-systems.
38
39

40
41 Yet another system is represented by cytoskeletal components neurofilaments (NFs, intermediate
42 filaments in non-neuronal cells). NFs have three components of different molecular weight, all three
43 having a rod region that forms the body of the cytoskeletal structure, and intrinsically disordered
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 projection domains of variable length. These highly repetitive (low-complexity) disordered domains
4
5 can attract each-other and form transient cross-bridges supporting a liquid crystal-like hydrogel
6
7 structure.²⁸⁸ Depending on external pressure, the hydrogel has several stable structures, which may
8
9 have different effect on the flexibility (physical stiffness) of the cytoskeleton.

10
11 These and other cases¹³³ suggest an interesting novel regulatory mechanism in the cell by multivalent
12
13 IDPs. These proteins can generate an isolated physical phase of unique physical and biochemical
14
15 properties, which might also bridge diverse length scales in the cell stemming from macromolecules
16
17 to organelles.
18
19

20 21 22 **7. Cellular biology of IDPs. An overview** 23 24 25 26 27

28 **7.1. Abundance of intrinsic disorder in various proteomes** 29 30 31

32
33 Intrinsic disorder is abundant in nature, and many proteins are disordered along their entire length.
34
35 Indeed, predictions on representative genomes from the three kingdoms of life (i.e. bacteria, archaea
36
37 and eukaryotes) confirmed the prevalence of disorder although the disorder amount differs
38
39 significantly between them.^{82,289} Apparently, there is an overall increase of disorder with the
40
41 complexity of an organism.^{82,289,290} In agreement, earlier predictions revealed a high disorder
42
43 percentage for eukaryotes (with 30% of IDRs longer than 40 consecutive amino acids) compared to
44
45 bacteria (4%) and archaea (2%).²⁸⁹ This point is convincingly illustrated by differences between the
46
47 genome and the proteome of two model organisms, *S. cerevisiae* and *E. coli*,²⁹¹ which underscore a
48
49 “large jump in putatively disordered proteins in multi-celled, rather than single-celled, organisms”.⁸²
50
51

52
53
54 It is to be noted, though, that the level disorder in both prokaryotes²⁹² and eukaryotes²⁹³ varies
55
56 widely, and comparing grand averages is of limited information only. In both kingdoms the level of
57
58
59
60

1
2
3 disorder seems to represent a strong adaptive trait that shows correlation with the environment²⁹²
4
5 or pathogenic lifestyle.²⁹³ In all, the signaling functions linked with the complexity of the organism
6
7 seem to represent a major factor determining the general prevalence of disorder in an organism, on
8
9 which various adaptive changes reflecting the environment and lifestyle of the organism, operate.
10

11
12
13 It has been proposed that motifs embedded in disordered regions of viral proteins are employed by
14
15 viruses to interact with host proteins, in order to hijack and manipulate these proteins for their
16
17 purposes.²⁹⁴ In line with this suggestion, it was shown that viral proteins, and in particular RNA virus
18
19 proteins, are enriched in disordered regions.²⁹⁵ In that study, the authors propose that beyond
20
21 affording a broad partnership, the wide occurrence of disordered regions in viral proteins could also
22
23 be related to the typical high mutation rates of RNA viruses, representing a strategy for buffering the
24
25 deleterious effects of mutations.²⁹⁵ Taking into account these considerations, as well as the
26
27 correlation between overlapping genes and disorder,²⁹⁶⁻²⁹⁸ it has been proposed²⁹⁹⁻³⁰² that the main
28
29 advantage of the abundance of disorder within viruses would reside in pleiotropy and genetic
30
31 compaction. Indeed, disorder provides a solution to reduce both genome size and molecular
32
33 crowding, where a single gene would (i) encode a single (regulatory) protein product that can
34
35 establish multiple interactions via its disordered regions and hence exert multiple concomitant
36
37 biological effects, and/or (ii) would encode more than one product by means of overlapping reading
38
39 frames. In fact, since disordered regions are less sensitive to structural constraints than ordered
40
41 ones, the occurrence of disorder within one or both protein products encoded by an overlapping
42
43 reading frame can represent a strategy to alleviate evolutionary constraints imposed by the overlap.
44
45 As such, disorder would confer to viruses the ability to "handle" overlaps, thus further expanding the
46
47 coding potential of viral genomes.
48
49
50
51
52

53
54
55 In agreement with these predictions, and following the first and seminal observations that disorder is
56
57 abundant in proteins of the replicative complex of paramyxoviruses,³⁰³⁻³⁰⁵ an increasing number of
58
59
60

1
2
3 experimental evidence has been gathered in the last decade pointing out the abundance of disorder
4
5 in viral proteins (for reviews see refs.³⁰⁶⁻³⁰⁹). In fact, a recent bioinformatics analysis of the abundance
6
7 of intrinsic disorder in various proteomes from viruses and the three domains of life revealed that
8
9 viruses are characterized by the widest spread of proteome disorder content, with the percentage of
10
11 disordered residues ranging from 7.3% in human coronavirus NL63 to 77.3% in Avian carcinoma
12
13 virus.³⁰⁹
14
15
16
17
18

19 **7.2. Involvement of structural disorder in biological processes**

20
21
22

23
24 As suggested above, there are many functional categories that are positively or negatively correlated
25
26 with structural disorder. In a comprehensive bioinformatics study carried out by Xie et al.^{249,251} a
27
28 positive correlation between the functional annotation of the SwissProt database and predicted
29
30 intrinsic disorder has been found (238 out of 710 SwissProt functional keywords), whereas a negative
31
32 correlation was found in 310 out of 710 functional keywords (the remainder 162 functional keywords
33
34 yielded ambiguity).
35
36
37

38
39 In some detail, the major functional categories which are significantly enriched in structural disorder
40
41 are: differentiation, transcription, transcription regulation, spermatogenesis, DNA condensation, cell
42
43 cycle, mRNA processing (splicing), mitosis, apoptosis, protein transport and meiosis, whereas major
44
45 categories depleted in structural disorder are: GMP biosynthesis, amino acid biosynthesis, transport,
46
47 electron transport, lipid biosynthesis, aromatic hydrocarbon catabolism, glycolysis, purine and
48
49 pyrimidine biosynthesis, carbohydrate metabolism, and biosynthesis of branched-chain amino acids
50
51 and lipopolysaccharide biosynthesis. These preferences are in line with a large-scale, unbiased
52
53 analysis of the proteome in mouse embryo fibroblasts that were under- and over-represented in
54
55 heat-stable proteins, followed by a rigorous analysis of GO terms (biological process, molecular
56
57 function, and cellular localization).³¹⁰ In all, the structural preferences of these categories underscore
58
59
60

1
2
3 the old wisdom in the IDP field that structural disorder is enriched in proteins involved in signaling
4
5 and regulatory functions and depleted in enzymes.³¹¹
6
7
8
9

10 11 **8. Distinctive features of IDP evolution** 12 13

14
15
16 Tracking the evolutionary history of a protein is very closely related to understanding the molecular
17
18 basis of its function. As it was already mentioned (section 7.1), the level of disorder increases with
19
20 the complexity of organisms,^{8,289,312} and it can undergo rapid changes in adaptation to external
21
22 conditions.^{292,293} These rapid evolutionary changes raise the question as to how genes encoding for
23
24 IDPs can arise. There are several possible mechanisms, such as *de novo* generation (i.e. from random
25
26 pieces of genomic DNA), lateral gene transfer and horizontal gene transfer.⁴² A possible mechanism
27
28 that has been confirmed through several examples in the literature is the generation of disordered
29
30 domains by gene duplication and module exchange (i.e. attaching a disordered domain to an already
31
32 existing protein). Indeed, many protein families exist that have common disordered domains, as for
33
34 instance the kinase inhibitory domain (KID) of Cdk inhibitors p21^{Cip1}, p27^{Kip1}, and p57^{Kip2},³¹³ the
35
36 catenin-binding domain (CBD) in T-cell factors LEF-1, Tcf3, and Tcf4,³¹⁴ and the WASP-homology
37
38 domain 2 (WH2) in actin-binding proteins thymosin- β 4, ciboulot, and WASP³¹⁵, as outlined in⁴² The
39
40 possible generality of disorder spreading by domain duplications and exchange between genes is also
41
42 underscored by the observation that about 14% of all Pfam domains are predicted to be mostly
43
44 disordered.²⁰³
45
46
47
48
49

50
51 Another possible mechanism is the expansion of internal repeat regions (i.e. regions coding micro
52
53 and mini-satellites). Indeed, an analysis of 126 known IDPs showed that the percentage of proteins
54
55 with tandemly repeated segments is much higher in IDPs (39%) than in SwissProt (14%), or human
56
57 (28%) proteins²⁸⁶ and short sequence repeats often occur in IDPs.²⁴ The genetic instability of
58
59
60

1
2
3 repetitive regions as witnessed by large interspecies variability and polymorphism, combined with
4
5 their frequent involvement in the function of IDPs, confers to IDPs an exceptional functional
6
7 variability and might represent a prevalent mechanism of the evolutionary spread of disorder^{239,289,316}
8
9 and/or rapid adaptive genomic sweeps.^{292,293}
10

11
12
13
14 By aligning sequences in which a globular domain and a disordered region are present, one could
15
16 compare the rates of their evolutionary change. Accordingly, Celeste Brown and colleagues,³¹⁷
17
18 performed a sequence alignment of homologs of 26 such families and calculated their pairwise
19
20 genetic distances, confirming a high evolutionary rate of IDPs with respect to globular ones in 19
21
22 families (with similar rates in 5 families and slower rates in 2 families).³¹⁷ The likely explanation for
23
24 the generally faster rates of evolution in IDPs is the lack of structural constraints (i.e. lack of a fixed 3-
25
26 D structure) that enables many substitutions.³¹⁷
27
28
29
30

31
32 Despite the general tendency of IDPs to evolve rapidly, certain IDPs are rather resistant to
33
34 evolutionary changes and hence they are conserved. Those regions were termed conserved disorder
35
36 predictions (CDPs) and are usually short. CDPs can be found in all kingdoms of life, but long ones are
37
38 almost 10 times more frequent in viruses and eukaryotes than in bacteria and archae.²⁰⁵ In a recent
39
40 comparative analysis, disorder was found to segregate into three distinct behavioral categories,
41
42 flexible disorder (conserved disorder with quickly evolving sequences), constrained disorder
43
44 (conserved disorder with conserved amino acid sequence, such as ribosomal proteins) and non-
45
46 conserved disorder.³¹⁸
47
48
49

50
51 These findings have to be elucidated in terms of the subtle balance between the disadvantage of the
52
53 effect of a mutation on impairing activity/function, and the activity of change in generating evolution
54
55 variability.^{319,320} In other words, evolutionary variability provides the raw material for selection
56
57 among functional variants, but too much variability works against retention of function. In general, in
58
59
60

1
2
3 the case of functions that stem directly from the disordered state (i.e. entropic chains), the amino
4 acid sequences are rather resistant to mutations as long as the mutations do not bring a major
5 transition in the conformational ensemble. This was formally demonstrated in the case of replication
6 protein A (RPA70), which contains two domains and a disordered linker. It was shown by NMR that
7 the dynamics of the linker, despite great evolution variability, is preserved as a key functional
8 feature.³²¹

9
10
11
12
13
14
15
16
17
18 In the case of molecular recognition functions, which require motifs of specificity-determinant
19 residues interspersed in highly disordered regions (see section 6.2),¹⁹¹ conservation poses a serious
20 challenge with respect to the generally high sequence variability observed in IDPs. Apparently, only
21 the very little fraction of direct recognition residues is subjected to evolutionary constraints. The
22 same applies to PTM sites, changes of which might have a deleterious effect on signaling and
23 regulation events.²³⁹ In agreement, Chen et al.³²² showed that phosphorylated Ser and Thr residues
24 tend to evolve more slowly than the non-phosphorylated ones.

35 **9. Intrinsic disorder and human diseases**

40 **9.1. Correlation of disorder and diseases**

41
42
43
44
45 A general correlation between intrinsic disorder and various diseases, such as cancer, diabetes,
46 amyloidoses, neurodegenerative and cardiovascular diseases, is supported by several bioinformatics
47 analyses (Fig. 20) and by detailed studies on individual proteins (see also section 6.6). These findings
48 suggest that protein conformational diseases may result not only from protein misfolding, but also
49 from misidentification, missignaling, and unnatural or non-native folding, which lead to the D^2
50 concept, i.e. disorder in disorders.³²³

1
2
3
4
5 In fact, many human diseases known as protein conformational or protein-misfolding diseases arise
6
7 from the failure of a specific peptide or protein to adopt its final conformational state (i.e. protein
8
9 misfolding), which leads to protein aggregation and/or fibril formation, loss of normal function, and
10
11 gain of toxic function. An increase in the propensity to misfold can be the result of interactions with
12
13 endogenous factors, such as chaperones, intracellular and extracellular matrixes, other proteins and
14
15 small molecules, or point mutations, or also impaired PTMs.³²³ The conversion of specific proteins
16
17 from their soluble functional states into stable, highly ordered, filamentous protein aggregates (i.e.
18
19 amyloid fibrils) and the deposition of these aggregated materials into a variety of tissues and organs
20
21 lead to a high number of misfolding diseases (as for instance, neurodegenerative diseases like
22
23 Parkinson and Alzheimer diseases). The fact that IDPs are key players in the development of such
24
25 pathological conditions is also in agreement with their involvement in key cellular functions (i.e.
26
27 signaling and regulation).³²⁴ It should be noted, however that the capacity to form fibrils is not
28
29 unique to IDPs: globular proteins can also form such aggregates (transthyretin for instance³²⁵) and it
30
31 has been suggested that practically all proteins have the capacity to form fibrils.³²⁶
32
33
34
35
36

37
38 Based on the high abundance of IDPs in all kingdoms of life, their important functions and their
39
40 ability to undergo a misfolding (i.e. amyloids), a novel view of the fate of a polypeptide chain inside
41
42 the cell has been reported by Uversky (see Fig. 21).³²⁷ A polypeptide chain can either fold to gain a
43
44 unique structure necessary for catalytic and transport activities, it can stay substantially non-folded,
45
46 or can misfold and, under some circumstances, can form amyloid-like fibrils.
47
48
49
50

51 **9.2. IDPs as drug targets**

52
53
54
55

56 Protein-protein interactions are attractive targets for drug development. If the 3-D structure of the
57
58 complex is available, small molecules can be designed that will bind to interfaces and interfere with
59
60

1
2
3 protein-protein interactions. The low affinity of molecules to their targets and their low specificity
4
5 towards specific interfaces (i.e. all active sites of proteases are similar), however, often limits the
6
7 biological use of these molecules.¹²⁹
8
9

10
11 In this regard, protein-protein interactions involving IDPs are of great interest to develop new
12
13 strategies for drug design. The “druggability” of IDPs is based on different observations: 1) the lack
14
15 of flat surfaces in IDP-complexes which are typically found in interactions implying globular proteins
16
17 2) the weakness of interactions involving IDPs due to the use of energy in organizing the disordered
18
19 partner thus facilitating tighter binding by a small drug, 3) the frequent involvement of hydrophobic
20
21 clusters in folding coupled to binding and the involvement of only one (or few) contiguous segments
22
23 making the interaction area small and compact, and 4) the ease of mimicking MoREs by appropriate
24
25 small molecules.³²⁸
26
27
28
29
30

31 In IDP-partner interactions, drugs can be designed either to target the binding site of the partner
32
33 through mimicking the disordered motif,³²⁸ or the IDP itself. As an example of the first approach,
34
35 nutlins have been identified to bind Mdm2 and hence block the binding of a disordered region of
36
37 p53.^{329,330} These compounds have the capacity to stabilize and activate the p53 pathway in cancer
38
39 cells, leading to cell-cycle arrest, apoptosis, and growth inhibition of human tumor xenografts in
40
41 nude mice.^{329,330}
42
43
44
45

46 In the inverse scenario, inhibitors bind to the disordered regions and interfere with their interactions.
47
48 c-Myc-Max complex provides an interesting example. c-Myc is a transcription factor involved in many
49
50 types of cancer. In order to bind to its targets, c-Myc must dimerize with Max by mutual folding
51
52 when their zipper domains interact to form a helical coiled-coil. Several inhibitors were found
53
54 through high throughput screening which are able to bind to a disordered region within the
55
56
57
58
59
60

1
2
3 monomeric c-Myc. Inhibitor binding induces only local conformational changes, preserves the overall
4
5 disorder of c-Myc, and inhibits dimerization with Max.^{129,331}
6
7
8
9
10

11 Abbreviations

14	1-D	one-dimensional
15		
16	3-D	three-dimensional
17		
18	AFM	atomic force microscopy
19		
20	appMM	apparent molecular mass
21		
22	AS	alternative splicing
23		
24	CD	circular dichroism
25		
26	CDF	cumulative distribution function
27		
28	CH-plot	charge-hydrophathy plot
29		
30	CDP	conserved disorder prediction
31		
32	CSI	chemical shift index
33		
34	DLS	dynamic light scattering
35		
36	EAM	ensemble allosteric model
37		
38	ELM	eukaryotic linear motif
39		
40	EM	electron microscopy
41		
42	ES	enzyme-substrate
43		
44	FRET	fluorescence resonance energy transfer
45		
46	FS	force spectroscopy
47		
48	GF	gel filtration
49		
50	HCA	hydrophobic clusters analysis
51		
52	HS	high speed
53		
54	HSQC	heteronuclear single quantum coherence
55		
56		
57		
58		
59		
60		

1		
2		
3	IDP	intrinsically disordered protein
4		
5	IDR	intrinsically disordered region
6		
7	KID	kinase inhibitory domain
8		
9	LC8	dynein light chain
10		
11	LM	linear motif
12		
13	MAP2	microtubule-associated protein 2
14		
15	MD	molecular dynamics
16		
17	MeV	measles virus
18		
19		
20	MG	molten globule
21		
22	MLA	machine-learning algorithms
23		
24	MoRE	molecular recognition élément
25		
26	MoRF	molecular recognition feature
27		
28		
29	MW	molecular-weight
30		
31	NMR	nuclear magnetic résonance
32		
33	NOE	nuclear Overhauser effect
34		
35	N-WASP	actin-regulatory neural Wiskott-Aldrich syndrome protein
36		
37	N _{TAIL}	C-terminal domain of measles virus nucleoprotein
38		
39	PDB	protein data bank
40		
41	pE-DB	protein ensemble database
42		
43		
44	PCS	primary contact site
45		
46	PMG	pre-molten globule
47		
48	PRE	paramagnetic relaxation enhancement
49		
50	PreSMo	pre-structured motif
51		
52	PSE	preformed structural élément
53		
54		
55	PTM	posttranslational modification
56		
57	RC	random coil
58		
59		
60		

1		
2		
3	RDC	residual dipolar coupling
4		
5	RPA70	replication protein A
6		
7	SAXS	small angle X-ray scattering
8		
9	SCS	secondary chemical shift
10		
11	SDS	sodium dodecyl sulfate
12		
13	SDS-PAGE	SDS polyacrylamide gel electrophoresis
14		
15	SEC	size-exclusion chromatography
16		
17	SLiM	short linear motif
18		
19		
20	SM	single-molecule
21		
22	TFE	trifluoroethanol
23		
24	TMAO	trimethylamine N-oxide
25		
26		
27	XD	X domain
28		
29		
30		

31 Acknowledgements

32
33
34
35 This work was carried out with the financial support of the Odysseus grant G.0029.12 from Research
36 Foundation Flanders (FWO) to P.T., and of the Agence Nationale de la Recherche, specific programs
37 "Physico-Chimie du Vivant", ANR-08-PCVI-0020-01, and "ASTRID", ANR-11-ASTR-003-01, to S.L.. The
38 work was also supported in part by the Russian Academy of Sciences "Molecular and Cellular
39 Biology" Program to V.N.U. The funders had no role in study design, data collection and analysis,
40 decision to publish, or preparation of the manuscript.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 Figure 5. Binary predictors of intrinsic disorder. (A) CDF analysis. The dark blue curve located above
4 the boundary represents the CDF curve of an ordered protein (β -glycosidase, UniProt ID : P07986),
5 whereas the dark red line located below the boundary corresponds to the CDF curve of an IDP
6 (secretogranin-1, UniProt ID: P23389). Here, δ_i and d_j (where i and j range from 1 to 7) are attributed
7 to the ordered and disordered protein, respectively, and represent the distances of points at the CDF
8 curve from the corresponding boundary points. The averaged distance of a given CDF curve from a
9 boundary line is calculated as $\Delta\text{CDF} = \frac{\sum_{i=1}^7 \delta_i}{7}$ or $\Delta\text{CDF} = \frac{\sum_{j=1}^7 d_j}{7}$. (B) CH-plot analysis. The dark red circle
10 located above the boundary corresponds to the same IDP as in (A), namely secretogranin-1 (UniProt
11 ID: P23389), the dark blue square located below the boundary represents the same ordered protein
12 as in (A), namely β -glycosidase (UniProt ID: P07986). (C) CH-CDF analysis. The dark red circle
13 corresponds to the disordered protein, secretogranin-1 (UniProt ID: P23389), whereas the dark blue
14 square represents the ordered protein β -glycosidase (UniProt ID: P07986). X-coordinates were
15 calculated as averaged distances of corresponding CDF curves from a boundary (a positive ΔCDF
16 distance corresponds to a protein predicted to be ordered by CDF analysis, while a negative ΔCDF
17 distance corresponds to a protein predicted to be disordered by CDF analysis, see plot A). Y-
18 coordinates were obtained as distances from spots corresponding to proteins to the boundary.
19 Positive and negative ΔCH distances correspond to proteins predicted by the CH-plot to be
20 disordered or ordered, respectively (see plot B). (D) CH-CDF plot for ribosomal proteins. Quadrants
21 contain differently disordered proteins: the light red quadrant contains extended IDPs (predicted to
22 be disordered by CDF and CH-plot analysis), the light pink quadrant contains native molten globules
23 and/or hybrid proteins containing ordered and disordered regions (predicted to be disordered by
24 CDF and ordered by CH-plot), the light blue quadrant contains globular proteins (predicted to be
25 ordered by both CDF and CH-plot analyses), whereas the light cyan quadrant contains proteins
26 predicted to be ordered by CDF and disordered by CH-plot.

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Figure 6. Quantitative ensemble description of an IDP. (A) Carton representation of four
4 conformational states of the α -MoRE of the intrinsically disordered C-terminal domain of the measles
5 virus nucleoprotein (N_{TAIL}), as derived by NMR spectroscopy through RDCs measurements. Modified
6 from ref.⁶² (B) Secondary chemical shift differences values for the $^{13}C\alpha$ atoms of N_{TAIL} alone (blue
7 bars) and in complex with the α -helical C-terminal domain of the phosphoprotein (P_{XD} ; red bars) with
8 respect to a random coil chemical shift standard. In the free form, the values for the region
9 encompassing the residues 90-110 (red bars) are shifted downfield (to larger ppm values) indicating a
10 transiently populated α -helix in this region that is stabilized in the presence of P_{XD} (blue bars).
11 Modified from ref.⁵⁸ (C) Comparison of experimental (blue) and back-calculated (red) DH-NH RDCs
12 from the model of N_{TAIL} shown in (A). Adapted with permission from ref.⁶² Copyright 2011 National
13 Academy of Sciences.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 Figure 7. NMR PRE measurements in α -synuclein. Difference between the measured peak intensities
30 (black bars) and the paramagnetic effect expected for a random coil (red line) suggests long-range N-
31 to C-terminal tertiary structures in agreement with the observed broadening of the peak intensities.
32 Adapted with permission from ref.³³² Copyright 2007 Elsevier.
33
34
35
36
37
38

39 Figure 8. NMR-based model allowing positioning of the disordered N_{TAIL} region of the measles virus
40 nucleoprotein within the viral nucleocapsid. Front (A) and top (B) views of 13 N_{TAIL} conformers
41 sampling an ensemble of conformations that point out from the surface of the viral nucleocapsid.
42 Adapted with permission from ref.⁶² Copyright 2011 National Academy of Sciences.
43
44
45
46
47
48
49

50 Figure 9. The protein quartet model of protein function. Function can arise from four different
51 conformations of the polypeptide chain, or from transitions between any of the states. Reprinted
52 with permission from ref.¹³ Copyright 2002 John Wiley & Sons, Inc.
53
54
55
56
57
58
59
60

1
2
3 Figure 10. Mechanisms to prevent IDPs from degradation by default. IDPs can be stabilized by
4 intermolecular interactions (A), interactions with gatekeepers (B), nanny proteins (C), DNA binding-
5 sites (D), “decoy” DNA binding-sites (E), intramolecular interactions (F), local folding at the N-
6 terminal region (G) or interactions with ribosome or ribosome-associated proteins (H). Reprinted
7 with permission from ref.¹¹⁷ Copyright 2011 John Wiley & Sons, Inc.
8
9
10
11
12

13
14
15
16 Figure 11. Continuum of disorder. Functional disordered segments can be as small as only a few
17 amino acids residues, or they can occupy rather long regions or ends. Different levels of order and
18 disorder. From left to right, no disorder; disordered N- and C-termini; disordered linker; disordered
19 loop; disordered domain; disordered protein with some residual structure; wholly disordered, mostly
20 collapsed protein; wholly disordered, extended protein. Corresponding disordered regions are shown
21 in red. Adapted with permission from ref.¹¹⁹ Copyright 2005 John Wiley & Sons, Inc.
22
23
24
25
26
27
28

29
30
31 Figure 12. Conformational behavior of proteins. Schematic representation of the physical principles
32 of molecule separation by size-exclusion chromatography. The porous column matrix is shown by
33 gray spheres. Small and large molecules are shown as pink and yellow spheres respectively. A vertical
34 arrow on the left indicate the migration direction through the SEC matrix. Large molecules migrate
35 faster than small molecules. At the right, relative hydrodynamic volumes occupied by a long
36 polypeptide chain in four different conformations: RC: random coil, PMG: premolten globule, MG:
37 molten globule and folded. Spheres show an increase in the Stokes radius when a protein possessing
38 the same length is progressively unfolded.
39
40
41
42
43
44
45
46
47
48

49
50 Figure 13. Spectroscopic features determined by CD. (A) CD spectra representing different protein
51 conformations are shown (α -helix, β -sheet and RC). RC-like conformations could be distinguished
52 from secondary structures by a large negative ellipticity at 290 nm and an ellipticity close to 0 at
53 222nm. A negative ellipticity at 222 nm reveals the presence of transient secondary structures.
54
55
56
57
58
59
60

1
2
3 Indeed, analysis of far-UV CD spectra in terms of double wavelength plot, $[\Theta]_{222}$ versus $[\Theta]_{200}$, (B),
4
5 allows IDPs to be subdivided in coil-like (green diamonds) and PMG-like subclasses (blue circles).¹³
6
7

8
9 Figure 14. Distinctive properties of IDPs: diversity and functional role.
10
11

12
13 Figure 15. Examples of structurally divergent MoREs. MoREs (red ribbons) and partners (green
14 surfaces) are shown. (A) An α -MoRF, Proteinase Inhibitor IA3, bound to Proteinase A (PDB entry
15 1DP5). (B) A β -MoRF, viral protein pVIc, bound to Human Adenovirus 2 Proteinase (PDB entry 1AVP).
16
17 (C) An i-MoRF, Amphiphysin, bound to α -adaptin C (PDB entry 1KY7). (D) A complex-MoRF, β -amyloid
18 precursor protein (β APP), bound to the PTB domain of the neuron specific protein X11 (PDB entry
19 1X11). Partner interfaces (gray surfaces) are also indicated. Reprinted with permission from ref.¹⁷
20
21
22
23
24
25

26 Copyright 2008 Springer.
27
28

29
30
31 Figure 16. Adaptability of IDPs to their partners. A chameleon sequence (the C-terminal region of
32 p53) is shown in complex with four different partners and exhibiting four different structural forms.
33
34 Reprinted with permission from ref.²¹³ Copyright 2008 Springer.
35
36
37

38
39
40 Figure 17. Schematic representation of coupled and decoupled affinity. For the binding of two
41 globular proteins, affinity and specificity are expected to be proportional (Upper). When an induced
42 folding occurs in case of IDPs (as for instance α -MoREs), affinity and specificity are decoupled due to
43 the entropic cost of the folding.²⁰
44
45
46
47

48
49
50 Figure 18. Proposed mechanism by Espinoza-Fonseca for the synergistic model. Left: once the IDP is
51 close enough to its target via “fly-casting”, conformer selection comes into play in that the target
52 chooses a specific conformational state of the preformed region (oval) among the populated ones.
53
54
55
56
57
58
59
60

1
2
3 Right: folding coupled to binding constitutes the final step where the MoRE is stabilized by the
4
5 partner in the fully bound complex. Reprinted with permission from ref.²²⁶ Copyright 2009 Elsevier.
6
7

8
9 Figure 19. Fuzzy complexes. Static (A) and dynamic (B-D) disorder in protein-protein interactions are
10
11 represented with the arrow indicating increasing disorder. The binding partner is rendered as a solid
12
13 grey surface whereas the colored ribbon corresponds to the IDP in complex possessing residual
14
15 disorder (dotted line). In the static model, there is more than one stable conformation whereas in
16
17 the dynamic model parts or the entirety of the IDP remain disordered in the bound form. Reprinted
18
19 with permission from ref.²³³ Copyright 2008 Elsevier Trends Journals.
20
21
22
23

24 Figure 20. Abundance of intrinsic disorder in disease-associated proteins. Percentages of disease-
25
26 associated proteins with ≥ 30 to ≥ 100 consecutive residues predicted to be disordered.
27
28 Corresponding data for signaling and ordered proteins (i.e. PDB_S25) are shown for the comparison.
29
30 Analyzed sets of disease-related proteins included 1786, 487, 689, and 285 proteins for cancer, CVD,
31
32 neurodegenerative disease and diabetes, respectively. Data were taken from ref.³²³
33
34
35
36

37 Figure 21. Folding, nonfolding or misfolding of a polypeptide chain. A polypeptide chain can either
38
39 gain a well-defined 3-D structure, or can stay non-folded in the three different states (i.e. RC, PMG
40
41 and MG), or also can misfold to form amyloid-like fibrils.³³³
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

- 1
2
3 (15) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* 2000, 41, 415.
4
5 (16) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.;
6
7 Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.;
8
9 Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.;
10
11 Garner, E. C.; Obradovic, Z. *J. Mol. Graph. Model.* 2001, 19, 26.
12
13 (17) Dunker, A. K.; Oldfield, C. J.; Meng, J.; Romero, P.; Yang, J. Y.; Chen, J. W.; Vacic, V.;
14
15 Obradovic, Z.; Uversky, V. N. *BMC Genomics* 2008, 9 Suppl 2, S1.
16
17 (18) Campen, A.; Williams, R. M.; Brown, C. J.; Meng, J.; Uversky, V. N.; Dunker, A. K.
18
19 *Protein Pept. Lett.* 2008, 15, 956.
20
21 (19) Williams, R. M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Young, J.;
22
23 Takayama, S.; Brown, C. J.; Dunker, A. K. *Pac. Symp. Biocomput.* 2001, 89.
24
25 (20) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K.
26
27 *Biochemistry* 2005, 44, 1989.
28
29 (21) Csizmok, V.; Szollosi, E.; Friedrich, P.; Tompa, P. *Mol. Cell. Proteomics* 2006, 5, 265.
30
31 (22) Romero, P.; Obradovic, Z.; Dunker, A. K. *FEBS Lett.* 1999, 462, 363.
32
33 (23) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins*
34
35 2001, 42, 38.
36
37 (24) Lise, S.; Jones, D. T. *Proteins* 2005, 58, 144.
38
39 (25) Longhi, S. *Methods Molecular Biology* 2010, 609, 307.
40
41 (26) Ferron, F.; Longhi, S.; Canard, B.; Karlin, D. *Proteins* 2006, 65, 1.
42
43 (27) Lieutaud, P.; Ferron, F.; Habchi, J.; Canard, B.; Longhi, S., *Advances in Protein and*
44
45 *Peptide Sciences*, Dunn, B., Ed., Bentham Science Publishers (eBook), 2013, pp
46
47 441-492.
48
49 (28) Bourhis, J. M.; Canard, B.; Longhi, S. *Curr. Protein Pept. Sci.* 2007, 8, 135.
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (29) Uversky, V. N.; Radivojac, P.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K.
4
5 *Methods Mol. Biol.* 2007, 408, 69.
6
7
8 (30) Prilusky, J.; Felder, C. E.; Zeev-Ben-Mordehai, T.; Rydberg, E. H.; Man, O.;
9
10 Beckmann, J. S.; Silman, I.; Sussman, J. L. *Bioinformatics* 2005, 21, 3435.
11
12 (31) Coeytaux, K.; Poupon, A. *Bioinformatics* 2005, 21, 1891.
13
14 (32) Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J. *Nucleic Acids Res.* 2003, 31,
15
16 3701.
17
18 (33) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo,
19
20 B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids*
21
22 *Res.* 2007, 35, D786.
23
24
25 (34) Fukuchi, S.; Sakamoto, S.; Nobe, Y.; Murakami, S. D.; Amemiya, T.; Hosoda, K.;
26
27 Koike, R.; Hiroaki, H.; Ota, M. *Nucleic Acids Res.* 2012, 40, D507.
28
29
30 (35) Di Domenico, T.; Walsh, I.; Martin, A. J.; Tosatto, S. C. *Bioinformatics* 2012, 28,
31
32 2080.
33
34
35 (36) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. *Structure*
36
37 *(Camb)* 2003, 11, 1453.
38
39
40 (37) Ward, J. J.; McGuffin, L. J.; Bryson, K.; Buxton, B. F.; Jones, D. T. *Bioinformatics* 2004,
41
42 20, 2138.
43
44
45 (38) Yang, Z. R.; Thomson, R.; McNeil, P.; Esnouf, R. M. *Bioinformatics* 2005, 21, 3369.
46
47 (39) Garbuzynskiy, S. O.; Lobanov, M. Y.; Galzitskaya, O. V. *Protein Sci.* 2004, 13, 2871.
48
49 (40) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. *Bioinformatics* 2005, 21, 3433.
50
51 (41) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. *J. Mol. Biol.* 2005, 347, 827.
52
53 (42) Tompa, P., *Structure and Function of Intrinsically Disordered Proteins*, CRC Press,
54
55 Taylor & Francis Group: Boca Raton, Florida, 2010.
56
57
58
59
60

- 1
2
3 (43) Callebaut, I.; Labesse, G.; Durand, P.; Poupon, A.; Canard, L.; Chomilier, J.;
4
5 Henrissat, B.; Mornon, J. P. *Cell. Mol. Life Sci.* 1997, 53, 621.
6
7
8 (44) Xue, B.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. *FEBS Lett.* 2009, 583, 1469.
9
10 (45) Mohan, A.; Sullivan, W. J., Jr.; Radivojac, P.; Dunker, A. K.; Uversky, V. N. *Mol.*
11
12 *Biosyst.* 2008, 4, 328.
13
14 (46) Huang, F.; Oldfield, C.; Meng, J.; Hsu, W. L.; Xue, B.; Uversky, V. N.; Romero, P.;
15
16 Dunker, A. K. *Pac. Symp. Biocomput.* 2012, 128.
17
18 (47) Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N. *Biochim.*
19
20 *Biophys. Acta* 2010, 1804, 996.
21
22
23 (48) Lieutaud, P.; Canard, B.; Longhi, S. *BMC Genomics* 2008, 9, S25.
24
25 (49) Blocquel, D.; Habchi, J.; Gruet, A.; Blangy, S.; Longhi, S. *Mol. Biosyst.* 2012, 8, 392.
26
27
28 (50) Kosol, S.; Contreras-Martos, S.; Cedeno, C.; Tompa, P. *Molecules* 2013, 18, 10802.
29
30
31 (51) Dyson, H. J.; Wright, P. E. *Chem. Rev.* 2004, 104, 3607.
32
33 (52) Bermel, W.; Bertini, I.; Chill, J.; Felli, I. C.; Haba, N.; Kumar, M. V. V.; Pierattelli, R.
34
35 *ChemBioChem* 2012, 13, 2425.
36
37 (53) Receveur-Bréchet, V.; Bourhis, J. M.; Uversky, V. N.; Canard, B.; Longhi, S. *Proteins*
38
39 2006, 62, 24.
40
41
42 (54) Eliezer, D. *Curr. Opin. Struct. Biol.* 2009, 19, 23.
43
44
45 (55) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* 2008, 128, 052204.
46
47 (56) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. *Biochemistry* 2012,
48
49 51, 2224.
50
51 (57) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* 2010, 132,
52
53 1270.
54
55
56
57
58
59
60

- 1
2
3 (58) Gely, S.; Lowry, D. F.; Bernard, C.; Ringkjøbing-Jensen, M.; Blackledge, M.;
4
5 Costanzo, S.; Darbon, H.; Daughdrill, G. W.; Longhi, S. *J. Mol. Recognit.* 2010, 23,
6
7 435.
8
9
10 (59) Jensen, M. R.; Markwick, P. R.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek,
11
12 S.; Bernado, P.; Blackledge, M. *Structure* 2009, 17, 1169.
13
14 (60) Dyson, H. J.; Wright, P. E. *Adv. Protein Chem.* 2002, 62, 311.
15
16
17 (61) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.;
18
19 Blackledge, M. *J. Am. Chem. Soc.* 2010, 132, 8407.
20
21 (62) Jensen, M. R.; Communie, G.; Ribeiro, E. A., Jr.; Martinez, N.; Desfosses, A.; Salmon,
22
23 L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; Ruigrok, R. W.; Blackledge, M. *Proc.*
24
25 *Natl. Acad. Sci. U.S.A.* 2011, 108, 9839.
26
27
28 (63) Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* 2009,
29
30 131, 18314.
31
32
33 (64) Fisher, C. K.; Stultz, C. M. *Curr. Opin. Struct. Biol.* 2011, 21, 426.
34
35
36 (65) Belle, V.; Rouger, S.; Costanzo, S.; Longhi, S.; Fournel, A., *Instrumental analysis of*
37
38 *intrinsically disordered proteins: assessing structure and conformation*, Uversky, V.
39
40 N., Longhi, S., Eds., John Wiley and Sons: Hoboken, New Jersey 2010.
41
42 (66) Habchi, J.; Martinho, M.; Gruet, A.; Guigliarelli, B.; Longhi, S.; Belle, V. *Methods Mol.*
43
44 *Biol.* 2012, 895, 361.
45
46
47 (67) Drescher, M. *Top. Curr. Chem.* 2012, 321, 91.
48
49 (68) Lorenzi, M.; Puppo, C.; Lebrun, R.; Lignon, S.; Roubaud, V.; Martinho, M.; Mileo, E.;
50
51 Tordo, P.; Marque, S. R.; Gontero, B.; Guigliarelli, B.; Belle, V. *Angew. Chem. Int. Ed.*
52
53 *Engl.* 2011, 50, 9108.
54
55
56
57
58
59
60

- 1
2
3 (69) Mileo, E.; Etienne, E.; Martinho, M.; Lebrun, R.; Roubaud, V.; Tordo, P.; Gontero, B.;
4
5 Guigliarelli, B.; Marque, S. R.; Belle, V. *Bioconjug. Chem.* 2013, 24, 1110.
6
7 (70) Wang, J.; Verkhivker, G. M. *Phys. Rev. Lett.* 2003, 90, 188101.
8
9 (71) Chu, X.; Gan, L.; Wang, E.; Wang, J. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, E2342.
10
11 (72) Wang, J.; Xu, L.; Wang, E. *Biophys. J.* 2007, 92, L109.
12
13 (73) Wang, Y.; Chu, X.; Longhi, S.; Roche, P.; Wang, E.; Wang, J. *Proc. Natl. Acad. Sci.*
14
15 *U.S.A.* 2013, in press.
16
17 (74) Bernado, P.; Svergun, D., *Instrumental Analysis of Intrinsically Disordered Proteins:*
18
19 *Assessing Structure and Conformation*, Uversky, V. N., Longhi, S., Eds., John Wiley
20
21 and Sons: Hoboken, New Jersey USA, 2010, pp 451-476.
22
23 (75) Ando, T.; Kodera, N. *Methods Mol. Biol.* 2012, 896, 57.
24
25 (76) Miyagi, A.; Tsunaka, Y.; Uchihashi, T.; Mayanagi, K.; Hirose, S.; Morikawa, K.; Ando,
26
27 T. *Chemphyschem* 2008, 9, 1859.
28
29 (77) Sandal, M.; Valle, F.; Tessari, I.; Mammi, S.; Bergantino, E.; Musiani, F.; Brucale, M.;
30
31 Bubacco, L.; Samori, B. *PLoS Biol.* 2008, 6, e6.
32
33 (78) Ferreon, A. C.; Moran, C. R.; Gambin, Y.; Deniz, A. A. *Methods Enzymol.* 2010, 472,
34
35 179.
36
37 (79) Ferreon, A. C.; Ferreon, J. C.; Wright, P. E.; Deniz, A. A. *Nature* 2013, 498, 390.
38
39 (80) Milles, S.; Koehler, C.; Gambin, Y.; Deniz, A. A.; Lemke, E. A. *Mol. Biosyst.* 2012, 8,
40
41 2531.
42
43 (81) Ando, T.; Uchihashi, T.; Kodera, N. *Annu. Rev. Biophys.* 2013, 42, 393.
44
45 (82) Dunker, A. K.; Obradovic, Z. *Nat. Biotechnol.* 2001, 19, 805.
46
47 (83) Greenblatt, J.; Li, J. *J. Biol. Chem.* 1982, 257, 362.
48
49 (84) Lisse, T.; Bartels, D.; Kalbitzer, H. R.; Jaenicke, R. *Biol. Chem.* 1996, 377, 555.
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (85) Kriwacki, R. W.; Wu, J.; Siuzdak, G.; Wright, P. E. *J. Am. Chem. Soc.* 1996, 118, 5320.
4
5 (86) Kriwacki, R. W.; Wu, J.; Tennant, L.; Wright, P. E.; Siuzdak, G. *J. Chromatogr. A*
6
7 1997, 777, 23.
8
9
10 (87) Fontana, A.; Polverino de Laureto, P.; Spolaore, B.; Frare, E.; Zambonin, M.,
11
12 *Instrumental analysis of intrinsically disordered proteins: assessing structure and*
13
14 *conformation*, Uversky, V., Longhi, S., Eds., John Wiley and Sons: Hoboken, New
15
16 Jersey, 2010, pp 569-626.
17
18
19 (88) Hernandez, M. A.; Avila, J.; Andreu, J. M. *Eur. J. Biochem.* 1986, 154, 41.
20
21 (89) Weinreb, P. H.; Zhen, W.; Poon, A. W.; Conway, K. A.; Lansbury, P. T., Jr.
22
23 *Biochemistry* 1996, 35, 13709.
24
25
26 (90) Etoh, Y.; Simon, M.; Green, H. *Biochem. Biophys. Res. Commun.* 1986, 136, 51.
27
28 (91) Lynch, W. P.; Riseman, V. M.; Bretscher, A. *J. Biol. Chem.* 1987, 262, 7429.
29
30 (92) Livernois, A. M.; Hnatchuk, D. J.; Findlater, E. E.; Graether, S. P. *Anal. Biochem.*
31
32 2009, 392, 70.
33
34
35 (93) Kalthoff, C. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2003, 786, 247.
36
37 (94) Cortese, M. S.; Baird, J. P.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* 2005, 4,
38
39 1610.
40
41
42 (95) Uversky, V. N.; Longhi, S., *Instrumental analysis of intrinsically disordered proteins*,
43
44 John Wiley & Sons Hoboken, 2010.
45
46 (96) Ellis, R. *J. Trends Biochem. Sci.* 2001, 26, 597.
47
48 (97) Minton, A. P. *Biophys. J.* 2005, 88, 971.
49
50 (98) Bolen, D. W.; Baskakov, I. V. *J. Mol. Biol.* 2001, 310, 955.
51
52 (99) Munishkina, L. A.; Cooper, E. M.; Uversky, V. N.; Fink, A. L. *J. Mol. Recognit.* 2004,
53
54 17, 456.
55
56
57
58
59
60

- 1
2
3 (100) Breydo, L.; Reddy, K. D.; Piai, A.; Felli, I. C.; Pierattelli, R.; Uversky, V. N. *Biochim.*
4
5 *Biophys. Acta* 2014, 1844, 346.
6
7 (101) Szasz, C. S.; Alexa, A.; Toth, K.; Rakacs, M.; Langowski, J.; Tompa, P. *Biochemistry*
8
9 2011, 50, 5834.
10
11 (102) Asher, G.; Tsvetkov, P.; Kahana, C.; Shaul, Y. *Genes Dev.* 2005, 19, 316.
12
13 (103) Sheaff, R. J.; Singer, J. D.; Swanger, J.; Smitherman, M.; Roberts, J. M.; Clurman, B. E.
14
15 *Mol. Cell* 2000, 5, 403.
16
17 (104) Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.;
18
19 Futaki, S.; Ito, Y.; Hiroaki, H.; Shirakawa, M. *Nature* 2009, 458, 106.
20
21 (105) Bodart, J. F.; Wieruszeski, J. M.; Amniai, L.; Leroy, A.; Landrieu, I.; Rousseau-
22
23 Lescuyer, A.; Vilain, J. P.; Lippens, G. *J. Magn. Reson.* 2008, 192, 252.
24
25 (106) Binolfi, A.; Theillet, F. X.; Selenko, P. *Biochem. Soc. Trans.* 2012, 40, 950.
26
27 (107) Waudby, C. A.; Camilloni, C.; Fitzpatrick, A. W.; Cabrita, L. D.; Dobson, C. M.;
28
29 Vendruscolo, M.; Christodoulou, J. *PLoS One* 2013, 8, e72286.
30
31 (108) McNulty, B. C.; Young, G. B.; Pielak, G. J. *J. Mol. Biol.* 2006, 355, 893.
32
33 (109) Belle, A.; Tanay, A.; Bitincka, L.; Shamir, R.; O'Shea, E. K. *Proc. Natl. Acad. Sci. U.S.A.*
34
35 2006, 103, 13004.
36
37 (110) Tompa, P.; Prilusky, J.; Silman, I.; Sussman, J. L. *Proteins* 2008, 71, 903.
38
39 (111) Prakash, S.; Tian, L.; Ratliff, K. S.; Lehotzky, R. E.; Matouschek, A. *Nat. Struct. Mol.*
40
41 *Biol.* 2004, 11, 830.
42
43 (112) Eralles, J.; Coffino, P. *Biochim. Biophys. Acta* 2014, 1843, 216.
44
45 (113) Hegyi, H.; Tompa, P. *PLoS Comput. Biol.* 2008, 4, e1000017.
46
47 (114) Gsponer, J.; Futschik, M. E.; Teichmann, S. A.; Babu, M. M. *Science* 2008, 322, 1365.
48
49 (115) Vavouri, T.; Semple, J. I.; Garcia-Verdugo, R.; Lehner, B. *Cell* 2009, 138, 198.
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (116) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. *Curr. Opin. Struct. Biol.*
4
5 2011, 21, 432.
6
7 (117) Suskiewicz, M. J.; Sussman, J. L.; Silman, I.; Shaul, Y. *Protein Sci.* 2011, 20, 1285.
8
9 (118) Simister, P. C.; Schaper, F.; O'Reilly, N.; McGowan, S.; Feller, S. M. *PLoS Biol.* 2011,
10
11 9, e1000591.
12
13 (119) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *J. Mol. Recognit.* 2005, 18, 343.
14
15 (120) Uversky, V. N. *Biochim. Biophys. Acta* 2013, 1834, 932.
16
17 (121) Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K.
18
19 *Proteins* 2003, 53 Suppl 6, 566.
20
21 (122) Le Gall, T.; Romero, P. R.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *J. Biomol.*
22
23 *Struct. Dyn.* 2007, 24, 325.
24
25 (123) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* 2005, 6, 197.
26
27 (124) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. *Proc. Natl. Acad. Sci.*
28
29 *U.S.A.* 2010, 107, 8183.
30
31 (125) Das, R. K.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, 13392.
32
33 (126) Tcherkasskaya, O.; Uversky, V. N. *Protein Pept. Lett.* 2003, 10, 239.
34
35 (127) Marsh, J. A.; Forman-Kay, J. D. *Biophys. J.* 2010, 98, 2383.
36
37 (128) Meng, W.; Lyle, N.; Luan, B.; Raleigh, D. P.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.*
38
39 2013, 110, 2123.
40
41 (129) Uversky, V. N.; Dunker, A. K. *Biochim. Biophys. Acta* 2010, 1804, 1231.
42
43 (130) Uversky, V. N.; Dunker, A. K. *Anal. Chem.* 2012, 84, 2096.
44
45 (131) Fasman, G. D., *Circular dichroism and conformational analysis of biomolecules*,
46
47 Plenum Press: New York, 1996.
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (132) Li, P.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.;
4
5 Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q. X.; Nixon, B. T.;
6
7 Rosen, M. K. *Nature* 2012, 483, 336.
8
9
10 (133) Tompa, P. *Intrinsically Disordered Proteins* 2013, 1, e24068.
11
12 (134) Bernado, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. J.
13
14 *Am. Chem. Soc.* 2005, 127, 17968.
15
16
17 (135) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am.*
18
19 *Chem. Soc.* 2007, 129, 5656.
20
21
22 (136) Schneider, R.; Huang, J. R.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon,
23
24 L.; Jensen, M. R.; Blackledge, M. *Mol. Biosyst.* 2012, 8, 58.
25
26
27 (137) Huang, J. R.; Grzesiek, S. *J. Am. Chem. Soc.* 2010, 132, 694.
28
29 (138) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J. R.; Jensen, M. R.; Segard, S.; Bernado, P.;
30
31 Charavay, C.; Blackledge, M. *Bioinformatics* 2012, 28, 1463.
32
33
34 (139) Bernado, P.; Svergun, D. I. *Mol. Biosyst.* 2012, 8, 151.
35
36 (140) Feldman, H. J.; Hogue, C. W. *Proteins* 2000, 39, 112.
37
38 (141) Feldman, H. J.; Hogue, C. W. *Proteins* 2002, 46, 8.
39
40 (142) Eyal, E.; Najmanovich, R.; McConkey, B. J.; Edelman, M.; Sobolev, V. *J. Comput.*
41
42 *Chem.* 2004, 25, 712.
43
44
45 (143) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. *Protein Sci.* 2003, 12, 2001.
46
47 (144) Svergun, D. I.; Barabero, C.; Koch, M. H. *J Appl Cryst* 1995, 28, 768.
48
49 (145) Shen, Y.; Bax, A. *J. Biomol. NMR* 2007, 38, 289.
50
51
52 (146) Daughdrill, G. W.; Kashtanov, S.; Stancik, A.; Hill, S. E.; Helms, G.; Muschol, M.;
53
54 Receveur-Brechot, V.; Ytreberg, F. M. *Mol. Biosyst.* 2012, 8, 308.
55
56
57
58
59
60

- 1
2
3 (147) Schwalbe, M.; Ozenne, V.; Bibow, S.; Jaremko, M.; Jaremko, L.; Gajda, M.; Jensen, M.
4
5 R.; Biernat, J.; Becker, S.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *Structure*
6
7 2013.
8
9
10 (148) Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I.
11
12 C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; Sussman, J.; Svergun, D. I.;
13
14 Uversky, V. N.; Vendruscolo, M.; Wishart, D.; Wright, P. E.; Tompa, P. *Nucleic Acids*
15
16 *Res.* 2014, 42, D326.
17
18
19 (149) Tompa, P. *Curr. Opin. Struct. Biol.* 2011, 21, 419.
20
21 (150) Tran, H. T.; Mao, A.; Pappu, R. V. *J. Am. Chem. Soc.* 2008, 130, 7380.
22
23 (151) Szollosi, E.; Bokor, M.; Bodor, A.; Perczel, A.; Klement, E.; Medzihradszky, K. F.;
24
25 Tompa, K.; Tompa, P. *J. Proteome Res.* 2008, 7, 2291.
26
27
28 (152) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond,
29
30 L.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A* 2010, 107, 14609.
31
32
33 (153) Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Muller-Spath, S.; Pfeil, S. H.;
34
35 Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.*
36
37 2012, 109, 17800.
38
39
40 (154) Brocca, S.; Samalikova, M.; Uversky, V. N.; Lotti, M.; Vanoni, M.; Alberghina, L.;
41
42 Grandori, R. *Proteins* 2009, 76, 731.
43
44
45 (155) Li, J.; Uversky, V. N.; Fink, A. L. *Biochemistry* 2001, 40, 11604.
46
47 (156) Uversky, V. N.; Li, J.; Souillac, P.; Jakes, R.; Goedert, M.; Fink, A. L. *J. Biol. Chem.*
48
49 2002, 277, 25.
50
51
52 (157) Guo, J. T.; Jaromczyk, J. W.; Xu, Y. *Proteins* 2007, 67, 548.
53
54 (158) Zhang, Y.; Stec, B.; Godzik, A. *Structure* 2007, 15, 1141.
55
56 (159) Hazy, E.; Tompa, P. *Chemphyschem* 2009, 10, 1415.
57
58
59
60

- 1
2
3 (160) Uversky, V. N. *Biochim. Biophys. Acta* 2011, 1814, 693.
4
5 (161) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z.
6
7 *Biochemistry* 2002, 41, 6573.
8
9
10 (162) Mukhopadhyay, R.; Hoh, J. H. *FEBS Lett.* 2001, 505, 374.
11
12 (163) Patel, S. S.; Belmont, B. J.; Sante, J. M.; Rexach, M. F. *Cell* 2007, 129, 83.
13
14 (164) Kriwacki, R. W.; Hengst, L.; Tennant, L.; Reed, S. I.; Wright, P. E. *Proc. Natl. Acad.*
15
16 *Sci. U.S.A* 1996, 93, 11504.
17
18
19 (165) Russo, A. A.; Jeffrey, P. D.; Patten, A. K.; Massague, J.; Pavletich, N. P. *Nature* 1996,
20
21 382, 325.
22
23 (166) Daughdrill, G. W.; Chadsey, M. S.; Karlinsey, J. E.; Hughes, K. T.; Dahlquist, F. W.
24
25 *Nat. Struct. Biol.* 1997, 4, 285.
26
27
28 (167) Peti, W.; Nairn, A. C.; Page, R. *FEBS J.* 2013, 280, 596.
29
30 (168) Peti, W.; Nairn, A. C.; Page, R. *Curr. Phys. Chem.* 2012, 2, 107.
31
32 (169) Tompa, P.; Szasz, C.; Buday, L. *Trends Biochem. Sci.* 2005, 30, 484.
33
34 (170) Lu, Y.; Bennick, A. *Arch. Oral Biol.* 1998, 43, 717.
35
36 (171) Kalmar, L.; Homola, D.; Varga, G.; Tompa, P. *Bone* 2012, 51, 528.
37
38 (172) Toth-Petroczy, A.; Oldfield, C. J.; Simon, I.; Takagi, Y.; Dunker, A. K.; Uversky, V. N.;
39
40 Fuxreiter, M. *PLoS Comput. Biol.* 2008, 4, e1000243.
41
42 (173) Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *Prog. Biophys. Mol. Biol.* 2008, 98, 85.
43
44 (174) Balazs, A.; Csizmok, V.; Buday, L.; Rakacs, M.; Kiss, R.; Bokor, M.; Udupa, R.;
45
46 Tompa, K.; Tompa, P. *FEBS J.* 2009, 276, 3744.
47
48 (175) Dosztanyi, Z.; Chen, J.; Dunker, A. K.; Simon, I.; Tompa, P. *J. Proteome Res.* 2006, 5,
49
50 2985.
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (176) Haynes, C.; Oldfield, C. J.; Ji, F.; Klitgord, N.; Cusick, M. E.; Radivojac, P.; Uversky, V.
4
5 N.; Vidal, M.; Iakoucheva, L. M. *PLoS Comput. Biol.* 2006, 2, e100.
6
7 (177) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. *FEBS J.*
8
9 2005, 272, 5129.
10
11 (178) Barbar, E. *Biochemistry* 2008, 47, 503.
12
13 (179) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.;
14
15 Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* 2004, 32, 1037.
16
17 (180) Radivojac, P.; Vacic, V.; Haynes, C.; Cocklin, R. R.; Mohan, A.; Heyen, J. W.; Goebel, M.
18
19 G.; Iakoucheva, L. M. *Proteins* 2010, 78, 365.
20
21 (181) Tompa, P.; Csermely, P. *FASEB J.* 2004, 18, 1169.
22
23 (182) Kim, T. D.; Paik, S. R.; Yang, C. H. *Biochemistry* 2002, 41, 13782.
24
25 (183) Bhattacharyya, J.; Das, K. P. *J. Biol. Chem.* 1999, 274, 15505.
26
27 (184) Chakrabortee, S.; Boschetti, C.; Walton, L. J.; Sarkar, S.; Rubinsztein, D. C.;
28
29 Tunnacliffe, A. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 18073.
30
31 (185) Kovacs, D.; Kalmar, E.; Torok, Z.; Tompa, P. *Plant Physiol.* 2008, 147, 381.
32
33 (186) Andley, U. P.; Mathur, S.; Griest, T. A.; Petrash, J. M. *J. Biol. Chem.* 1996, 271,
34
35 31973.
36
37 (187) Pasta, S. Y.; Raman, B.; Ramakrishna, T.; Rao Ch, M. *J. Biol. Chem.* 2002, 277,
38
39 45821.
40
41 (188) Santner, A. A.; Croy, C. H.; Vasanwala, F. H.; Uversky, V. N.; Van, Y. Y.; Dunker, A. K.
42
43 *Biochemistry* 2012, 51, 7250.
44
45 (189) Reichmann, D.; Xu, Y.; Cremers, C. M.; Ilbert, M.; Mittelman, R.; Fitzgerald, M. C.;
46
47 Jakob, U. *Cell* 2012, 148, 947.
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (190) Neduva, V.; Linding, R.; Su-Angrand, I.; Stark, A.; Masi, F. D.; Gibson, T. J.; Lewis, J.;
4
5 Serrano, L.; Russell, R. B. *PLoS Biol.* 2005, 3, e405.
6
7 (191) Fuxreiter, M.; Tompa, P.; Simon, I. *Bioinformatics* 2007, 23, 950.
8
9 (192) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Romero, P.; Uversky, V. N.; Dunker, A. K.
10
11 *Biochemistry* 2005, 44, 12454.
12
13 (193) Neduva, V.; Russell, R. B. *FEBS Lett.* 2005, 579, 3342.
14
15 (194) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J. Mol. Biol.* 2004, 338, 1015.
16
17 (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.*
18
19 *Protein Pept. Sci.* 2012, 13, 34.
20
21
22
23 (196) Fuxreiter, M. *Mol. Biosyst.* 2012, 8, 168.
24
25 (197) Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.;
26
27 Uversky, V. N. *J. Mol. Biol.* 2006, 362, 1043.
28
29 (198) Vacic, V.; Oldfield, C. J.; Mohan, A.; Radivojac, P.; Cortese, M. S.; Uversky, V. N.;
30
31 Dunker, A. K. *J. Proteome Res.* 2007, 6, 2351.
32
33 (199) Puntervoll, P.; Linding, R.; Gemund, C.; Chabanis-Davidson, S.; Mattingsdal, M.;
34
35 Cameron, S.; Martin, D. M.; Ausiello, G.; Brannetti, B.; Costantini, A.; Ferre, F.;
36
37 Maselli, V.; Via, A.; Cesareni, G.; Diella, F.; Superti-Furga, G.; Wyrwicz, L.; Ramu, C.;
38
39 McGuigan, C.; Gudavalli, R.; Letunic, I.; Bork, P.; Rychlewski, L.; Kuster, B.; Helmer-
40
41 Citterich, M.; Hunter, W. N.; Aasland, R.; Gibson, T. J. *Nucleic Acids Res.* 2003, 31,
42
43 3625.
44
45 (200) Davey, N. E.; Shields, D. C.; Edwards, R. J. *Nucleic Acids Res.* 2006, 34, 3546.
46
47 (201) Fuxreiter, M.; Tompa, P., *Fuzziness: structural disorder in protein complexes*,
48
49 Springer Science + Business Media, LLC Landes Bioscience: New York, USA, 2012.
50
51 (202) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* 2002, 12, 54.
52
53
54
55
56
57
58
59
60

- 1
2
3 (203) Tompa, P.; Fuxreiter, M.; Oldfield, C. J.; Simon, I.; Dunker, A. K.; Uversky, V. N.
4
5 *Bioessays* 2009, 31, 328.
6
7 (204) Chen, J. W.; Romero, P.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* 2006, 5, 888.
8
9 (205) Chen, J. W.; Romero, P.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* 2006, 5, 879.
10
11 (206) Xue, B.; Brown, C. J.; Dunker, A. K.; Uversky, V. N. *Biochim. Biophys. Acta* 2013,
12
13 1834, 725.
14
15 (207) Dosztanyi, Z.; Meszaros, B.; Simon, I. *Bioinformatics* 2009, 25, 2745.
16
17 (208) Cilia, E.; Pancsa, R.; Tompa, P.; Lenaerts, T.; Vranken, W. F. *Nat. Commun.* 2013, 4,
18
19 2741.
20
21 (209) Meszaros, B.; Simon, I.; Dosztanyi, Z. *PLoS Comput. Biol.* 2009, 5, e1000376.
22
23 (210) Hammes, G. G.; Chang, Y. C.; Oas, T. G. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106,
24
25 13737.
26
27 (211) Wang, Y.; Tang, C.; Wang, E.; Wang, J. *PLoS Comput. Biol.* 2012, 8, e1002471.
28
29 (212) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* 2009, 19, 31.
30
31 (213) Oldfield, C. J.; Meng, J.; Yang, J. Y.; Yang, M. Q.; Uversky, V. N.; Dunker, A. K. *BMC*
32
33 *Genomics* 2008, 9 Suppl 1, S1.
34
35 (214) Pancsa, R.; Fuxreiter, M. *IUBMB Life* 2012, 64, 513.
36
37 (215) Tsai, C. D.; Ma, B.; Kumar, S.; Wolfson, H.; Nussinov, R. *Crit. Rev. Biochem. Mol. Biol.*
38
39 2001, 36, 399.
40
41 (216) Tsai, C. J.; Ma, B.; Sham, Y. Y.; Kumar, S.; Nussinov, R. *Proteins* 2001, 44, 418.
42
43 (217) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 17327.
44
45 (218) Fersht, A. R.; Sato, S. *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 7976.
46
47 (219) Kiefhaber, T.; Bachmann, A.; Jensen, K. S. *Curr. Opin. Struct. Biol.* 2012, 22, 21.
48
49 (220) Rogers, J. M.; Steward, A.; Clarke, J. *J. Am. Chem. Soc.* 2013, 135, 1415.
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (221) Shammas, S. L.; Travis, A. J.; Clarke, J. *J. Phys. Chem. B* 2013.
4
5 (222) Giri, R.; Morrone, A.; Toto, A.; Brunori, M.; Gianni, S. *Proc. Natl. Acad. Sci. U.S.A.*
6
7 2013, 110, 14942.
8
9 (223) Huang, Y.; Liu, Z. *J. Mol. Biol.* 2009, 393, 1143.
10
11 (224) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* 2000, 97,
12
13 8868.
14
15 (225) Csizmok, V.; Bokor, M.; Banki, P.; Klement, E.; Medzihradzsky, K. F.; Friedrich, P.;
16
17 Tompa, K.; Tompa, P. *Biochemistry* 2005, 44, 3955.
18
19 (226) Espinoza-Fonseca, L. M. *Biochem. Biophys. Res. Commun.* 2009, 382, 479.
20
21 (227) Radhakrishnan, I.; Perez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.;
22
23 Wright, P. E. *Cell* 1997, 91, 741.
24
25 (228) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* 2007, 447, 1021.
26
27 (229) Savvides, S. N.; Raghunathan, S.; Futterer, K.; Kozlov, A. G.; Lohman, T. M.;
28
29 Waksman, G. *Protein Sci.* 2004, 13, 1942.
30
31 (230) Sigalov, A. B.; Zhuravleva, A. V.; Orekhov, V. Y. *Biochimie* 2007, 89, 419.
32
33 (231) Sigalov, A. B.; Kim, W. M.; Saline, M.; Stern, L. J. *Biochemistry* 2008, 47, 12942.
34
35 (232) Communie, G.; Habchi, J.; Yabukarski, F.; Blocquel, D.; Schneider, R.; Tarbouriech,
36
37 N.; Papageorgiou, N.; Ruigrok, R. W.; Jamin, M.; Jensen, M. R.; Longhi, S.;
38
39 Blackledge, M. *PLoS Pathog.* 2013, 9, e1003631.
40
41 (233) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* 2008, 33, 2.
42
43 (234) Meszaros, B.; Tompa, P.; Simon, I.; Dosztanyi, Z. *J. Mol. Biol.* 2007, 372, 549.
44
45 (235) Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* 1999, 285, 2177.
46
47 (236) Shammas, S. L.; Rogers, J. M.; Hill, S. A.; Clarke, J. *Biophys. J.* 2012, 103, 2203.
48
49 (237) Huang, Y.; Liu, Z. *Chemistry (Easton)* 2013, 19, 4462.
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (238) Wong, E. T.; Na, D.; Gsponer, J. *PLoS Comput. Biol.* 2013, 9, e1003192.
4
5 (239) Brown, C. J.; Johnson, A. K.; Dunker, A. K.; Daughdrill, G. W. *Curr. Opin. Struct. Biol.*
6
7 2011, 21, 441.
8
9 (240) Davey, N. E.; Cowan, J. L.; Shields, D. C.; Gibson, T. J.; Coldwell, M. J.; Edwards, R. J.
10
11 *Nucleic Acids Res.* 2012, 40, 10628.
12
13 (241) Espinoza-Fonseca, L. M. *Mol. Biosyst.* 2012, 8, 237.
14
15 (242) Fuxreiter, M.; Tompa, P. *Trends Biochem. Sci.* 2009, 34, 3.
16
17 (243) Polverini, E.; Rangaraj, G.; Libich, D. S.; Boggs, J. M.; Harauz, G. *Biochemistry* 2008,
18
19 47, 267.
20
21 (244) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr. *Angew. Chem. Int. Ed. Engl.*
22
23 2005, 44, 7342.
24
25 (245) Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G. *Nat. Methods* 2007, 4, 798.
26
27 (246) Yang, X. *J. Oncogene* 2005, 24, 1653.
28
29 (247) Faradonbeh, M. Z.; Gharechahi, J.; Mollamohammadi, S.; Pakzad, M.; Taei, A.;
30
31 Rassouli, H.; Baharvand, H.; Salekdeh, G. H. *Mol. Biosyst.* 2012, 8, 1833.
32
33 (248) Mersfelder, E. L.; Parthun, M. R. *Nucleic Acids Res.* 2006, 34, 2653.
34
35 (249) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.;
36
37 Uversky, V. N. *J. Proteome Res.* 2007, 6, 1917.
38
39 (250) Dunker, A. K.; Brown, C. J.; Obradovic, Z. *Adv. Protein Chem.* 2002, 62, 25.
40
41 (251) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.;
42
43 Obradovic, Z. *J. Proteome Res.* 2007, 6, 1882.
44
45 (252) Uversky, V. N. *Curr. Pharm. Des.* 2013, 19, 4191.
46
47 (253) Sambrook, J. *Nature* 1977, 268, 101.
48
49 (254) Black, D. L. *Annu. Rev. Biochem* 2003, 72, 291.
50
51
52
53
54
55
56
57
58
59
60

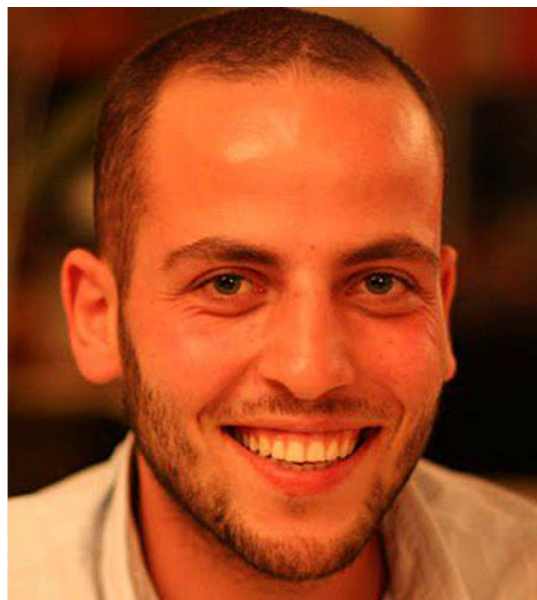
- 1
2
3 (255) Boue, S.; Letunic, I.; Bork, P. *Bioessays* 2003, 25, 1031.
4
5 (256) Ast, G. *Nat. Rev. Genet.* 2004, 5, 773.
6
7 (257) Kelemen, O.; Convertini, P.; Zhang, Z.; Wen, Y.; Shen, M.; Falaleeva, M.; Stamm, S.
8
9 *Gene* 2013, 514, 1.
10
11 (258) Stamm, S.; Ben-Ari, S.; Rafalska, I.; Tang, Y.; Zhang, Z.; Toiber, D.; Thanaraj, T. A.;
12
13 Soreq, H. *Gene* 2005, 344, 1.
14
15 (259) Brett, D.; Hanke, J.; Lehmann, G.; Haase, S.; Delbruck, S.; Krueger, S.; Reich, J.; Bork,
16
17 P. *FEBS Lett.* 2000, 474, 83.
18
19 (260) Johnson, J. M.; Castle, J.; Garrett-Engele, P.; Kan, Z.; Loerch, P. M.; Armour, C. D.;
20
21 Santos, R.; Schadt, E. E.; Stoughton, R.; Shoemaker, D. D. *Science* 2003, 302, 2141.
22
23 (261) Graveley, B. R. *Trends Genet.* 2001, 17, 100.
24
25 (262) Minneman, K. P. *Mol. Interv.* 2001, 1, 108.
26
27 (263) Thai, T. H.; Kearney, J. F. *J. Immunol.* 2004, 173, 4009.
28
29 (264) Thai, T. H.; Kearney, J. F. *Adv. Immunol.* 2005, 86, 113.
30
31 (265) Scheper, W.; Zwart, R.; Baas, F. *Neurogenetics* 2004, 5, 223.
32
33 (266) Romero, P. R.; Zaidi, S.; Fang, Y. Y.; Uversky, V. N.; Radivojac, P.; Oldfield, C. J.;
34
35 Cortese, M. S.; Sickmeier, M.; LeGall, T.; Obradovic, Z.; Dunker, A. K. *Proc. Natl.*
36
37 *Acad. Sci. U.S.A.* 2006, 103, 8390.
38
39 (267) Buljan, M.; Chalancon, G.; Dunker, A. K.; Bateman, A.; Balaji, S.; Fuxreiter, M.; Babu,
40
41 M. M. *Curr. Opin. Struct. Biol.* 2013, 23, 443.
42
43 (268) Liu, S.; Cheng, C. *Wiley Interdiscip. Rev. RNA* 2013, 4, 547.
44
45 (269) Skotheim, R. I.; Nees, M. *Int. J. Biochem. Cell Biol.* 2007, 39, 1432.
46
47 (270) Fu, R. H.; Liu, S. P.; Huang, S. J.; Chen, H. J.; Chen, P. R.; Lin, Y. H.; Ho, Y. C.; Chang,
48
49 W. L.; Tsai, C. H.; Shyu, W. C.; Lin, S. Z. *Cell Transplant.* 2013, 22, 653.
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (271) Niblock, M.; Gallo, J. M. *Biochem. Soc. Trans.* 2012, 40, 677.
4
5 (272) Mills, J. D.; Janitz, M. *Neurobiol. Aging* 2012, 33, 1012.e11.
6
7 (273) Lara-Pezzi, E.; Gomez-Salinerro, J.; Gatto, A.; Garcia-Pavia, P. *J. Cardiovasc. Transl.*
8
9 *Res.* 2013, 6, 945.
10
11 (274) Venables, J. P.; Burn, J. *Nucleic Acids Res.* 2006, 34, e103.
12
13 (275) Hastings, M. L.; Resta, N.; Traum, D.; Stella, A.; Guanti, G.; Krainer, A. R. *Nat. Struct.*
14
15 *Mol. Biol.* 2005, 12, 54.
16
17 (276) Orban, T. I.; Olah, E. *Mol. Pathol.* 2003, 56, 191.
18
19 (277) Uversky, V. N. *Front. Biosci. (Landmark Ed)* 2009, 14, 5188.
20
21 (278) Koshland, D. E., Jr.; Nemethy, G.; Filmer, D. *Biochemistry* 1966, 5, 365.
22
23 (279) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* 1965, 12, 88.
24
25 (280) Hilser, V. J.; Thompson, E. B. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 8311.
26
27 (281) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. *Annu. Rev. Biophys.* 2012, 41, 585.
28
29 (282) Garcia-Pino, A.; Balasubramanian, S.; Wyns, L.; Gazit, E.; De Greve, H.; Magnuson,
30
31 R. D.; Charlier, D.; van Nuland, N. A.; Loris, R. *Cell* 2010, 142, 101.
32
33 (283) Motlagh, H. N.; Hilser, V. J. *Proc. Natl. Acad. Sci. U.S.A.* 2012, 109, 4134.
34
35 (284) Follis, A. V.; Chipuk, J. E.; Fisher, J. C.; Yun, M. K.; Grace, C. R.; Nourse, A.; Baran, K.;
36
37 Ou, L.; Min, L.; White, S. W.; Green, D. R.; Kriwacki, R. W. *Nat. Chem. Biol.* 2013, 9,
38
39 163.
40
41 (285) Gruet, A.; Dosnon, M.; Vassena, A.; Lombard, V.; Gerlier, D.; Bignon, C.; Longhi, S. *J.*
42
43 *Mol. Biol.* 2013, 425, 3495.
44
45 (286) Tompa, P. *Bioessays* 2003, 25, 847.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (287) Kato, M.; Han, T. W.; Xie, S.; Shi, K.; Du, X.; Wu, L. C.; Mirzaei, H.; Goldsmith, E. J.;
4
5 Longgood, J.; Pei, J.; Grishin, N. V.; Frantz, D. E.; Schneider, J. W.; Chen, S.; Li, L.;
6
7 Sawaya, M. R.; Eisenberg, D.; Tycko, R.; McKnight, S. L. *Cell* 2012, 149, 753.
8
9
10 (288) Beck, R.; Deek, J.; Safinya, C. R. *Biochem. Soc. Trans.* 2012, 40, 1027.
11
12 (289) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* 2004,
13
14 337, 635.
15
16 (290) Schad, E.; Tompa, P.; Hegyi, H. *Genome Biol.* 2011, 12, R120.
17
18 (291) Tompa, P.; Dosztanyi, Z.; Simon, I. *J. Proteome Res.* 2006, 5, 1996.
19
20 (292) Burra, P. V.; Kalmar, L.; Tompa, P. *PLoS One* 2010, 5, e12069.
21
22 (293) Pancsa, R.; Tompa, P. *PLoS One* 2012, 7, e34687.
23
24 (294) Davey, N. E.; Trave, G.; Gibson, T. J. *Trends Biochem. Sci.* 2011.
25
26 (295) Tokuriki, N.; Oldfield, C. J.; Uversky, V. N.; Berezovsky, I. N.; Tawfik, D. S. *Trends*
27
28 *Biochem. Sci.* 2009, 34, 53.
29
30
31 (296) Jordan, I. K.; Sutter, B. A.; McClure, M. A. *Mol. Biol. Evol.* 2000, 17, 75.
32
33 (297) Narechania, A.; Terai, M.; Burk, R. D. *J. Gen. Virol.* 2005, 86, 1307.
34
35 (298) Rancurel, C.; Khosravi, M.; Dunker, K. A.; Romero, P. R.; Karlin, D. *J. Virol.* 2009, 83,
36
37 10719.
38
39 (299) Bourhis, J. M.; Canard, B.; Longhi, S. *Virology* 2006, 344, 94.
40
41 (300) Longhi, S. *Protein Pept. Lett.* 2010, 17, 930.
42
43 (301) Longhi, S.; Oglesbee, M. *Protein Pept. Lett.* 2010, 17, 961.
44
45 (302) Habchi, J.; Longhi, S. *Mol. Biosyst.* 2012, 8, 69.
46
47 (303) Karlin, D.; Longhi, S.; Receveur, V.; Canard, B. *Virology* 2002, 296, 251.
48
49 (304) Longhi, S.; Receveur-Brechot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.;
50
51 Yeo, R.; Finet, S.; Canard, B. *J. Biol. Chem.* 2003, 278, 18638.
52
53
54
55
56
57
58
59
60

- 1
2
3 (305) Karlin, D.; Ferron, F.; Canard, B.; Longhi, S. *J. Gen. Virol.* 2003, 84, 3239.
4
5 (306) Xue, B.; Williams, R. W.; Oldfield, C. J.; Goh, G. K.; Dunker, A. K.; Uversky, V. N.
6
7 *Protein Pept. Lett.* 2010, 17, 932.
8
9
10 (307) Uversky, V. N.; Longhi, S., *Flexible viruses: structural disorder in viral proteins*,
11
12 Uversky, V. N., Longhi, S.; John Wiley and Sons: Hoboken, 2012.
13
14 (308) Alves, C.; Cunha, C. *Future Virol.* 2012, 7, 1183.
15
16
17 (309) Xue, B.; Dunker, A. K.; Uversky, V. N. *J. Biomol. Struct. Dyn.* 2012, 30, 137.
18
19 (310) Galea, C. A.; High, A. A.; Obenauer, J. C.; Mishra, A.; Park, C. G.; Punta, M.;
20
21 Schlessinger, A.; Ma, J.; Rost, B.; Slaughter, C. A.; Kriwacki, R. W. *J. Proteome Res.*
22
23 2009, 8, 211.
24
25
26 (311) Tompa, P. *Trends Biochem. Sci.* 2012, 37, 509.
27
28 (312) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. *Genome Inform.*
29
30 *Ser. Workshop Genome Inform.* 2000, 11, 161.
31
32
33 (313) Galea, C. A.; Wang, Y.; Sivakolundu, S. G.; Kriwacki, R. W. *Biochemistry* 2008, 47,
34
35 7598.
36
37 (314) Huber, A. H.; Stewart, D. B.; Laurents, D. V.; Nelson, W. J.; Weis, W. I. *J. Biol. Chem.*
38
39 2001, 276, 12301.
40
41
42 (315) Paunola, E.; Mattila, P. K.; Lappalainen, P. *FEBS Lett.* 2002, 513, 92.
43
44 (316) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. *J. Mol.*
45
46 *Biol.* 2002, 323, 573.
47
48 (317) Brown, C. J.; Takayama, S.; Campen, A. M.; Vise, P.; Marshall, T. W.; Oldfield, C. J.;
49
50 Williams, C. J.; Dunker, A. K. *J. Mol. Evol.* 2002, 55, 104.
51
52 (318) Bellay, J.; Han, S.; Michaut, M.; Kim, T.; Costanzo, M.; Andrews, B. J.; Boone, C.;
53
54 Bader, G. D.; Myers, C. L.; Kim, P. M. *Genome Biol.* 2011, 12, R14.
55
56
57
58
59
60

- 1
2
3 (319) Tokuriki, N.; Tawfik, D. S. *Science* 2009, 324, 203.
4
5 (320) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. *PLoS Comput. Biol.* 2008, 4,
6
7 e1000002.
8
9
10 (321) Olson, K. E.; Narayanaswami, P.; Vise, P. D.; Lowry, D. F.; Wold, M. S.; Daughdrill, G.
11
12 *W. J. Biomol. Struct. Dyn.* 2005, 23, 113.
13
14 (322) Chen, S. C.; Chen, F. C.; Li, W. H. *Mol. Biol. Evol.* 2010, 27, 2548.
15
16
17 (323) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* 2008, 37, 215.
18
19 (324) Uversky, V. N. *Curr. Alzheimer Res.* 2008, 5, 260.
20
21 (325) Ribeiro-Silva, C.; Gilberto, S.; Gomes, R. A.; Mateus, L.; Monteiro, E.; Barroso, E.;
22
23 Coelho, A. V.; da Costa, G. A.; Freire, A. P.; Cordeiro, C. *Amyloid* 2011, 18, 191.
24
25 (326) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem* 2006, 75, 333.
26
27 (327) Uversky, V. N.; Oldfield, C. J.; Midic, U.; Xie, H.; Xue, B.; Vucetic, S.; Iakoucheva, L.
28
29 M.; Obradovic, Z.; Dunker, A. K. *BMC Genomics* 2009, 10 Suppl 1, S7.
30
31 (328) Cheng, Y.; Legall, T.; Oldfield, C. J.; Mueller, J. P.; Van, Y. Y.; Romero, P.; Cortese, M.
32
33 S.; Uversky, V. N.; Dunker, A. K. *Trends Biotechnol.* 2006, 24, 435.
34
35 (329) Vassilev, L. T. *Cell Cycle* 2004, 3, 419.
36
37 (330) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.;
38
39 Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. A. *Science* 2004, 303, 844.
40
41 (331) Metallo, S. J. *Curr. Opin. Chem. Biol.* 2010, 14, 481.
42
43 (332) Sung, Y. H.; Eliezer, D. J. *Mol. Biol.* 2007, 372, 689.
44
45 (333) Uversky, V. N. *Cell. Mol. Life Sci.* 2003, 60, 1852.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 has research groups both in Budapest and Brussels. He was promoter of 10 doctorates, published
4
5 121 papers, 10 book chapters, and the first monograph of the field “Structure and function of
6
7 intrinsically disordered proteins” (2009) by Taylor and Francis, Inc (CRC Press). He received about
8
9 5500 citations and his H-index is 32.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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".



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

Page 95 Submitted to Chemical Reviews

1

2

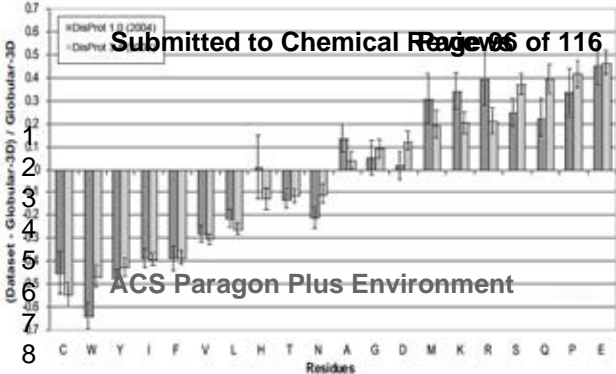
3

4

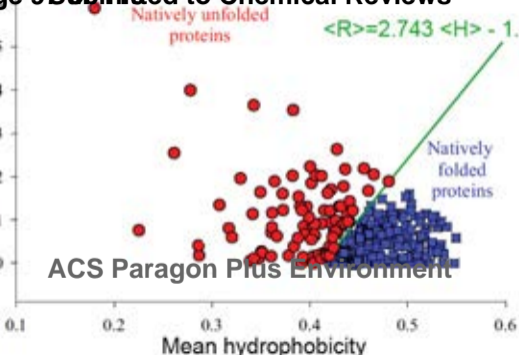
ACS Paragon Plus Environment

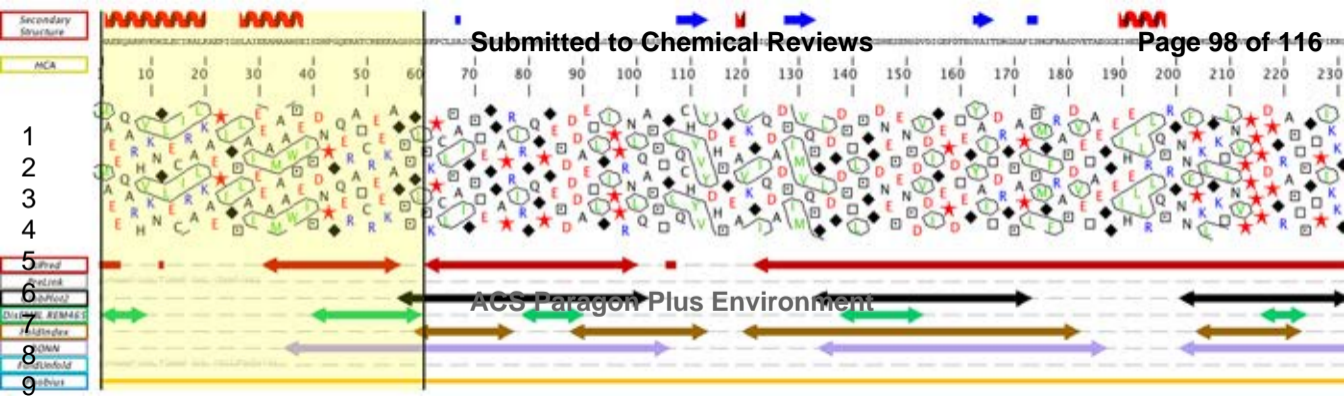
B

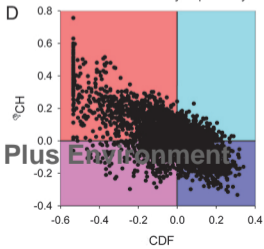
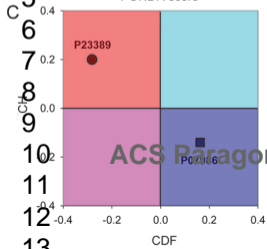
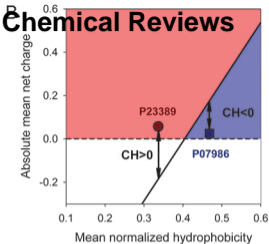
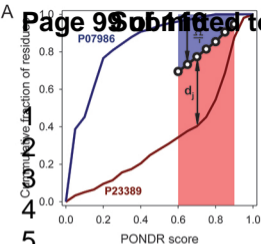


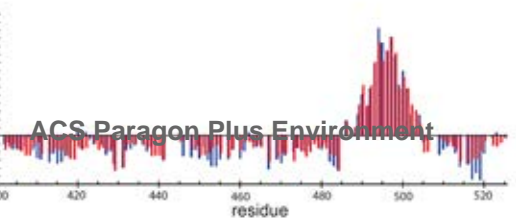
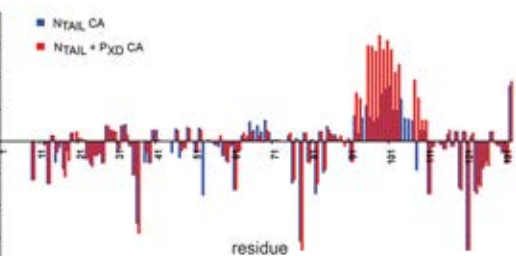
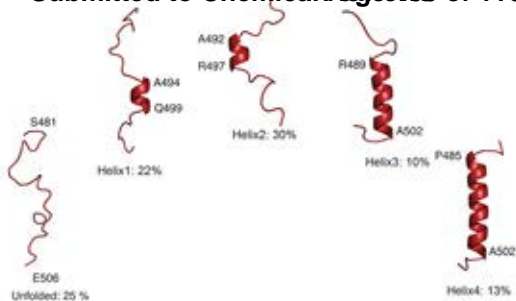


1
2
3
4
5
6
7
8
9
10



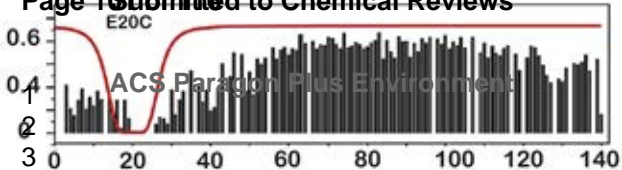




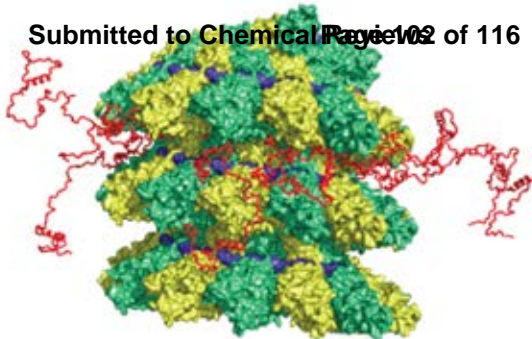


α -synuclein

Page 1 Submitted to Chemical Reviews

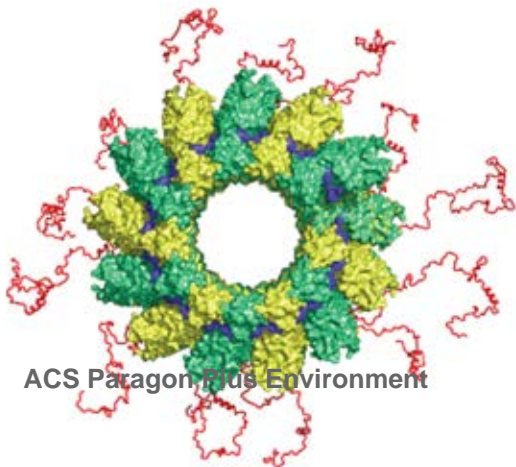


A

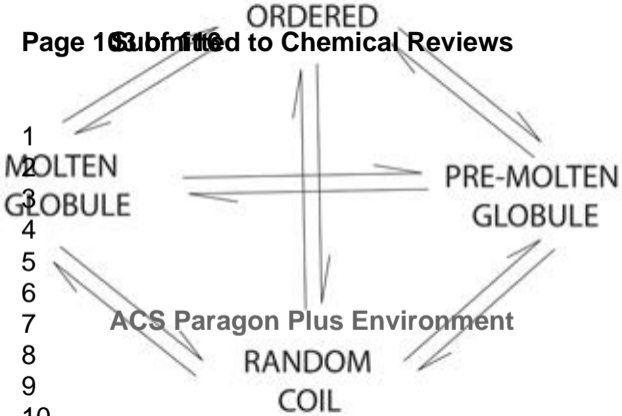


1
2
3
4
5
6
7
8
9

B



11
12
13
14
15
16
17
18
19
20
21
22
23
24



A



B



C



D



E



F



G



H



I



J



ACS Paragon IPAs Environment
Submitted to Chemical Abstracts of 116

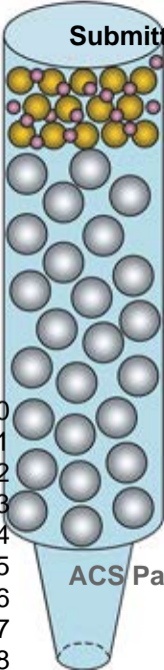
1

2

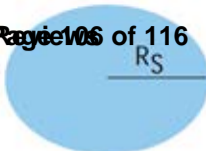
3

Disorder

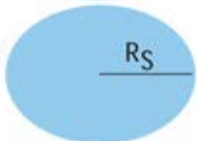


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

RC



PMG



MG

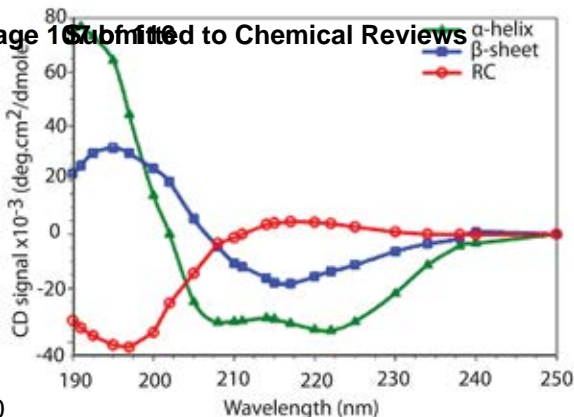


Folded

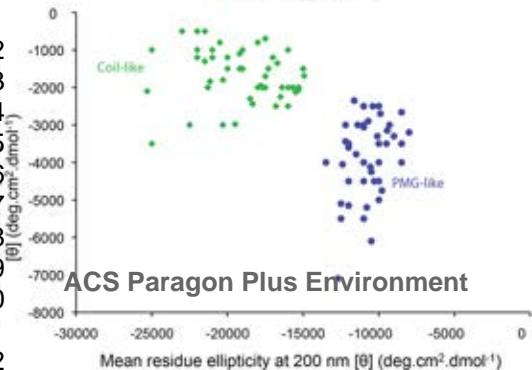


ACS Paragon Plus Environment

A

1
2
3
4
5
6
7
8
9

B

10
11
12
13
14
15
16
17
18
19
20
21
22

ACS Paragon Plus Environment

1
2
3
4
5
6
7

Amino acid sequence

Non-folding

Folding

Flexible ensembles

MG

RC

3D-structure

Molecular

Functions

Effectors
Scaffolds
Assemblers
Chaperones
Display sites

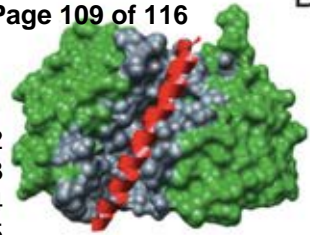
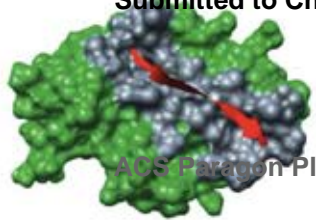
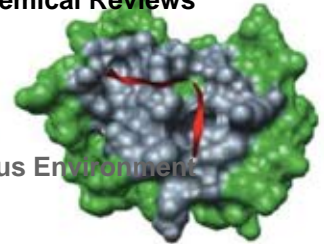
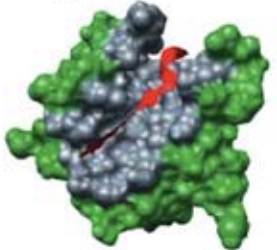
Functions

Entropic chains
Flexible linkers

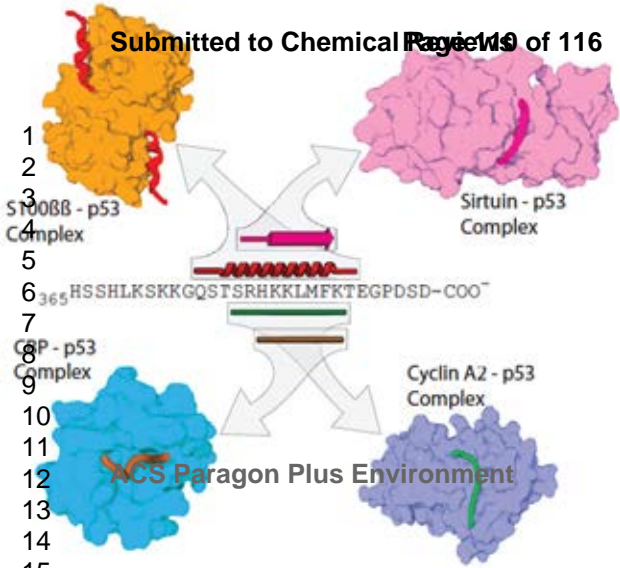
ACS Paragon Plus Environment

Functions

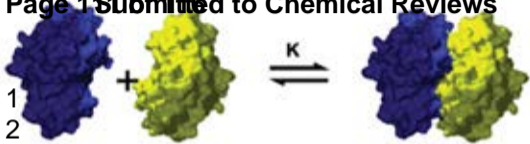
Catalysis
Transport
Biosynthesis
Metabolism

A**Page 109 of 116**1
2
3
4
5
6**B****Submitted to Chemical Reviews****C****D**

ACS Paragon Plus Environment



Page 1 Submitted to Chemical Reviews

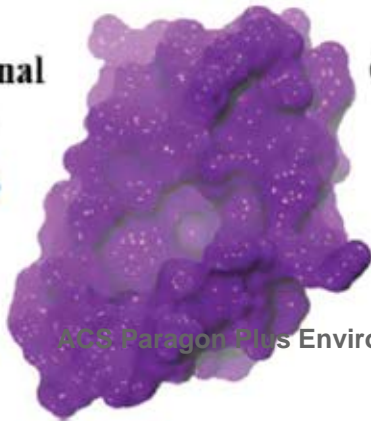


3 Decoupled affinity and specificity

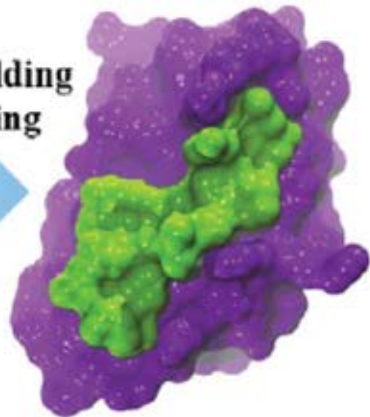


1
2
3
4
5
6
7
8
9
10
11
12
13
14

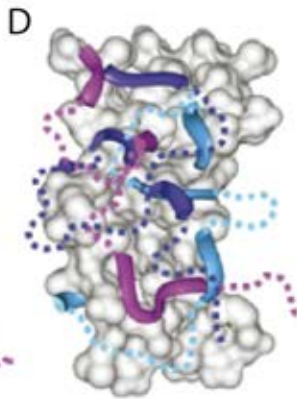
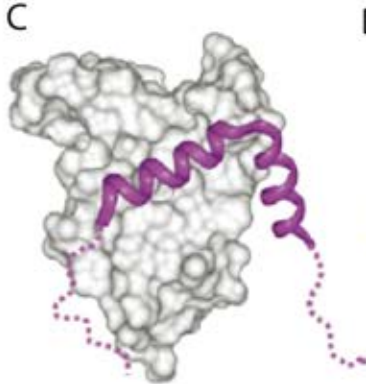
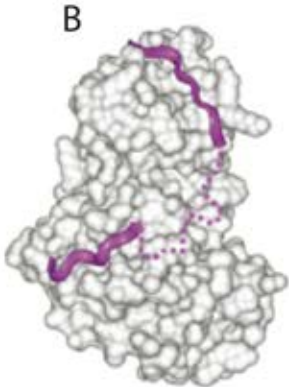
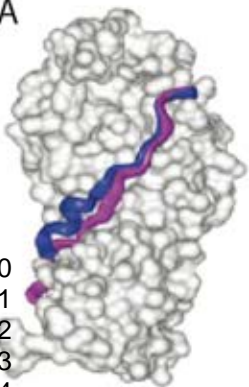
Conformational Selection



Coupled folding and binding



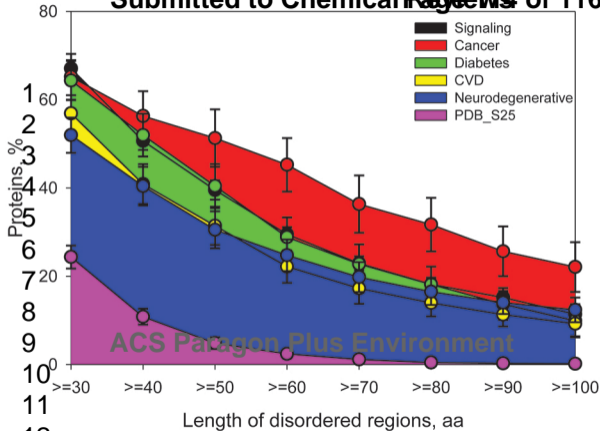
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

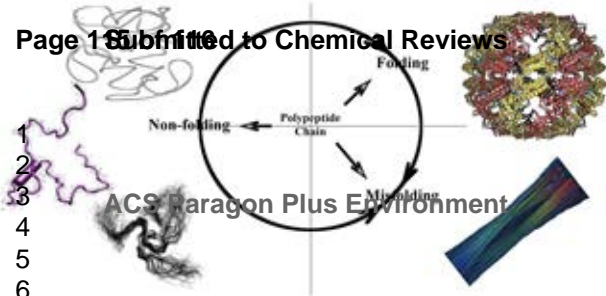


ACS Paragon Plus Environment

Disorder



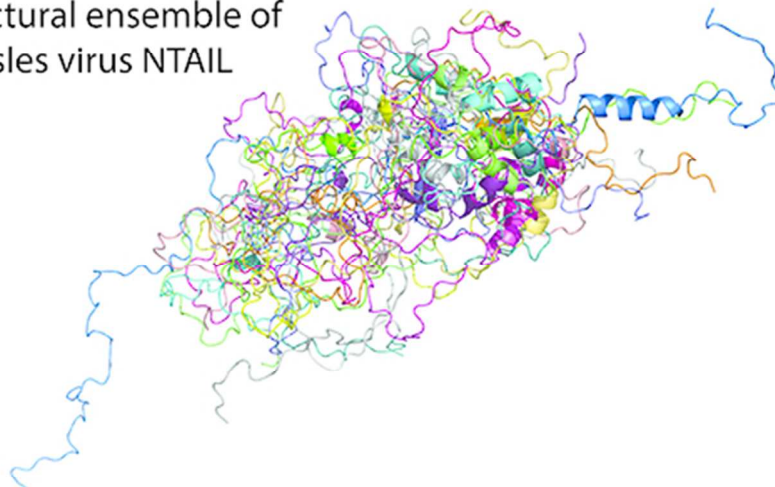




1
2
3
4
5
6

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Structural ensemble of
measles virus NTAIL



Ribbon representation of the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (NTAIL) as a conformational ensemble (accession number 7AAC in the Protein Ensemble Database). Only 15 out of the 10,000 conformers generated by Flexible-Meccano are depicted.
50x37mm (300 x 300 DPI)