Review

Intrinsically disordered proteins: a 10-year recap

Peter Tompa[1,2]

[1] VIB Department of Structural Biology, Vrije Universiteit Brussel, Brussels, Belgium
[2] Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

The suggestion that the native state of many proteins is intrinsically disordered (or, as originally termed, unstructured) is now integral to our general view of protein structure and function. A little more than 10 years ago, however, such challenge to the almost dogmatic 'structure–function paradigm' was pure heresy due to the overwhelming evidence that structure determines function. A decade of steady progress turned skepticism around: this 10-year recap review outlines the situation a decade ago and the major directions of the breathtaking advance achieved by experimental and computational approaches. I show that the evidence for the generality and importance of this phenomenon is now so insurmountable that it demands the inclusion of 'unstructural' biology into mainstream biology and biochemistry textbooks.

Structural disorder taking over

Over the past century evidence steadily accumulated that a well-defined structure is the prerequisite of protein function. Basic biology and biochemistry textbooks that explain biological phenomena at the molecular level exquisitely rely on this notion, the 'structure–function paradigm'. Although deviations from this norm were always apparent, they had been invariably neglected or ignored. Only around the turn of the millennium was it eventually formally raised in several conceptual papers [1–3], among them one in *TiBS* [4], that many proteins or regions of proteins are intrinsically disordered (IDPs, or as originally termed, unstructured) under native, functional conditions. Although it was based on rather limited evidence, the groundwork of the new field was laid and sparked an immediate rapid expansion. Many visionary elements of the new paradigm stood the test of time and it is now beyond doubt that 'unstructural' biology has become an integral part of molecular biology [5]. Here, I outline the most important developments, acknowledging that, due to an incredible pace of progress, one can only survey the most crucial observations and concepts, directing the reader for more comprehensive coverage to several recent reviews [6–8] and a textbook [9].

Expanding evidence for disorder

As suggested [4], the transition in paradigm was enforced by scattered experimental observations of disorder in a few dozen proteins [1–3], whereas structural biology was at that time based on the ~18 000 structures deposited in the Protein Data Bank (PDB) (which has now grown to more than 80 000 [10]). Intriguingly, thousands of the structures in PDB are now known to contain disordered chains that become structured only in the presence of the partner, and there are also many regions that are actually missing from electron density maps [11]. Such evidence added to that obtained by other techniques [mostly NMR and circular dichroism (CD)], which enabled the creation of DisProt, a database dedicated to structural disorder [11]. DisProt now contains more than 1300 curated IDPs/IDRs; the known resistance to heat- and/or chemical denaturation provided further proteomic-scale evidence for disorder, through mass-spectrometry-based identification of 95 (*Arabidopsis thaliana* [12]) to 1320 (*Mus musculus* [13]) potentially disordered proteins.

The identification of many IDPs also enabled the development of sophisticated bioinformatic algorithms for predicting disorder from sequence, which further advanced the field. Ten years ago, only predictor of natural disordered regions (PONDR) was available [14]; today, one can use any of about 50 predictors, which are based on several different principles [15]. Structural disorder represents different states, limiting the prediction accuracy by any single approach; therefore, predictors have also been recently combined into metapredictors, as exemplified by PONDR-FIT [16]. In addition, the accuracy of predictors is now regularly assessed as part of the critical assessment of structure prediction (CASP) experiment, which shows that the best predictors approach 85% accuracy [17].

Based on predictions, we know that structural disorder is abundant in all species, and due to its strong correlation with regulatory and signaling functions, its level is significantly higher in eukaryotes than in prokaryotes [18]. By conservative estimates, about 10–35% of prokaryotic and about 15–45% of eukaryotic proteins contain significant disorder, that is, long disordered regions at least 30 residues in length [9]; such data are deposited in a prediction-based database of disorder [19]. Although it has become almost commonplace in the field that structural disorder increases with the complexity of the organism, the highest levels are not witnessed in the most complex metazoan eukaryotes (e.g., in humans), but in single-celled eukaryotes that lead a host-changing lifestyle [18].

IDPs are not unstructured but ... pliable?

Although the initial name 'unstructured' implied that IDPs might completely lack structure, it was apparent even 10

Corresponding author: Tompa, P. (ptompa@vub.ac.be).
Keywords: unstructured protein; unstructural biology; structure–function relationship; disorder in disease; drug development.

 http://dx.doi.org/10.1016/j.tibs.2012.08.004

years ago [4] that they have potentially function-related short- and long-range structural organization, which eventually called upon a change in terminology. At that time, high-resolution data were rather limited, thus the concept was mostly phrased from the global structural level, which suggested that IDPs fall into coil-like, pre-molten globule-type and molten-globule types [20].

Structural disorder can now be studied in great detail by several dozen experimental techniques [21], and the most spectacular advance has been achieved through the application of multidimensional NMR. This approach is often complemented by other structural techniques, such as small-angle X-ray scattering (SAXS), which is combined with advanced computational data integration based upon molecular dynamics (MD) simulations (Figure 1). These new approaches [22] enabled the characterization of the full structural ensemble of several dozen IDPs (Table 1). Full resonance assignment of an IDP as long as 441 residues (human τ protein) is now feasible [23], including even proline residues, which has been enabled by carbon-sensitive detection [24]. NMR parameters that are sensitive to local structure, such as chemical shift, heteronuclear Overhauser effect (hetNOE), relaxation, and residual dipolar coupling (RDC) values, can be determined; these can be complemented by long-range structural constraints obtained in paramagnetic relaxation enhancement (PRE) NMR measurements and SAXS experiments. From a large number of possible disordered conformers, a limited number of structural states are then selected [22] to describe experimental data collectively (Table 1). Such ensemble solutions can even be provided in complexed (p53 tetramer [25], Figure 1a) or partner-bound stoichiometric inhibitor of cyclin-dependent kinase 1-B-type cyclin (Cdk1-Clb) (Sic1) bound to Skp1–cullin–F-box ubiquitin ligase (SCF^{Cdc4}) [26], Figure 1b) states. If the calculations are based on MD simulations, we can take into account the energetics and dynamics of the conformational ensemble. For example, such an approach (Table 1) has shown that $p27^{Kip1}$ preferentially samples conformations that are rather similar to its partner-bound state [27], and it has also enabled the mapping of the free energy landscape of the structural ensemble of α-synuclein [28]. Undoubtedly, these structural studies represent the first step toward achieving a quantitative structural description of the function of IDPs.

IDPs are now also studied at the single-molecule level. Single-molecule fluorescence resonance energy transfer (smFRET) measurements, for example, extended previous studies based on NMR, electron microscopy and SAXS, to show multiple long-range interactions between the N-terminal domain and DNA binding domain of p53 [29]. The superior sensitivity of smFRET even enabled the characterization of the dynamics of assembly of soluble NSF-attachment protein receptor (SNARE) complexes in live cells [30]. Atomic-force microscopy (AFM) also enables either the visualization of conformational changes [31] or the study of the energetics and dynamics of the structural ensemble of IDP molecules. For example, mechanical unfolding of α-synuclein molecules by AFM showed that the molecule exists in three major conformations: disordered, soluble oligomeric, and a 'β-like' state [32], which could be informative for amyloid formation in Parkinson's disease.

These and many other studies made the term unstructured obsolete. The distinguishing and unifying feature of these proteins – if any – is their inability to fold into a unique and stable tertiary structure. They display a vast array of function-related structural organization, therefore, the field has settled on the term 'disordered'. This term applies to both short and long regions that are part of a larger folded

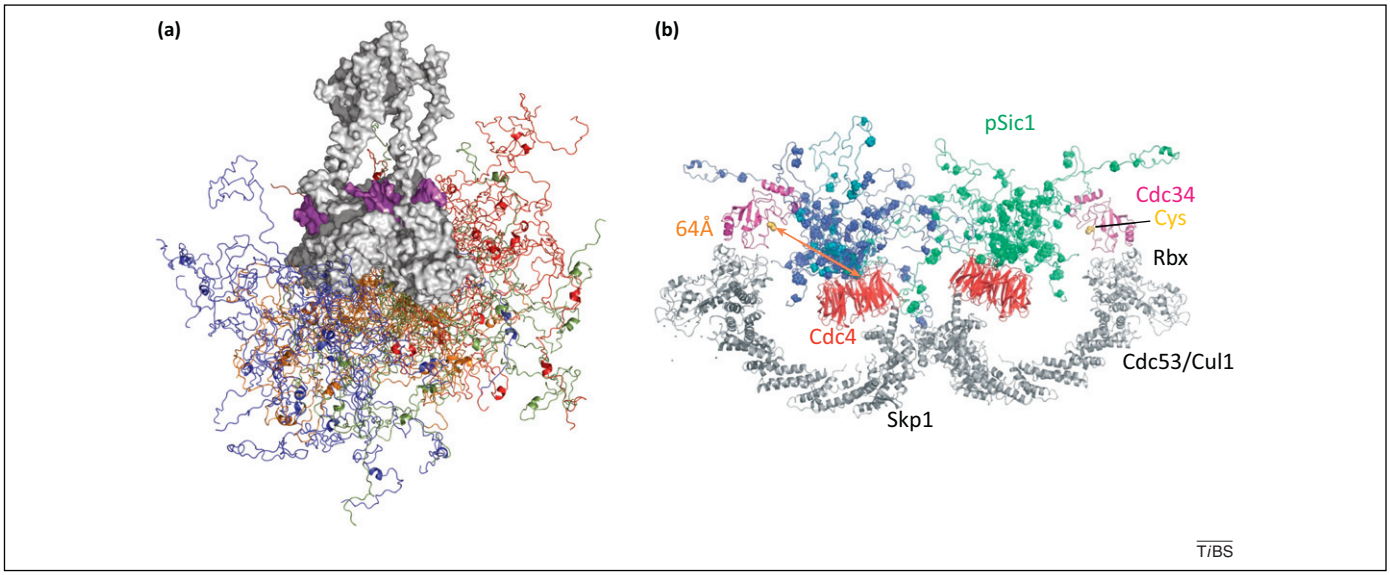

Figure 1. Structures of IDPs without and with their binding partners. **(a)** The structure of p53 was resolved by a combination of several techniques. The protein is a tetramer composed of structured DNA-binding and tetramerization domains (gray space-filling models) and a disordered transactivator domain (shown as an ensemble of 20 conformations in different colors for each molecule in the tetramer). The structure of the folded domains was resolved by X-ray crystallography, whereas the structural ensemble of the transactivator domain was calculated from distance constraints derived from NMR and small-angle X-ray scattering (SAXS) measurements (figure adapted, with permission, from [25]). **(b)** Structural ensemble of stoichiometric inhibitor of Cdk1-Clb (Sic1) cyclin-dependent kinase (Cdk) inhibitor in complex with the SCF^{Cdc4}/cell division cycle 34 (Cdc34) ubiquitin ligase complex [SCF is made up of subunits Skp1, Cdc53 (cullin1) and Rbx; it binds substrate Sic1 via a Cdc4 adaptor subunit and the E2 ubiquitin-conjugating enzyme Cdc34 via its Rbx subunit]. The dynamic ensemble (of two phosphorylated Sic1 molecules, blue and green, bound to the dimer) was calculated from distance constraints obtained by NMR and SAXS, and is superimposed on the structure of the dimer bound to Cdc34 (magenta). Phosphorylated Sic1 binds through cell division cycle 4 (Cdc4) (red); other subunits of the ubiquitin ligase complex are shown in gray (adapted, with permission, from [26]).

Table 1. Select cases for which a structural ensemble of IDPs has been described

Protein	Ensemble method	Parameters	Refs
$p27^{Kip1}$	Molecular dynamics	Chemical shift, R_h (AUC) SAXS	[27]
p53	Flexible-Meccano Accelerated molecular dynamics (AMD)	RDC, SAXS	[25]
Measles virus nucleocapsid	Flexible-Meccano, ASTEROIDS	RDC, chemical shifts, relaxation rates, SAXS/SANS, EM	[80]
Sic1	ENSEMBLE	Chemical shift, ^{15}N R_2, RDC, PRE, SAXS	[26]
α-Synuclein	Molecular dynamics	PRE	[28]
tau protein	XPLOR-NIH	Chemical shift, scalar coupling, RDC, relaxation rate, PRE, NMR diffusion rate	[23]

The structural ensemble of the given protein was determined by the (combination of) methods indicated. The methods mostly rely on parameters derived from NMR and SAXS, such as: RDC, residual dipolar coupling; ^{15}N R_2, amide proton relaxation rate; PRE, distance restraints from paramagnetic relaxation enhancement; Kratky plot from SAXS. Complementing these approaches are small-angle neutron scattering (SANS), dynamic light scattering (DLS), and analytical ultracentrifugation (AUC), providing further hydrodynamic parameters (R_h).

protein (loops and linkers) and proteins that entirely lack a folded structure. This latter category is sometimes also termed (natively) unfolded. Terms such as 'pliable', 'malleable', and others have also been suggested in the literature, but have not gained general recognition.

Disorder exists *in vivo*

A decade ago, a major unsolved mystery of the disorder field was whether structural disorder exists *in vivo*, or if it is an *in vitro* artifact caused by isolation and high dilution of the protein in the test tube. If this were the case, crowding elicited by extreme macromolecular concentrations and/or folding induced by physiological binding partners would actually make them fold *in vivo*.

Since then, several studies have suggested that these factors do not force IDPs to fold (fully) in the cell. *In vitro*, macromolecular crowding can be elicited by up to a 400 mg/ml concentration of Dextran or Ficoll, which does not make plant dehydrins [33] or regulatory IDPs [34] fold. Isotope-labeled proteins can also be studied by multidimensional NMR in live cells. Such in-cell NMR experiments demonstrated that τ protein that has been microinjected into *Xenopus* oocytes [35], or α-synuclein that is overexpressed in *Escherichia coli* cells [36], are largely disordered. Functional studies have provided additional indirect evidence: for example, the chaperone function associated with the structural disorder of late embryogenesis abundant (LEA) proteins *in vitro* can also be observed *in vivo*, attesting to their structural disorder in a living cell [37]. A similar conclusion was also drawn from the noted sensitivity of IDPs to ubiquitin-independent degradation by the 20S proteasome both *in vitro* and *in vivo* [38], a phenomenon termed 'degradation by default'. In all, the evidence seems overwhelming now that structural disorder also exists *in vivo* and it is truly the native, functional state of these proteins.

Functional modes of disordered proteins

The most important question of the field, therefore, is the physiological function and functional mode of IDPs/IDRs. As a result of breathtaking advances in comparative evolutionary and experimental structure–function studies, it is now clear that structural disorder provides multiple functional advantages, and IDP functions either directly stem from their disorder (entropic chains) or from molecular recognition, when they undergo induced folding (disorder-to-order transition) upon binding to a partner molecule [4]. Several of their specific functional modalities, such as adaptability in binding, high functional density, weak but specific binding, and frequent regulation by post-translational modification, have been formally demonstrated. The induced folding concept stimulated much research and led to a virtual explosion in the field of short motifs. It was previously known that many types of protein–protein interactions and enzymatic modifications are mediated by short sequence elements (consensus sites), which gave way to the much broader concept of short linear motifs (SLiMs) or eukaryotic linear motifs (ELMs) [39]. Motifs display extreme evolutionary agility, represent extremely versatile short recognition modules, and are usually bound by recognition domains of the partner molecule (e.g., PxxP motifs are bound by SH3 domains). It was formally shown that most motifs are found in intrinsically disordered regions (IDRs) [40]. On a more structural basis, the related concepts of preformed structural elements (PSEs [41]) and molecular recognition features (MoRFs [42]) are suggested to be short disordered regions that sample structured states within the conformational ensemble, and become fully ordered upon binding to the partner. Binding by motifs is usually weak, transient, and possibly of limited specificity [39], which can be made stronger and/or more specific by either cooperating with flanking regions, combining several motifs, or utilizing longer disordered domains [43,44].

An often-raised issue with respect to IDP binding is whether folding occurs before or after binding (termed conformational selection and induced folding, respectively). This issue has strong parallels with the classical question of the mechanism of protein folding; in that case, it is phrased as diffusion and collision of preformed secondary structural elements in a framework model, or hydrophobic collapse followed by the formation of secondary structure. In the case of IDP binding, detailed studies suggest that folding can occur both before and after binding in distinct cases [45]. For example, NMR ^{15}N relaxation dispersion studies of the binding of the phosphorylated kinase-inducible domain (pKID) of cyclic AMP response element-binding protein (CREB) to the KID-interaction domain (KIX) of CREB-binding protein (CBP) [46] shows that binding proceeds through an ensemble of disordered encounter complexes dominated by nonspecific hydrophobic contacts, only to complete folding in the final bound state.

Two unexpected and important concepts have also emerged in the binding-folding paradigm. One is the intriguing observation that two IDPs/IDRs can bind each other in a process of mutual (synergistic) folding [47]. The other, even more confounding finding, is that recognition can proceed even in the lack of folding in the bound state [48]. The many examples of such 'fuzzy' interactions, such as Sic1 binding to SCF^{Cdc4} [26] (Figure 1b), and up-frameshift factor 2 (UPF2) binding to UPF1 [49], demand the extension of the concept of structural disorder to the bound state.

Irrespective of the actual mechanism of binding, an IDP/IDR might have a greater capture radius than a globular protein and it can bind at a relatively larger distance followed by reeling on to the partner, potentially enhancing the rate of binding by a 'fly-casting' mechanism [50]. This mechanism seems to operate in the assembly of large multiprotein complexes, such as in nonsense-mediated decay (NMD), for example, which is triggered by the productive interaction of UPF proteins. The interaction is initiated by the long-disordered C-terminal domain of UPF2 initially binding UPF1 and bringing various parts of the complex in proximity [49]. A related phenomenon might operate in the spatial search by transcription factors in sequence-specific DNA recognition, termed the 'monkey-bar' binding mechanism [51].

These functional modes and functional advantages demonstrate the involvement of structural disorder in protein function, as approached from a molecular function perspective. Protein function can also be viewed from a biological process point of view, as discussed in the next section.

Function of disordered proteins

The functional role of structural disorder from a biological process view addresses what type of cellular functions benefit most from the lack of a stable structure. This question was addressed in several large bioinformatics studies. As a result, IDPs are generally thought to be involved in processes of signaling and regulation, and in-depth correlation analysis [52] of 710 Swiss-Prot functional keywords suggested significant positive correlation with 238 functions and negative correlation with 302 functions (Table 2). Most of the functions that correlate with the presence of long disordered regions are related to regulation via transcription and translation, whereas functions that correlate with the lack of disorder are dominated by enzymatic catalysis.

Beyond broad correlations, detailed structure–function studies of individual proteins have provided the most convincing demonstration of the mechanistic involvement of disorder in function. Some of the most influential recent studies approach the resolution, accuracy, and detail of traditional studies of ordered proteins, such as enzymes; I present here the most illustrative cases (Table 3).

It is suggested that chaperone function in general correlates with structural disorder [53], and its most detailed experimental demonstration has been presented by virtue of studies of heat shock protein (Hsp)-33; a chaperone that functions in oxidative stress [54]. Hsp-33 has its own redox-sensor domain, which unfolds upon oxidative insult and becomes competent for binding misfolded client proteins.

Table 2. Correlation and anticorrelation of structural disorder with Swiss-Prot functional categories

Top functions that correlate with long disorder[a]	Top functions that anticorrelate with long disorder
Differentiation	GMP biosynthesis
Transcription	Amino acid biosynthesis
Transcription regulation	Transport
Spermatogenesis	Electron transport
DNA condensation	Lipid A biosynthesis
Cell cycle	Aromatic hydrocarbons catabolism
mRNA processing	Glycolysis
mRNA splicing	Purine biosynthesis
Mitosis	Pyrimidine biosynthesis
Apoptosis	Carbohydrate metabolism
Protein transport	Branched-chain amino acid biosynthesis
Meiosis	Lipopolysaccharide biosynthesis

Structural disorder of Swiss-Prot proteins was predicted and the correlation and anticorrelation of proteins with long (≥30 consecutive residues) disordered regions in 710 different biological process functional categories was investigated [52]. The table lists the top 12 functional categories that show significant positive (out of 238) or negative (out of 302) correlation with structural disorder.

[a]The system is not fully orthogonal (e.g., please note overlaps of cell cycle, mitosis, and meiosis).

Through binding and induced folding, Hsp-33 stabilizes more structured folding intermediates, and upon returning to reducing conditions it hands them over to ATP-dependent foldases. Its transitions between ordered and disordered states control substrate binding and release, and thus serve the function of an energy-independent chaperone [53].

An irreparably misfolded protein is tagged for degradation by the ubiquitin-proteasome system (UPS). In a seminal work, it was shown that the yeast nuclear E3 ubiquitin ligase San1 uses intrinsically disordered N- and C-terminal domains for directly recognizing misfolded substrates [55]. Within these IDRs, there are several short, conserved regions of a local hydrophobic nature flanked by very flexible regions, which enable San1 to adaptively recognize many differently shaped misfolded substrates.

Fine tuning of the kinetics and thermodynamics of structural transitions of IDPs between free and bound states also gives rise to complex allosteric signal integration phenomena (Figure 2). Disorder-to-order transitions can optimize allosteric coupling [56], as manifested in signal integration by the allosteric switch Wiskott–Aldrich syndrome protein (WASP), which responds to signals coming from cell division cycle 42 (Cdc42), phosphatidylinositol 4,5-bisphosphate ($PtdInsP_2$), and phosphorylation to induce cytoskeleton reorganization [57]. The protein has a closed, autoinhibited state in which its C-terminal VCA (verprolin, cofilin-homology, and acidic) region is kept quiescent by the N-terminal GTPase-binding domain (GBD). The above signals make the two regions dissociate, and in the active, open state VCA can bind the actin regulatory protein 2/3 (Arp2/3) complex and initiate actin polymerization (Figure 2).

Even more subtle is regulation in bacterial toxin/antitoxin (T/A) systems, in which, depending on T:A stoichiometry, binding of the disordered antitoxin (CcdA or Phd) to the toxin (CcdB or Doc) results in either inhibition of the toxin,

Table 3. The most detailed studies linking structural disorder with protein function

Protein (class)	Function	Role of disorder	Refs
San1	E3 ubiquitin ligase	Recognition of misfolded substrate	[55]
Hsp-33	Redox chaperone	Holding misfolded structure	[54]
E1A	Viral oncoprotein	Binding of host factors CBP and pRb, deregulation of host signaling	[73]
PhD	Bacterial antitoxin	Allosteric regulation of bacterial toxin	[58,59]
Sic1	Cyclin-dependent kinase inhibitor	'Polyelectrostatic' interaction with SCF (Cdc4) ubiquitin ligase	[26]
WASP	Regulator of actin polymerization	Intra- and intermolecular interactions, allosteric regulation	[57]
p27	Cyclin-dependent kinase inhibitor	'Signaling conduit' regulation of targeted degradation	[81]
CREB	General transcription coactivator	Interaction through induced folding by a wide range of transcription factors	[46,47]
LEA proteins	Stress response protein in plants and animals	Chaperone function in abiotic stress via 'entropy transfer' or 'space filling'	[53,82]

allosteric release (rejuvenation) of inhibition, or transition of the T/A operon from a repressed to derepressed state [58,59].

In describing all these functional modes, we approach IDPs by chemical concepts and terms. A recent development in the field addresses IDPs from a physical perspective, via their involvement in phase transitions, which may help bridge the length scales of proteins (angstrom) to that of organelles and cells (micrometers). The interactions between multivalent repetitive disordered proteins, such as WASP and its established biological partners non-catalytic region of tyrosine kinase adaptor protein (NCK) and phosphorylated nephrin, can produce sharp liquid–liquid-demixing phase separations, which generate micrometer-sized liquid droplets [60]. The phase transition depends on the valency of the disordered polymeric chain, which is regulated by post-translational modification, and can alter the activity of the protein. A similar transition to a hydrogel state has also been observed in the low-complexity regions of RNA-binding proteins [61].

Proteome- and systems-level descriptions and understanding of protein disorder have also advanced considerably. As suggested, IDPs/IDRs often function by binding accompanied by induced folding (molecular recognition) [6–8] mediated by SLiMs/ELMs, and in several recent works it has been shown that the functional and evolutionary agility of IDPs can be ascribed to the inclusion or exclusion of such motifs in RNA maturation; that is, by alternative splicing, alternative promoter usage, and RNA editing, the alternative isoforms thus generated promote the functional diversification of the proteome [62]. This mechanism can result in a change in diverse functional attributes, such as subcellular localization, protein–protein interaction, phase transitions and even opposing (dominant negative) function, as also demonstrated by studying tissue-specific forms of alternative splicing. These protein isoforms tend to occupy central positions in interaction networks and their pattern of interaction partners tend to significantly differ [63]; that is, structural disorder and encoded motifs have a strong potential to define and redefine wiring of cellular signaling pathways.

Structural disorder in disease

Ten years ago, we surmised that structural disorder might be involved in diseases [4], primarily because of its enrichment in a few important cases, such as α-synuclein [64]. Structural disorder was then confirmed and/or studied in great detail in many other important disease-associated proteins, such as p53 [25], τ protein [23], and cystic fibrosis transmembrane conductance regulator (CFTR) [65], and was also substantiated by several genome-scale bioinformatics studies. In these, a significant enrichment of structural disorder was found in proteins involved in cancer, neurodegenerative diseases, cardiovascular diseases, and diabetes, leading to the formulation of disorder in disease (D^2) concept [66,67]. A comprehensive bioinformatics analysis of 406 human proteins, such as breakpoint cluster region-Abelson leukemia (Bcr-Abl) and CBP-mixed lineage leukemia (CBP-MLL), which are both generated by chromosomal translocation and gene fusion in cancer [68], substantiated the hypothesis that structural disorder enables the cellular existence of oncogenic protein chimeras. This downside of structural disorder can also appear in dosage sensitivity of genes that cause trouble if overexpressed [69]. This phenomenon might be linked with the binding promiscuity of IDPs, which can be restrained only by their tight regulation at the transcriptional, RNA and protein levels [70].

Perhaps offering a further rationale for such tight regulation, structural disorder is also heavily involved in the formation of insoluble and intractable aggregates, known as amyloids [67]. As a result of the structural exposure of their polypeptide chain, amyloidogenic proteins, which can either cause disease or a heritable advantageous change in phenotype, have a high level of disorder. Gln/Asn-rich proteins are particularly prone to form amyloids: Asn promotes assembly of potentially functional self-templating amyloids, whereas richness in Gln seems to favor the formation of toxic nonamyloid conformers [71].

Structural disorder is also crucial in the action of pathogens. For example, virus entry, replication, and budding are based on deregulating the signaling of the host cell, which is orchestrated through interactions of viral proteins with key host regulatory proteins. In most cases viruses use motif-mimicry for this purpose, that is, short motifs in disordered regions mimicking host protein SLiMs [72]. For instance, adenovirus early region 1A (E1A) oncoprotein transforms host cells by simultaneously recruiting CBP/

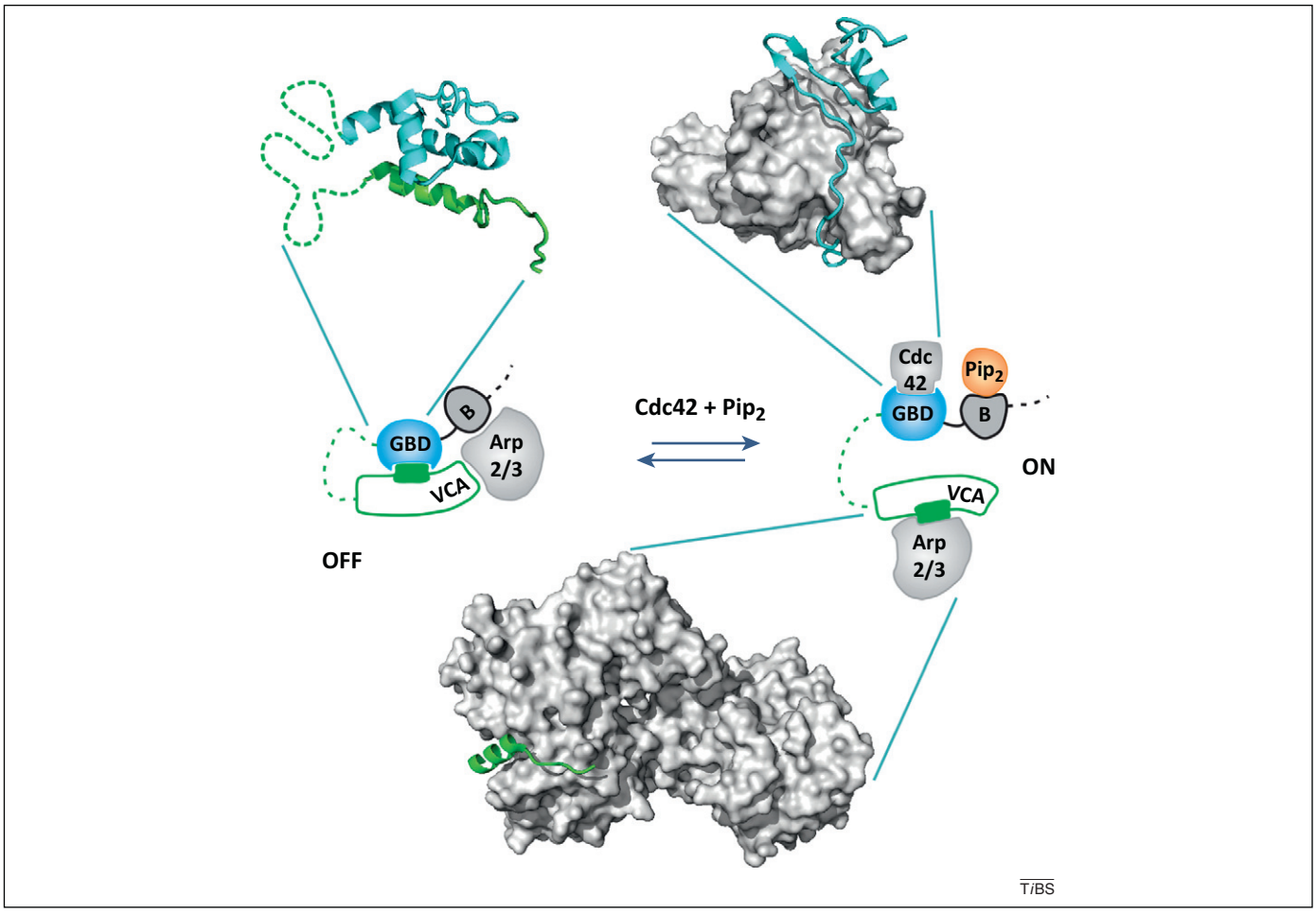

Figure 2. Allosteric regulatory changes in Wiskott–Aldrich syndrome protein (WASP). Structural disorder enables complex signal integration in regulatory proteins, as illustrated by WASP, a modular signaling protein that regulates actin polymerization. In its closed (inactive, OFF) conformation, the autoregulatory regions GTPase binding domain (GBD) (blue) and basic motif (B) bind and inhibit the remote output region, VCA (verprolin, cofilin-homology and acidic) domain (green), connected to them by a long disordered region (broken lines). Input signals cell division cycle 42 (Cdc42) and phosphatidylinositol 4,5-bisphosphate ($PtdInsP_2$) bind GBD and the B motif, respectively, and activate the switch by disrupting autoinhibitory interactions. The VCA domain can then bind to actin regulatory protein 2/3 (Arp2/3) and lead to constitutive activation of actin polymerization (adapted from [57]; PDB structures are: WASP alone: 1ej5, WASP with Cdc42: 1cee, WASP with Arp2/3: 2a3z).

p300 and phosphorylated retinoblastoma protein (pRb) into a tertiary pRb:E1A:CBP complex [73] that overrides the cell-cycle checkpoints of the host. Another outstanding example of the adaptive potential of IDPs in pathogens is represented by the *Mycobacterium tuberculosis* prokaryotic ubiquitin-like protein (Pup), which is analogous but unrelated to eukaryotic ubiquitin. This IDP is used for tagging proteins destined for degradation; pupylated proteins are then recruited by the proteasome-assisted ATPase (Mpa), which differs mechanistically but bears strong analogy to substrate recognition by the ubiquitin–proteasome system [74].

Such detailed structural–functional studies of IDPs promise a more thorough understanding of the cause and development of various disease states. More importantly, they also raise hope of developing remedies against the ensuing adverse conditions.

IDPs in drug development

Thus, the involvement of IDPs in disease makes them prime targets for drug development, which was not entirely apparent a decade ago. Of course, IDPs have no enzymatic activity and thus cannot be attacked in the way traditional drugs function, which usually target active sites or the ligand-binding pockets of enzymes and/or receptors [75]. However, as suggested above, they are often engaged in protein–protein interactions, which might be interfered with via small molecules. The interfaces of IDPs seem well suited for small-molecule interference, because they usually bind their partner through SLiMs/PSEs/MoRFs in a way resembling the binding of substrates and/or inhibitors to the actives sites of enzymes. This IDP partner-targeting approach has been suggested for drug development [76], and has already been demonstrated by the success of nutlins, which inhibit the p53-murine double minute 2 (MDM2) interaction and reactivate the p53 pathway in cancer cells [77]. Although far more difficult to rationalize in structural terms, recently it was also shown that IDPs themselves can be targeted by small molecule interactors, as demonstrated for the oncoprotein c-Myc [78] and amyloid precursor protein (APP) in Alzheimer's disease [79]. However, the potential of this approach in a cellular context still needs to be demonstrated.

Concluding remarks

The field of structural disorder has been – and still is – developing at a rapid pace. Intriguingly, many of the basic concepts such as its prevalence, functional associations, and functional advantages were correctly foreseen, yet the pace of discovery surpassed all expectations. The most notable new developments are: (i) the adaptation of structural techniques to the detailed description of their structural ensemble *in vitro* and *in vivo*; (ii) the advanced state of structural–functional studies of many IDPs at unprecedented accuracy; and (iii) the recognition of their involvement in disease, from which novel drugs can be developed. In all, the field has matured to take a strong stand in molecular structural biology, and it is almost evident that, by the turn of the next decade, unstructural biology will take its due place in mainstream biochemistry and molecular biology textbooks.

Acknowledgements

This work was supported by the Odysseus grant G.0029.12 from the Research Foundation Flanders (FWO). The help of Dr Simone Kosol in preparing the figures is highly appreciated.

References

1 Dunker, A.K. *et al.* (2001) Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26–59
2 Uversky, V.N. *et al.* (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41, 415–427
3 Wright, P.E. and Dyson, H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331
4 Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27, 527–533
5 Tompa, P. (2011) Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* 21, 419–425
6 Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208
7 Uversky, V.N. and Dunker, A.K. (2010) Understanding protein non-folding. *Biochim. Biophys. Acta* 1804, 1231–1264
8 Dyson, H.J. (2011) Expanding the proteome: disordered and alternatively folded proteins. *Q. Rev. Biophys.* 44, 467–518
9 Tompa, P. (2009) *Structure and Function of Intrinsically Disordered Proteins*, CRC Press
10 Berman, H.M. *et al.* (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* 10, 98
11 Sickmeier, M. *et al.* (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* 35, D786–D793
12 Irar, S. *et al.* (2006) Towards the identification of late-embryogenic-abundant phosphoproteome in *Arabidopsis* by 2-DE and MS. *Proteomics* 6 (Suppl. 1), S175–S185
13 Galea, C.A. *et al.* (2009) Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *J. Proteome Res.* 8, 211–226
14 Garner, E. *et al.* (1998) Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization. *Genome Inform. Ser. Workshop Genome Inform.* 9, 201–213
15 He, B. *et al.* (2009) Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 19, 929–949
16 Xue, B. *et al.* (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* 1804, 996–1010
17 Monastyrskyy, B. *et al.* (2011) Evaluation of disorder predictions in CASP9. *Proteins* 79 (Suppl. 10), 107–118
18 Pancsa, R. and Tompa, P. (2012) Structural disorder in eukaryotes. *PLoS ONE* 7, e34687
19 Fukuchi, S. *et al.* (2012) IDEAL: Intrinsically Disordered proteins with Extensive Annotations and Literature. *Nucleic Acids Res.* 40, D507–D511
20 Uversky, V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11, 739–756
21 Eliezer, D. (2009) Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 19, 23–30
22 Fisher, C.K. and Stultz, C.M. (2011) Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 21, 426–431
23 Mukrasch, M.D. *et al.* (2009) Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* 7, e34
24 Bermel, W. *et al.* (2009) Speeding up (13)C direct detection biomolecular NMR spectroscopy. *J. Am. Chem. Soc.* 131, 15339–15345
25 Wells, M. *et al.* (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5762–5767
26 Mittag, T. *et al.* (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* 18, 494–506
27 Sivakolundu, S.G. *et al.* (2005) Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J. Mol. Biol.* 353, 1118–1128
28 Allison, J.R. *et al.* (2009) Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J. Am. Chem. Soc.* 131, 18314–18326
29 Huang, F. *et al.* (2009) Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20758–20763
30 Sakon, J.J. and Weninger, K.R. (2010) Detecting the conformation of individual proteins in live cells. *Nat. Methods* 7, 203–205
31 Kodera, N. *et al.* (2010) Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature* 468, 72–76
32 Sandal, M. *et al.* (2008) Conformational equilibria in monomeric alpha-synuclein at the single-molecule level. *PLoS Biol.* 6, e6
33 Mouillon, J.M. *et al.* (2008) Mimicking the plant cell interior under water stress by macromolecular crowding: disordered dehydrin proteins are highly resistant to structural collapse. *Plant Physiol.* 148, 1925–1937
34 Szasz, C.S. *et al.* (2011) Protein disorder prevails under crowded conditions. *Biochemistry* 50, 5834–5844
35 Bodart, J.F. *et al.* (2008) NMR observation of Tau in *Xenopus* oocytes. *J. Magn. Reson.* 192, 252–257
36 McNulty, B.C. *et al.* (2006) Macromolecular crowding in the *Escherichia coli* periplasm maintains alpha-synuclein disorder. *J. Mol. Biol.* 355, 893–897
37 Chakrabortee, S. *et al.* (2010) Catalytic and chaperone-like functions in an intrinsically disordered protein associated with desiccation tolerance. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16084–16089
38 Tsvetkov, P. *et al.* (2008) Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. *Proteins* 70, 1357–1366
39 Gould, C.M. *et al.* (2009) ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.* 38, D167–D180
40 Fuxreiter, M. *et al.* (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 23, 950–956
41 Fuxreiter, M. *et al.* (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* 338, 1015–1026
42 Mohan, A. *et al.* (2006) Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* 362, 1043–1059
43 Stein, A. and Aloy, P. (2008) Contextual specificity in peptide-mediated protein interactions. *PLoS ONE* 3, e2524
44 Tompa, P. *et al.* (2009) Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays* 31, 328–335
45 Onitsuka, M. *et al.* (2008) Mechanism of induced folding: both folding before binding and binding before folding can be realized in staphylococcal nuclease mutants. *Proteins* 72, 837–847
46 Sugase, K. *et al.* (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447, 1021–1025
47 Demarest, S.J. *et al.* (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415, 549–553
48 Tompa, P. and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33, 2–8
49 Clerici, M. *et al.* (2009) Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.* 28, 2293–2306

50 Shoemaker, B.A. *et al.* (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8868–8873
51 Vuzman, D. *et al.* (2010) Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J. Mol. Biol.* 396, 674–684
52 Xie, H. *et al.* (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* 6, 1882–1898
53 Tompa, P. and Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* 18, 1169–1175
54 Reichmann, D. *et al.* (2012) Order out of disorder: working cycle of an intrinsically unfolded chaperone. *Cell* 148, 947–957
55 Rosenbaum, J.C. *et al.* (2011) Disorder targets misorder in nuclear quality control degradation: a disordered ubiquitin ligase directly recognizes its misfolded substrates. *Mol. Cell* 41, 93–106
56 Hilser, V.J. and Thompson, E.B. (2007) Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8311–8315
57 Leung, D.W. and Rosen, M.K. (2005) The nucleotide switch in Cdc42 modulates coupling between the GTPase-binding and allosteric equilibria of Wiskott-Aldrich syndrome protein. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5685–5690
58 De Jonge, N. *et al.* (2009) Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. *Mol. Cell* 35, 154–163
59 Garcia-Pino, A. *et al.* (2010) Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* 142, 101–111
60 Li, P. *et al.* (2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483, 336–340
61 Kato, M. *et al.* (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149, 753–767
62 Weatheritt, R.J. and Gibson, T.J. (2012) Linear motifs: lost in (pre)translation. *Trends Biochem. Sci.* 37, 333–341
63 Buljan, M. *et al.* (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* 46, 871–883
64 Weinreb, P.H. *et al.* (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35, 13709–13715
65 Baker, J.M. *et al.* (2007) CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat.Struct. Mol. Biol.* 14, 738–745
66 Uversky, V.N. *et al.* (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* 37, 215–246
67 Raychaudhuri, S. *et al.* (2009) The role of intrinsically unstructured proteins in neurodegenerative diseases. *PLoS ONE* 4, e5566
68 Hegyi, H. *et al.* (2009) Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins. *PLoS Comput. Biol.* 5, e1000552
69 Vavouri, T. *et al.* (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* 138, 198–208
70 Gsponer, J. *et al.* (2008) Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* 322, 1365–1368
71 Halfmann, R. *et al.* (2011) Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. *Mol. Cell* 43, 72–84
72 Davey, N.E. *et al.* (2011) How viruses hijack cell regulation. *Trends Biochem. Sci.* 36, 159–169
73 Ferreon, J.C. *et al.* (2009) Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13260–13265
74 Pearce, M.J. *et al.* (2008) Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science* 322, 1104–1107
75 Cohen, M.S. *et al.* (2005) Structural bioinformatics-based design of selective, irreversible kinase inhibitors. *Science* 308, 1318–1321
76 Cheng, Y. *et al.* (2006) Rational drug design via intrinsically disordered protein. *Trends Biotechnol.* 24, 435–442
77 Vassilev, L.T. *et al.* (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303, 844–848
78 Metallo, S.J. (2010) Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* 14, 481–488
79 Kukar, T.L. *et al.* (2008) Substrate-targeting gamma-secretase modulators. *Nature* 453, 925–929
80 Jensen, M.R. *et al.* (2011) Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9839–9844
81 Grimmler, M. *et al.* (2007) Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases. *Cell* 128, 269–280
82 Chakrabortee, S. *et al.* (2007) Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18073–18078