FEBS Letters 579 (2005) 3346–3354 FEBS 29507

Minireview

The interplay between structure and function in intrinsically unstructured proteins

Peter Tompa*

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, P.O. Box 7, H-1518 Budapest, Hungary

Accepted 29 March 2005

Available online 8 April 2005

Edited by Péter Friedrich

Abstract Intrinsically unstructured proteins (IUPs) are common in various proteomes and occupy a unique structural and functional niche in which function is directly linked to structural disorder. The evidence that these proteins exist without a well-defined folded structure in vitro is compelling, and justifies considering them a separate class within the protein world. In this paper, novel advances in the rapidly advancing field of IUPs are reviewed, with the major attention directed to the evidence of their unfolded character in vivo, the interplay of their residual structure and their various functional modes and the functional benefits their malleable structural state provides. Via all these details, it is demonstrated that in only a couple of years after its conception, the idea of protein disorder has already come of age and transformed our basic concepts of protein structure and function.

Keywords: Natively unfolded protein; Intrinsically disordered protein; Protein disorder in vivo; Functional classification; Residual structure

1. Introduction

Our traditional view of protein structure–function relationship is rooted in the notion that function critically depends on a well-defined 3D structure. In a recent surge of reports, however, it has been shown that for many proteins and protein domains the functional state is intrinsically unstructured. Sporadic data go back to more than a decade [1–3] but it has only recently been that the generality of the phenomenon was noted [4]. Since then, the field is in a steady progress, as attested by many individual examples and numerous reviews [5–12]. The structure of intrinsically unstructured proteins (IUPs) resembles the denatured states of ordered proteins, best described as an ensemble of rapidly interconverting alternative structures characterized by differing backbone torsion angles. By bioinformatic estimations, these proteins are common in various proteomes and their frequency increases with increasing complexity of the organisms [7,13,14]. The functional importance of protein disorder is also underscored by that it dominates in proteins associated with signal transduction, cell-cycle regulation, gene expression and chaperone action [7,14–17]. The widespread occurrence and importance of these proteins has called for re-assessing the classical structure–function paradigm [4]. The field of protein disorder is already too wide to be covered in a single review. Thus, I survey herein some of the most interesting recent developments with respect to the evidence of the unfolded character of IUPs in vivo, their distinct and unique functional modes, the functional implications of their residual structure and the functional benefits structural disorder, as opposed to order, provides.

2. Disorder is the native state of IUPs

For almost 200 proteins and protein domains [18], the lack of a unique 3D structure has been convincingly demonstrated by using three techniques mostly, X-ray crystallography, multidimensional nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy. In some cases this evidence is complemented by other techniques, such as Fourier-transformed infrared spectroscopy (FTIR) and Raman optical activity (ROA) spectroscopy, hydrodynamic techniques (small angle X-ray scattering, ultracentrifugation and gel-filtration), differential scanning calorimetry and some indirect approaches, such as proteolytic sensitivity, heat stability and anomalous sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) mobility [7–10,19].

The vast majority of this evidence, however, has come from studying IUPs in highly diluted solutions in vitro. This may cast doubt on their disorder in vivo, as the crowding effect elicited by extreme macromolecular concentrations (up to 400 mg/ml) in living cells may significantly shift their conformational

*Fax: +361 466 5465.
E-mail address: tompa@enzim.hu (P. Tompa).

Abbreviations: CBP, CREB-binding protein; CD, circular dichroism; CREB, cAMP response element binding protein; Cdk, cyclin-dependent kinase; CST, calpastatin; DHPR, dihydropyridine receptor; FT-IR, Fourier-transformed infrared spectroscopy; IUP, intrinsically unstructured protein; KID, kinase-inducible domain; MAP2, microtubule-associated protein 2; MoRE, molecular recognition element; NACP, non-A beta component of Alzheimer's disease amyloid plaque (also termed α-synuclein); NMR, nuclear magnetic resonance; PCS, primary contact site; PEVK, region rich in Pro, Glu, Val and Lys; PP II, polyproline II helix; RNAP II, DNA-dependent RNA polymerase II; ROA, Raman optical activity; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis

0014-5793/$30.00
doi:10.1016/j.febslet.2005.03.072

equilibrium towards a folded state [20]. In fact, the intrinsically unstructured [21] inhibitor of the transcription factor sigma28, when expressed in *Escherichia coli*, undergoes significant ordering, as demonstrated by NMR [22]. Although such a tendency to get ordered in vivo could be a general feature of IUPs, an array of considerations warrant that their function is intimately linked with their lack of a compact fold in vivo.

A prime argument that IUPs basically differ from globular proteins in vivo relates to the predictability of structural disorder from sequence. It is evident that IUPs identified in vitro have a distinct amino acid composition, in that they are enriched in disorder-promoting amino acids (A, R, G, Q, S, P, E and K) and depleted in order-promoting amino acids (W, C, F, I, Y, V, L and N) [6]. Other manifestations of this distinct character is that they are usually characterized by a high net charge and low mean hydrophobicity [5] and their amino acid composition inversely correlates with β-aggregation propensity [23]. Based on these sequence attributes, a range of bioinformatic predictors, such as PONDR [10], DISOPRED [14] and GLOBPLOT [24] have been developed. These predictors perform at a level comparable to the best secondary structure prediction algorithms. A different algorithm, IUPred [25], estimates the total pairwise interresidue interaction energy of sequences, which is significantly smaller for IUPs than for globular controls. As this predictor has not been trained to recognize disordered sequences, its correct assessment of IUPs substantiates that the lack of a stable structure is their *intrinsic* property. In all, the success of disorder predictors confirms that IUPs are basically different from ordered proteins, i.e. their anomalous structural behavior is not an in vitro artefact.

Another point to make is that the question of a crowding-induced compact fold in vivo is irrelevant with extracellular IUPs, which by definition do not experience a crowded environment under physiological conditions. The best-studied examples are milk casein(s), salivary proline-rich glycoproteins and bacterial fibronectin-binding proteins [9]. In addition, direct structural studies have been conducted for some IUPs under crowded conditions. In these, evidence is mostly against overall folding with only a marginal tendency to form structure [26–28]. Consistent with this limited tendency to adopt structure is that IUPs are not fully unstructured [15] but contain local recognition elements of appreciable tendency to be preorganized [29], which may gain significant stability under crowding [22,30].

The entire issue of structural organization can also be largely dismissed for those IUPs, for which function directly stems from the disordered state and thus in vivo foldedness is out of question (entropic chains [7,9], cf. Fig. 1 and Table 1). By definition, their function cannot be fulfilled by a rigid structure but it is associated with the ability of the polypeptide chain to rapidly fluctuate among alternative states in a conformational ensemble. The region rich in Pro, Glu, Val and Lys (PEVK) in titin, an entropic spring in muscle [31], the projection domain of microtubule-associated protein 2 (MAP2) [32], an entropic bristle that provides spacing in the cytoskeleton and the FG repeat region of nucleoporins, which regulate transport through the nuclear pore complex via spatial exclusion and specific recognition of transport proteins [33], exemplify this behavior.

An additional argument against a compact structure of IUPs in vivo comes from their mode of binding to their partners. Most often, IUPs function by molecular recognition, i.e. via transient or permanent binding to a structured partner (Fig. 1 and Table 1). In several cases (cf. [29]) the extended, open, structure of the IUP in the bound state is known (Fig. 2). For these proteins, assuming a compact state prior to binding that had to unfold to adopt the structure seen in the complex makes no sense. Pertinent to this point is that certain complexes simply cannot be assembled from rigid components due to topological constraints: the IUP wraps around its partner and thus its flexibility is inevitable to reach the final state (Fig. 2). Furthermore, some IUPs can bind several different partners in a process termed binding promiscuity [34] or one-to-many signaling [6] and it has been suggested that the IUP in these cases may adopt different structures. This structural malleability has actually been demonstrated for the C-terminal domain of DNA-dependent RNA polymerase II (RNAP II) (Table 1) bound to either RNA guanylyl transferase Cgt1 or peptidyl-proline isomerase Pin1 [35] and the HIF-1α-interaction domain bound to either the TAZ1 domain of cAMP response element binding protein (CREB)-binding protein (CBP) [36] or the asparagine hydroxylase FIH [37]. This behavior is incompatible with a unique structure (Fig. 2).

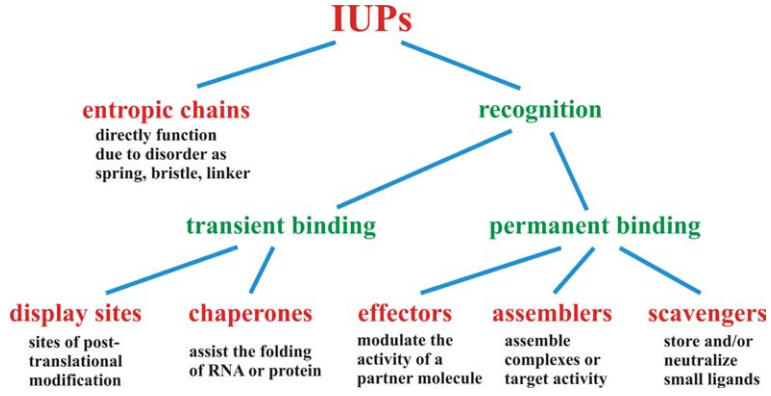

Fig. 1. Functional classification scheme of IUPs. The function of IUPs stems either directly from their capacity to fluctuate freely about a large configurational space (entropic chain functions) or ability to transiently or permanently bind partner molecule(s). For each functional class, a short definition of function is given. More extended description and examples are found in the text, Table 1 and the references cited.

Table 1
Functional classification of IUPs

Protein (IUP)	Function	Structure in complex (PDB)
Entropic chains		
Neurofilament-H KSP domain	Entropic bristle (spacing in neurofilament lattice)	
Nup2p FG repeat region	Gating in nuclear pore complex	
tau/MAP2 projection domain	Entropic bristle (spacing in cytoskeleton)	MAP tau – Pin1 WW (1I8H)
Titin PEVK domain	Entropic spring (passive force in muscle)	
K channel N-terminal region	Entropic clock/inactivation gate	
Display sites		
SNAP-25	Cleavage by neurotoxin	SNAP-25 – BoNT/A (1XTG)
CREB KID	Regulation by phosphorylation	CREB KID – CBP KIX (1KDX)
MAP2 microtubule-binding domain	Regulation by phosphorylation	
Casein	Turnover by proteasome	
tau	Turnover by proteasome	
α-Synuclein (NACP)	Turnover by proteasome	
Cyclin B N-terminal domain	Ubiquitination	
Chaperones		
α-Synuclein (NACP)	Protein chaperone	
Casein	Protein chaperone	
Nucleocapsid protein 7/9	RNA chaperone	
Ribosomal S12	RNA chaperone	
Prion protein N-terminal domain	RNA chaperone	
Effectors		
CITED2	Regulation of hypoxic response	CITED2 – CBP TAZ1 (1P4Q)
Securin	Inhibition/activation of separase in anaphase	
Calpastatin	Inhibition/activation of calpain	Calpastatin – calpain (1NX0)
$p21^{Cip1}/27^{Kip1}$	Inhibition/activation of cyclin-dependent kinases	p27 – CycA/Cdk2 (1JSU)
4EBP1	Inhibitor of eukaryotic translation initiation	
PKI	Inhibition of cAMP-dependent protein kinase	PKI – PKA (1APM)
PP I1, DARPP32	Inhibition of phosphorylase phosphatase	
FlgM	Inhibition of sigma28 transcription factor	FlgM – sigma28 (1RP3, 1SC5)
Stathmin/RB3	Microtubule disassembly	RB3 – tubulin (1FFX)
IA_3	Inhibition of aspartic proteinase A	IA_3-proteinase A (1DPJ)
DHPR II–III loop C fragment	Inhibition/activation of ryanodine receptor	
Assemblers		
Caldesmon	Actin polymerization, bundling	
Bob1	B-cell specific expression of Ig genes	
L7/L12	ribosome assembly/stability	
FnBP	Adherence to fibronectin in bacterial invasion	FnBPA – fibronectin (1O9A)
CREB trans-activator domain	Assembly of transcription preinitiation complex	CREB KID – CBP KIX (1KDX)
E-cadherin intracellular domain	Signaling in cell adhesion	E-cadherin – β-catenin (1I7X)
p53	Tumor suppressor transcription factor	p53 – MDM2 (1YCQ)
RNAP II C-terminal domain	Transcription of protein-coding genes	RNAP II CTD – mRNA capping enzyme Cgt1 (1P16)
SV40 virus coat protein	Virus assembly	SV40 coat (1SVA)
Tcf3/4	T-cell specific transcription factor	Tcf3/4 – β-catenin (1G3J, 1JPW)
SARA SBD	Smad anchoring to TGF receptor	SARA SBD – Smad2 MH2 (1DEV)
Ciboulot	Actin polymerization/assembly	Ciboulot – G actin (1SQK)
Lambda N	Translation antitermination	Lambda N – NusA (1U9L)
Thymosin β	Actin polymerization/assembly	Thymosin β-G actin (1T44)
HIF-1α	Regulation of hypoxic response	HIF-1α – CBP TAZ1 (1L8C), HIF-1α – FIH (1H2K)
Measles virus nucleoprotein C-terminal domain	Template for RNA synthesis	Nucleoprotein – phosphoprotein (1T6O)
$p21^{Cip1}$	Assembly of cyclin-Cdk complex	p21 – PCNA (1AXC)
Scavengers		
Casein	Inhibition of calcium phosphate precipitation in milk	
Salivary proline-rich glycoprotein	Neutralization of plant tannins	
Desiccation stress protein (Dsp) 16	Water retention in dehydration	

IUPs can be classified in terms of their functional modes into six broad categories, as put forward in [9,17]. A limited set of examples is shown here, further cases can be found in the original references. The physiological function of the proteins is given, and their structure bound to a partner is referred to. It is of note that most structures are known for effectors and assemblers, which function via permanent binding to partner molecule(s).

An indirect observation contrasting a compact, folded structure in vivo is the high evolutionary rate of IUPs (cf. also [7]). Evolutionary changes in sequence are limited by constraints on residues involved in functional/structural interactions, which keep the level of non-synonymous (K_A) vs. synonymous (K_S) mutations in such regions low, on

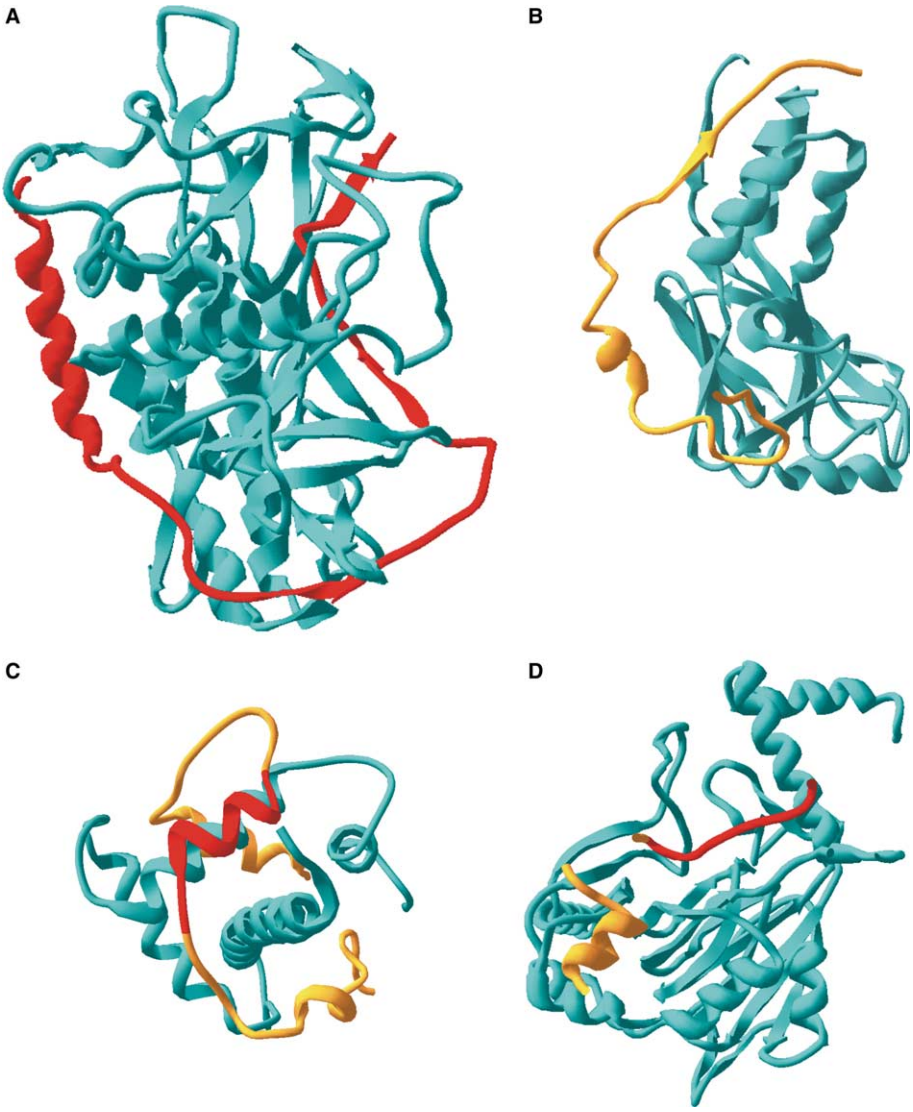

Fig. 2. Some IUPs wrap around their partner upon binding. For some IUPs, the structure in the complexed state is known from X-ray crystallography or NMR (IUP shown in yellow or red). The structures (PDB code in parenthesis) shown are: (A) SNAP-25 bound to BoNT/A (1XTG); (B) SARA SBD domain bound to Smad2 MH2 domain (1DEV), (C) HIF-1α interaction domain bound to the TAZ1 domain of CBP (1L8C) and (D) HIF-1α interaction domain bound to asparagine hydroxylase FIH (1H2K). Please note that the region of HIF-1α interaction domain, which adopts a different structure in the two complexes, is marked in red. Further structures of bound IUPs are referred to in Table 1. The structures have been visualized by the Swiss-PDB viewer.

the order of 0.1–0.2 [38]. In a recent comparative study, the pairwise genetic distances within disordered (IUP) and ordered regions of 26 protein families was found to differ significantly, disordered regions evolving significantly faster in 19 families, and more slowly in 2 families only. For the sex-determining transcription factor, SRY, its Gln-rich transactivator domain evolves much faster (K_A/K_S = 0.4–0.8) than its globular DNA-binding domain (K_A/K_S = 0.1–0.2) [39,40]. Casein has also been noted for its anomalous evolutionary behavior, as its translated region has much higher mutation rate than its non-translated region. This apparent contradiction strongly argues against significant structural constraints in this IUP (cf. [3]). Overall, these IUPs are subject to much less structural constraints in their native state than their structured counterparts, i.e. they, by all probability, lack a well-defined structure in vivo.

3. Functional modes that benefit from structural disorder

Thus, a large body of evidence supports that IUPs do lack a 3D structure in vivo. In the following chapter, it will be shown that they not only tolerate this structural state but structural disorder actually predisposes them for special functional modes in which they take advantage of it. In general, their function either directly stems from the protein's ability to fluctuate over an ensemble of structural states, or it is realized via binding to one or several partner molecule(s) in a structurally adaptive process (Fig. 1 and Table 1). These functional capacities are exploited in many molecular settings and thus IUPs may fulfil many different functions [7]. Functional disorder has been noted in proteins that can bind RNA, DNA, other protein(s) or even small ligands. It has also been observed that disorder correlates with the sites of post-translational modification, such

as phosphorylation [41] and ubiquitination/proteasomal degradation [42,43]. Predictions in various functional classes of proteins have shown that disorder is primarily associated with signal transduction, cell-cycle regulation and gene expression [14,16] and thus it is often implicated in cancer [16]. Recent studies have unveiled the high incidence and functional importance of disorder in endocytosis [44] and in RNA- and protein chaperones [17]. By considering unifying mechanistic details of their various modes of action, the many different functions of IUPs actually segregate into only six general categories [9,17]. Although novel IUPs are identified regularly, this classification scheme (Fig. 1 and Table 1) appears suited to accommodate most examples known today [18].

The first general functional class of IUPs is that of *entropic chains*, the function of which stems directly from their ensemble of structural states of similar conformational energies. These proteins subclassified as entropic springs, bristles/spacers, linkers, clocks, etc. either generate force against structural changes or influence the orientation/localization of attached domains [7].

In the other five classes, IUPs function via molecular recognition, i.e. they permanently or transiently bind another macromolecule or small ligand(s). Of those transiently binding their partner(s), *display sites* function in post-translational modification. It is dictated by common sense that the action of a modifying enzyme requires flexibility of the substrate, which enables transient but specific interaction with the active site of the modifying enzyme. Pertinent to this function is the success of disorder-based prediction of phosphorylation sites [41] and an array of recent observations of the cleavage of non-ubiquitinated, disordered proteins, such as casein [45], tau [46] and $p21^{Cip1}$ [47] by the 20S proteasome and the disorder-targeted ubiquitination of securin and cyclin B in cell cycle regulation [42]. A novel subclass within this category is *chaperones*, as unveiled by a recent statistical analysis. It was found that RNA chaperones have a much higher incidence of disorder than any other functional class: 40% of their residues fall into long disordered regions (>30 residues), whereas the same number is 15% for protein chaperones [17]. Further, the function of many, or possibly all, of these proteins depends directly on disorder in a way that the disordered segment serves for either recognizing, solubilizing or loosening the structure of the misfolded ligand. To account for these mechanistic details, an entropy transfer model of disorder in chaperone function has been suggested [17].

Disordered proteins that function by permanent partner binding belong to either of the three classes of *effectors, assemblers and scavengers*. *Effectors* bind and modify the activity of their partner enzyme [9]. Their action is mostly inhibitory, but in light of recent data they may also activate another protein, demonstrating their extreme structural and functional versatility. The classical effector protein [9], $p21^{Cip1}$ and its homologue, $p27^{Kip2}$ have been shown recently not only to inhibit cyclin-dependent kinases (Cdks), but also to be able to assemble the cyclin-Cdk complex leading to Cdk activation [48]. Another such ambiguous example is the disordered C fragment of dihydropyridine receptor (DHPR) II–III loop, which can bind the ryanodine receptor in two distinct conformations, one inhibiting but the other activating it [49]. Calpastatin (CST), the disordered inhibitor of the calcium-activated protease, calpain, can be fragmented in a way that converts it from a potent inhibitor to an activator of the enzyme [50].

The next class is that of *assemblers*, which assemble multiprotein complexes and/or target the activity of attached domains [9]. Such proteins/domains have been noted in the assembly of the ribosome, cytoskeleton, transcription preinitiation complex and the chromatin, for example. The unusual complexity of interaction networks supported by such disordered assembly domains have been recently demonstrated within the partners of CBP, a multidomain transcription coactivator, which forms complexes with a variety of partners [11].

The third subclass within this category, *scavengers*, store and/or neutralize small ligands. The classical examples of this mode of action are casein(s), which prevent calcium phosphate precipitation in the milk by capturing small seeds as they form and salivary proline-rich glycoproteins, which form tight complexes with tannins that can resist harsh conditions encountered in the digestive tract (cf. [9]).

In general, this classification scheme appears suitable for systemizing the diverse functional modes of IUPs. Its notable aspect is that the various functional modes are not exclusive as different domains within the same protein, or even the very same region, may be involved in distinct functional modes. For example, as shown, $p21^{Cip1}$/$p27^{Kip1}$ may both inhibit and activate Cdk(s), via either an effector or assembler mechanism. The effector securin (Table 1) is an inhibitor of separase, but it is also required for the activation of the enzyme via a chaperone-like action [51]. As a final example, one might recall the HIF-1α interaction domain, which can alternatively bind to the TAZ1 domain of CBP in an assembly function, but also in a different conformation to the active site of asparagine hydroxylase FIH as a display site [11].

4. Function-related structural organization in IUPs

Their remarkable functional diversity and occasional ambiguity, combined with an exceptional specificity [5–12], raise doubts with respect to the fully disordered nature of IUPs. As limited structural data implied initially their lack of secondary and tertiary structure, prior to recognizing their functional importance their anomalous behavior has simply been equated with a complete lack of structural order [52,53]. In light of rapidly accruing data on the structure of these proteins, however, this simplistic view is no longer tenable. To explain the highly specialized and elaborate functional modes of IUPs [9], their significant, and often function-related, residual structure needs to be invoked.

A small amount of repetitive secondary structure is evident upon deconvoluting the CD spectra of many IUPs: α and/or β structure on the order of 10–20% has been ascertained in caseins [3], α-synuclein (NACP) [54], stathmin [55], $p21^{Cip1}$ [34], CST [56] and CREB kinase-inducible domain (KID) [57], for example. Also indicative of structural order is if the spectra of their fragments are not additive due to long-range interactions, as for stathmin [55] and CST [58], or the CD spectrum shifts toward the random-coil state upon heating/denaturation, such as for caseins [3,54] and CST [56]. Some secondary structure is also shown by FTIR for NACP [53]. In certain cases, the presence of polyproline II (PP II) helix, an extended and fully hydrated secondary-structural motif often implicated in molecular recognition [59], can also be inferred from the CD spectrum. There are clear signs of this

motif in tau (25%) [60], casein (23%) [61], stathmin [55], MAP2 [58] and RNAP II [62], for example. ROA measurements have also shown the presence of PP II helix in some IUPs, such as casein, NACP, tau [63] and wheat gluten [64]. NMR, the single most powerful technique for studying the structure and dynamics of IUPs, provides structural information via secondary chemical shift, residual dipolar coupling and long-range NOE upon spin labeling [19]. The application of these techniques has revealed sequence-specific transient secondary structures in FlgM [21], CREB KID [65] and $p27^{Kip1}$ [66], amongst others.

Thus, many IUPs exhibit significant and potentially functional structural organization. The generality of this issue has been addressed by Uversky, who compiled a great deal of relevant hydrodynamic and CD data. IUPs have been shown to fall into coil-like and a premolten-globule-like classes, with significant residual structure in the latter [8]. This observation has been interpreted in terms of a Protein Quartet model, which states that proteins may exist in any of four alternative conformational states, ordered, molten globule, premolten globule and random coil, and function stems from any of these or their interconversion [8]. This model is an extension of the previous Protein Trinity proposal [67] that reckoned with ordered, molten globule and random coil states in a similar manner.

Unexpectedly, the issue of the role of residual structure in IUP function can also be approached by limited proteolysis. This technique is traditionally used to probe the topology of globular proteins and their folding intermediates [68], as proteases generally attack spatially exposed and flexible sites. Under conditions of extremely low protease concentrations, however, IUPs also undergo limited proteolysis, which implies their non-random structural organization. As seen for caldesmon [69], CREB KID [57], stathmin [70] and recently for BRCA1 [71], MAP2 and CST [58], the location of the preferential cleavage site(s) correlate with their domain organization. An appealing interpretation of this observation is that transient short- and/or long-range structural organization ensures the spatial exposure of certain regions in these IUPs. This is of particular relevance for their binding functions as the large-scale binding-coupled folding of IUPs is hardly compatible with a fully disordered structure prior to binding. Rather, it may be anticipated that IUPs exploit some sort of structural preorganization in effectively recognizing their partner and initiating the subsequent induced folding process. In fact, such a mode of action has been suggested for FlgM [21], CREB KID [65], GCN4 [72], CST [58,73] and MAP2 [58], for example.

To approach the issue of structural preorganization, the actual bound structures (cf. Table 1) have been compared to the inherent structural preferences of IUPs, assessed by secondary-structure predictions [29]. It was shown that the prediction accuracy of IUP structures is commensurable with that of their ordered partners, which suggests a strong preference of IUPs for the structure they adopt in the bound state. This implies the presence of preformed structural elements, which may limit the conformational search accompanying folding. A special case of such elements is termed *primary contact sites* (PCSs) [58], i.e. structurally primed, exposed recognition motifs that dock to the partner and lead to the formation of a native-like encounter complex. The presence of such sites has been inferred in MAP2 and CST and suggested in several other IUPs [58]. These sites are conceptually closely related to anchor sites thus far reported for globular proteins [74], molecular recognition elements (MoREs) associated with short ordered motifs apparent in disorder patterns [10] and hot spots also implicated in protein–protein interactions [75]. Although the underlying concepts are closely related, a good deal of kinetic/thermodynamic work will be needed to sort these things out, since a PCS/anchor site is defined in kinetic terms as a recognition element that forms the initial contact with the partner, whereas a hot spot/MoRE is more of a thermodynamic term that signifies the region in the molecular interface that contributes the major part of the free energy of binding. It is to be noted that both may be interpreted in terms of the current "fly-casting" [76] model of IUP recognition, which suggests that IUPs make use of their folding funnel in binding to their partner. This mechanism invokes both the greater capture radius of IUPs and the mechanistic coupling of the recognition process to folding, in which pre-formed, exposed, recognition elements may be effective mechanistic devices.

5. Unique functional features endowed by disorder

The multifarious functioning of IUPs assumes that the lack of an ordered structure contributes in many ways to their mechanisms of action. In fact, their highly malleable structure endows them with functional features unparalleled by ordered proteins. The major benefits of structural disorder, as covered in several recent reviews [6–9,11,12], are the separation of specificity from binding strength, increased speed of interaction, the ability to bind distinct partners and effective regulation by degradation. Here, novel examples and extensions of these features are presented.

The advantage of the great conformational freedom of IUPs is most evident with entropic chains, which may exert a long-range, entropic exclusion of other proteins or cellular constituents in spacer functions (MAP2 [27]), and also in gating (nucleoporins [33]). Another molecular setting where such regions abound is in multidomain proteins, where globular domains are often separated by flexible linkers. These regions enable much freedom in orientational search [11] that permits the recognition of distant and/or discontinuous determinants on the target. Fully disordered IUPs also exploit this unique feature. Their extended structure enables them to contact their partner(s) over a large binding surface for a protein of the given size, which allows the same interaction potential to be realized by shorter proteins overall, encoded by a more economical genome [77]. In addition, the flexibility itself is instrumental to the assembly process itself, as certain complexes cannot be assembled from rigid components due to topological constraints (cf. Fig. 2).

A unique consequence of the structural flexibility of IUPs is their capacity to adapt to the structure of distinct partners, which enables an exceptional plasticity in cellular responses. An amply characterized case for this behavior is the Cdk inhibitor $p21^{Cip1}$, which can interact with CycA-Cdk2, CycE-Cdk2, CycD-Cdk4 complexes [34], the Rho kinase [78] and apoptosis signal-regulating kinase 1 [79] under different conditions; further examples can be found in [9]. The open, extended structure of IUPs also enables an increased speed of interaction. It has been noted that macromolecular association rates are highly enhanced by an initial, relatively non-specific, association enabled by flexible recognition segments, mechanistically

formulated in the "fly-casting" [76] or "protein fishing" [80] mechanisms of molecular recognition.

Another prominent feature of IUPs is that their extreme proteolytic sensitivity, in principle, allows for an effective control via rapid turnover. In fact, protein disorder prevails in signaling, regulatory and cancer-associated proteins, known to be short-lived proteins subject to rapid turnover [14,16]. Furthermore, disorder itself constitutes an integral part of the proteasomal destruction signal in two distinct ways. On the one hand, non-ubiquitinated IUPs may be directly degraded by the 20S proteasome, as shown for $p21^{Cip1}$ [47] and tau [46], for example. On the other hand, this mechanism may also play a more subtle regulatory role, by processing disordered segments in multidomain proteins and releasing the flanking, constitutively activated globular domains due to the endoproteolytic activity of the proteasome [81]. Disorder may also constitute part of the signal to the ubiquitination system itself [42] as the regions of securin and cyclin B recognized by the ubiquitination machinery have been shown recently to be natively unfolded. Furthermore, ubiquitination of unstructured regions may also directly stimulate the activity of proteins, as shown for certain transcription factors [82]. Intriguingly, ubiquitination in these cases not only signals destruction but it is also mandatory for activation. Thus, disorder may be involved in a very specific regulatory feature in which ubiquitination "licenses" activation to the destruction of the protein targeted.

6. Outlook

The history of intrinsically unstructured/disordered proteins is a short, yet already a very influential, one. These heretic proteins, which defy the once general structure–function paradigm that tied protein function to a well-defined 3D structure, prevail in all organisms studied thus far. They not only tolerate the lack of a stable structure but their structural disorder predisposes them to such elaborate functional modes that pale even the perfection of globular enzymes. As their unusual actions keep surprising us, their functional versatility has already transformed our basic concepts of protein structure and function.

Acknowledgment: This work was supported by the International Senior Research Fellowship GR067595 from the Wellcome Trust and a Bolyai János Fellowship.

References

[1] Lynch, W.P., Riseman, V.M. and Bretscher, A. (1987) Smooth muscle caldesmon is an extended flexible monomeric protein in solution that can readily undergo reversible intra- and intermolecular sulfhydryl cross-linking. A mechanism for caldesmon's F-actin bundling activity. J. Biol. Chem. 262, 7429–7437.
[2] Sigler, P.B. (1988) Transcriptional activation. Acid blobs and negative noodles. Nature 333, 210–212.
[3] Holt, C. and Sawyer, L. (1993) Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the alpha(s1)-, beta- and kappa-caseins. J. Chem. Soc. Faraday Trans. 89, 2683–2692.
[4] Wright, P.E. and Dyson, H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. J. Mol. Biol. 293, 321–331.
[5] Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) Why are natively unfolded proteins unstructured under physiologic conditions?. Proteins 41, 415–427.
[6] Dunker, A.K., et al. (2001) Intrinsically disordered protein. J. Mol. Graphics Modelling 19, 26–59.
[7] Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. and Obradovic, Z. (2002) Intrinsic disorder and protein function. Biochemistry 41, 6573–6582.
[8] Uversky, V.N. (2002) Natively unfolded proteins: A point where biology waits for physics. Protein Sci. 11, 739–756.
[9] Tompa, P. (2002) Intrinsically unstructured proteins. Trends Biochem. Sci. 27, 527–533.
[10] Bracken, C., Iakoucheva, L.M., Romero, P.R. and Dunker, A.K. (2004) Combining prediction, computation and experiment for the characterization of protein disorder. Curr. Opin. Struct. Biol. 14, 570–576.
[11] Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208.
[12] Fink, A.L. (2005) Natively unfolded proteins. Curr. Opin. Struct. Biol. 15, 35–41.
[13] Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C. and Brown, C.J. (2000) Intrinsic protein disorder in complete genomes. Genome Inform. Ser. Workshop Genome Inform. 11, 161–171.
[14] Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. 337, 635–645.
[15] Uversky, V.N. (2002) What does it mean to be natively unfolded?. Eur. J. Biochem. 269, 2–12.
[16] Iakoucheva, L., Brown, C., Lawson, J., Obradovic, Z. and Dunker, A. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J. Mol. Biol. 323, 573–584.
[17] Tompa, P. and Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones. FASEB J. 18, 1169–1175.
[18] Vucetic, S., et al. (2005) DisProt: a database of protein disorder. Bioinformatics 21, 137–140.
[19] Dyson, H.J. and Wright, P.E. (2004) Unfolded proteins and protein folding studied by NMR. Chem. Rev. 104, 3607–3622.
[20] Ellis, R.J. (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem. Sci. 26, 597–604.
[21] Daughdrill, G.W., Hanely, L.J. and Dahlquist, F.W. (1998) The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. Biochemistry 37, 1076–1082.
[22] Dedmon, M.M., Patel, C.N., Young, G.B. and Pielak, G.J. (2002) FlgM gains structure in living cells. Proc. Natl. Acad. Sci. USA 99, 12681–12684.
[23] Linding, R., Schymkowitz, J., Rousseau, F., Diella, F. and Serrano, L. (2004) A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J. Mol. Biol. 342, 345–353.
[24] Linding, R., Russell, R.B., Neduva, V. and Gibson, T.J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 31, 3701–3708.
[25] Dosztányi, Z., Csizmok, V., Tompa, P. and Simon, I. (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J. Mol. Biol. 347, 827–839.
[26] Flaugh, S.L. and Lumb, K.J. (2001) Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1). Biomacromolecules 2, 538–540.
[27] Morar, A.S., Olteanu, A., Young, G.B. and Pielak, G.J. (2001) Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome *c*. Protein Sci. 10, 2195–2199.
[28] Qu, Y. and Bolen, D.W. (2002) Efficacy of macromolecular crowding in forcing proteins to fold. Biophys. Chem. 101–102, 155–165.
[29] Fuxreiter, M., Simon, I., Friedrich, P. and Tompa, P. (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol. 338, 1015–1026.
[30] Sorenson, M.K., Ray, S.S. and Darst, S.A. (2004) Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/

FlgM reveals an intact sigma factor in an inactive conformation. Mol. Cell 14, 127–138.
[31] Trombitas, K., Greaser, M., Labeit, S., Jin, J.P., Kellermayer, M., Helmes, M. and Granzier, H. (1998) Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments. J. Cell Biol. 140, 853–859.
[32] Mukhopadhyay, R. and Hoh, J.H. (2001) AFM force measurements on microtubule-associated proteins: the projection domain exerts a long-range repulsive force. FEBS Lett. 505, 374–378.
[33] Denning, D.P., Patel, S.S., Uversky, V., Fink, A.L. and Rexach, M. (2003) Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. Proc. Natl. Acad. Sci. USA 100, 2450–2455.
[34] Kriwacki, R.W., Hengst, L., Tennant, L., Reed, S.I. and Wright, P.E. (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc. Natl. Acad. Sci. USA 93, 11504–11509.
[35] Fabrega, C., Shen, V., Shuman, S. and Lima, C.D. (2003) Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell 11, 1549–1561.
[36] Dames, S.A., Martinez-Yamout, M., Guzman, R.N., Dyson, H.J. and Wright, P.E. (2002) De Structural basis for Hif-1 alpha/CBP recognition in the cellular hypoxic response. Proc. Natl. Acad. Sci. USA 99, 5271–5276.
[37] Elkins, J.M., Hewitson, K.S., McNeill, L.A., Seibel, J.F., Schlemminger, I., Pugh, C.W., Ratcliffe, P.J. and Schofield, C.J. (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. J. Biol. Chem. 278, 1802–1806.
[38] Hurst, L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 18, 486.
[39] Tucker, P.K. and Lundrigan, B.L. (1993) Rapid evolution of the sex-determining locus in Old World mice and rats. Nature 364, 715–717.
[40] Whitfield, L.S., Lovell-Badge, R. and Goodfellow, P.N. (1993) Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature 364, 713–715.
[41] Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z. and Dunker, A.K. (2004) The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32, 1037–1049.
[42] Cox, C.J., Dutta, K., Petri, E.T., Hwang, W.C., Lin, Y., Pascal, S.M. and Basavappa, R. (2002) The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. FEBS Lett. 527, 303–308.
[43] Prakash, S., Tian, L., Ratliff, K.S., Lehotzky, R.E. and Matouschek, A. (2004) An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat. Struct. Mol. Biol. 11, 830–837.
[44] Dafforn, T.R. and Smith, C.J. (2004) Natively unfolded domains in endocytosis: hooks, lines and linkers. EMBO Rep. 5, 1046–1052.
[45] Davies, K.J. (2001) Degradation of oxidized proteins by the 20S proteasome. Biochimie 83, 301–310.
[46] David, D.C., Layfield, R., Serpell, L., Narain, Y., Goedert, M. and Spillantini, M.G. (2002) Proteasomal degradation of tau protein. J. Neurochem. 83, 176–185.
[47] Sheaff, R.J., Singer, J.D., Swanger, J., Smitherman, M., Roberts, J.M. and Clurman, B.E. (2000) Proteasomal turnover of p21Cip1 does not require p21Cip1 ubiquitination. Mol. Cell 5, 403–410.
[48] Olashaw, N., Bagui, T.K. and Pledger, W.J. (2004) Cell cycle control: a complex issue. Cell Cycle 3, 263–264.
[49] Haarmann, C.S., Green, D., Casarotto, M.G., Laver, D.R. and Dulhunty, A.F. (2003) The random-coil 'C' fragment of the dihydropyridine receptor II–III loop can activate or inhibit native skeletal ryanodine receptors. Biochem. J. 372, 305–316.
[50] Tompa, P., Mucsi, Z., Orosz, G. and Friedrich, P. (2002) Calpastatin subdomains A and C are activators of calpain. J. Biol. Chem. 277, 9022–9026.
[51] Uhlmann, F. (2003) Separase regulation during mitosis. Biochem. Soc. Symp. 70, 243–251.
[52] Gast, K., et al. (1995) Prothymosin alpha: a biologically active protein with random coil conformation. Biochemistry 34, 13211–13218.
[53] Weinreb, P.H., Zhen, W., Poon, A.W., Conway, K.A. and Lansbury Jr., P.T. (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry 35, 13709–13715.
[54] Kim, T.D., Ryu, H.J., Cho, H.I., Yang, C.H. and Kim, J. (2000) Thermal behavior of proteins: heat-resistant proteins and their heat-induced secondary structural changes. Biochemistry 39, 14839–14846.
[55] Wallon, G., Rappsilber, J., Mann, M. and Serrano, L. (2000) Model for stathmin/OP18 binding to tubulin. EMBO J. 19, 213–222.
[56] Hackel, M., Konno, T. and Hinz, H. (2000) A new alternative method to quantify residual structure in 'unfolded' proteins. Biochim. Biophys. Acta 1479, 155–165.
[57] Richards, J.P., Bachinger, H.P., Goodman, R.H. and Brennan, R.G. (1996) Analysis of the structural properties of cAMP-responsive element-binding protein (CREB) and phosphorylated CREB. J. Biol. Chem. 271, 13716–13723.
[58] Csizmok, V., Bokor, M., Banki, P., Klement, É., Medzihradszky, K.F., Friedrich, P., Tompa, K. and Tompa, P. (2005) Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. Biochemistry 44, 3955–3964.
[59] Williamson, M.P. (1994) The structure and function of proline-rich regions in proteins. Biochem. J. 297, 249–260.
[60] Uversky, V.N., Winter, S., Galzitskaya, O.V., Kittler, L. and Lober, G. (1998) Hyperphosphorylation induces structural modification of tau-protein. FEBS Lett. 439, 21–25.
[61] Andrews, A.L., Atkinson, D., Evans, M.T.A., Finer, E.G., Green, J.P., Phillips, M.C. and Robertson, R.N. (1979) The conformation and aggregation of bovine beta-casein A. I. Molecular aspects of thermal aggregation. Biopolymers 18, 1105–1121.
[62] Bienkiewicz, E.A., Woody, A. and Woody, R.W. (2000) Moon conformation of the RNA polymerase II C-terminal domain: circular dichroism of long and short fragments. J. Mol. Biol. 297, 119–133.
[63] Syme, C.D., Blanch, E.W., Holt, C., Jakes, R., Goedert, M., Hecht, L. and Barron, L.D. (2002) A Raman optical activity study of rheomorphism in caseins, synucleins and tau. Eur. J. Biochem. 269, 148–156.
[64] Blanch, E.W., Kasarda, D.D., Hecht, L., Nielsen, K. and Barron, L.D. (2003) New insight into the solution structures of wheat gluten proteins from Raman optical activity. Biochemistry 42, 5665–5673.
[65] Hua, Q.X., Jia, W.H., Bullock, B.P., Habener, J.F. and Weiss, M.A. (1998) Transcriptional activator–coactivator recognition: nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding. Biochemistry 37, 5858–5866.
[66] Bienkiewicz, E.A., Adkins, J.N. and Lumb, K.J. (2002) Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). Biochemistry 41, 752–759.
[67] Dunker, A.K. and Obradovic, Z. (2001) The protein trinity – linking function and disorder. Nat. Biotechnol. 19, 805–806.
[68] Fontana, A., Polverino de Laureto, P., Filippis, V., Scaramella, E. and Zambonin, M. (1997) De Probing the partly folded states of proteins by limited proteolysis. Fold. Des. 2, R17–R26.
[69] Marston, S.B. and Redwood, C.S. (1991) The molecular anatomy of caldesmon. Biochem. J. 279, 1–16.
[70] Redeker, V., Lachkar, S., Siavoshian, S., Charbaut, E., Rossier, J., Sobel, A. and Curmi, P.A. (2000) Probing the native structure of stathmin and its interaction domains with tubulin. Combined use of limited proteolysis, size exclusion chromatography, and mass spectrometry. J. Biol. Chem. 275, 6841–6849.
[71] Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabartty, A. and Arrowsmith, C.H. (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein–protein and protein–DNA interactions?. J. Mol. Biol. 345, 275–287.
[72] Hollenbeck, J.J., McClain, D.L. and Oakley, M.G. (2002) The role of helix stabilizing residues in GCN4 basic region folding and DNA binding. Protein Sci. 11, 2740–2747.

[73] Mucsi, Z., Hudecz, F., Hollosi, M., Tompa, P. and Friedrich, P. (2003) Binding-induced folding transitions in calpastatin subdomains A and C. Protein Sci. 12, 2327–2336.
[74] Rajamani, D., Thiel, S., Vajda, S. and Camacho, C.J. (2004) Anchor residues in protein–protein interactions. Proc. Natl. Acad. Sci. USA 101, 11287–11292.
[75] Bogan, A.A. and Thorn, K.S. (1998) Anatomy of hot spots in protein interfaces. J. Mol. Biol. 280, 1–9.
[76] Shoemaker, B.A., Portman, J.J. and Wolynes, P.G. (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc. Natl. Acad. Sci. USA 97, 8868–8873.
[77] Gunasekaran, K., Tsai, C.J., Kumar, S., Zanuy, D. and Nussinov, R. (2003) Extended disordered proteins: targeting function with less scaffold. Trends Biochem. Sci. 28, 81–85.
[78] Tanaka, H., Yamashita, T., Asada, M., Mizutani, S., Yoshikawa, H. and Tohyama, M. (2002) Cytoplasmic p21Cip1/WAF1 regulates neurite remodeling by inhibiting Rho-kinase activity. J. Cell Biol. 158, 321–329.
[79] Asada, M., Yamada, T., Ichijo, H., Delia, D., Miyazono, K., Fukumuro, K. and Mizutani, S. (1999) Apoptosis inhibitory activity of cytoplasmic p21(Cip1/WAF1) in monocytic differentiation. EMBO J. 18, 1223–1234.
[80] Evans, P.R. and Owen, D.J. (2002) Endocytosis and vesicle trafficking. Curr. Opin. Struct. Biol. 12, 814–821.
[81] Liu, C.W., Corboy, M.J., DeMartino, G.N. and Thomas, P.J. (2003) Endoproteolytic activity of the proteasome. Science 299, 408–411.
[82] Salghetti, S.E., Caudy, A.A., Chenoweth, J.G. and Tansey, W.P. (2001) Regulation of transcriptional activation domain function by ubiquitin. Science 293, 1651–1653.