Multisteric regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery

Peter Tompa[1,2]

[1]VIB Department of Structural Biology, Vrije Universiteit Brussel, Brussels, Belgium, and [2]Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary,

Table of Contents

1. Introduction

Allostery is a classical regulatory mechanism of proteins in which a signal at "another site" ("allos stereos" – another object in Greek) modifies the activity/function of a protein. The phenomenon has its roots in cooperativity, when binding of a ligand to an oligomeric protein increases the affinity to another one, i.e. binding events "cooperate".[1] Later it was recognized that an unrelated regulatory ligand can also modulate enzyme activity,[2] which gave rise to the concept of allostery as mostly known today. The underlying molecular mechanisms have been traditionally interpreted in terms of concerted changes in the structure of oligomeric proteins,[3,4] also recognizing that conformational changes can channel such regulatory information between distinct sites even in a monomeric protein.[5-8] Whereas it had always been appreciated that allostery stems from the conformational plasticity of proteins, due to the success and explanatory power of structural biology, our view of allostery is dominated by static views giving dynamics only secondary emphasis. However, since energetic coupling between remote sites can be achieved without structural changes, allosteric communication might also arise from altering protein dynamics.[6,7,9] In fact, with recognizing the generality of structural disorder of proteins and the landscape theory of protein structure, a "new view" of allostery started to emerge, in which emphasis is placed on ligand-induced shifts in the conformational ensemble of the protein. The ensuing changes in ligand binding/catalytic activity may stem from coupled folding transitions of distinct binding sites,[10-12] or remodeling of the conformational landscape to entropically favor a particular downstream binding/catalytic event.[13,14] Whereas the ensemble allostery model (EAM) provides a phenomenological framework of allostery arising from disorder-to-order transitions,[10,12] here we show that this model can mostly rationalize the behavior of simple regulatory proteins, whereas allostery and related terms (e.g. interdomain allostery, intramolecular-, intrasteric- and active-site directed regulation, motif exposure, autoinhibition and signaling conduit function) are increasingly used in the case of multidomain regulatory proteins with extended structural disorder. Through the example of p53, Wiskott-Aldrich syndrome protein (WASP), p27[Kip1], calcineurin, Sap97, Hemopoietic cell kinase (Hck), Vav-family guanine nucleotide exchange factors (Vav), SMAD specific E3 ubiquitin protein ligase 2 (Smurf2), stoichiometric inhibitor of Cdk1 (Sic1), E26 transformation-specific-1 (Ets-1), androgen receptor (AR) and others, we show how remodeling of conformational ensembles mediates long-range coupling between distinct functional sites, to entropically favor a particular downstream signaling (binding or catalytic) event. Due to a combination of multiple regulatory sites (motifs and domains) and complex conformational ensembles enabled by structural disorder, unlike classical allosteric switches, these regulatory proteins can integrate and interpret multiple incoming signals. As suggested for CBP/p300, they may act as "versatile molecular interpreters that can parse and/or conjugate the regulatory words, phrases, and sentences of the genome".[15] We anticipate that research into the structural and thermodynamic underpinning of this extended concept of allostery (termed here multistery) will result in a unifying mathematical model that will incorporate EAM as a special case. These future studies will represent one of the most interesting areas of the structural biology of intrinsically disordered proteins/regions (IDPs/IDRs), and they will eventually enable the rationalization of the role of complex, modular proteins in orchestrating complex cell signaling events.

2. Classical mechanisms of allostery

2.1. Models of cooperativity

The concept of allostery is rooted in the phenomenon of cooperativity, initially recognized as the interdependence of binding of identical ligands (termed homotropic) to oligomeric proteins, such as four oxygen molecules to a tetrameric hemoglobin.[1] The ensuing sigmoidal binding isotherm cannot be described by a simple saturation, rather it shows signs of "cooperation" between ligands. If binding of one ligand weakens that of the others, we may also speak about negative cooperativity. To elucidate the underlying mechanistic changes, two models have been suggested, which even today form the basis of our textbook wisdom of this phenomenon. Both models assume changes in the quaternary structure of oligomeric proteins that result in two conformational states, T (tense) and R (relaxed), with the substrate binding preferentially to the relaxed state. The models concur on the sequence of events (Figure 1A) and the stringency of conformational coupling between subunits. In the Monod-Wyman-Changeux (MWC), or concerted, model, the subunits are strongly coupled and the entire oligomer can only occur in either of all-T or all-R states. Binding of one ligand shifts the equilibrium to the all-R state, thus promoting the binding of subsequent ligand(s).[4,16] Also central to this model is conformational selection, i.e. that both states exist prior to binding of the ligand, which then stabilizes (selects) the all-R state. In the Koshland-Nemethy-Filmer (KNF), or sequential, model, the structural state of the oligomer is dominated by the all-T state, in which binding of the first ligand induces a transition to an R state only in that subunit.[3] The coupling of the subunits is not strong enough to provoke a conformational change in the other subunits, but it reduces the free energy of their subsequent transition to the R state.

2.2. Structural and dynamic basis of allostery

Initially, these phenomenological models emphasized kinetics/thermodynamics of coupled binding (Figure 1A) and, only later, through detailed structural studies, were the underlying conformational changes described at atomic detail.[17-20] The mathematical formalism was easily extended to accommodate heterotropic (occurring between different ligands) and/or negative cooperativity.[5,16] From a more theoretical point of view, it became clear that the essence of all related phenomena is the communication (energetic coupling) between remote binding sites enabled by conformational changes within a protein or a complex.

Energetic coupling requires signal propagation through the structure of the protein, which might be thought as a series of coupled local conformational changes of adjacent residues connecting remote sites (Figure 1B). Thus, the initial description of cooperativity has given way to the more general concept of allostery (from Greek allos - another and stereos - object). By observing that different ligands can be coupled, the conformational coupling is mediated through the matrix of a protein structure,[21,22] and many monomeric proteins show this behavior,[9,23] the concept of pathways or sectors, i.e. the set of interacting residues mediating the effect, emerged.[23] This view suggests that

the entire structure may collectively undergo a transition, i.e. the pathway is encoded by the entire structure.[24]

It also became clear that allosteric coupling might occur without a discernible conformational change, by way of altering the dynamics of the protein only[6,21,25] (Figure 1C) to favor a particular downstream binding/catalytic event. By combining elements of classical cooperativity (subunit interactions) and conformational coupling augmented by structural disorder (novel allostery), recently it was also reported that domain-domain interactions enabled by a disordered linker[13,26] might permit the formation of an energetic pathway traversing separate domains through their interface (Figure1D).

All the classical cases have been studied in great detail in terms of binding kinetics, thermodynamics and structural mechanisms (Table 1). The functional diversity, regulatory detail, and structural underpinning of all these cases underscore the general sentiment of the field that due to conformational coupling, practically any protein can be allosteric.[9] Having said this, however, one has to be aware that all the foregoing "classical" cooperative/allosteric phenomena (cf. Table 1) are interpreted within the confines of the classical structure-function paradigm, which assumes that a well-defined structure is the prerequisite for function. Whereas this view has been instrumental in understanding a wealth of cellular/molecular phenomena, recent observations have questioned its universality.

2.3. Biology of allostery

Allostery is a fundamental regulatory device of the cell, used to tune its activity in response to external and internal cues. The handful of diverse classic examples (Table 1) illustrates the breadth and generality of this phenomenon. The cases show a wide variety of molecular functions, ranging from energy metabolism, signal transduction, stress response, transcription regulation, and antibiotic resistance. The proteins also greatly vary in molecular design, ranging from monomers[8] to oligomers[19] of multiple subunits,[27] even demonstrating subunit dissociation as a mechanistic feature,[28] representing homotropic and heterotropic ligand interactions, positive and negative coupling mechanisms. Given the great variety in all aspects, it is not unreasonable to conclude that allostery represents one of the most important, widespread and versatile regulatory mechanisms of the cell.

Table 1 *Classical allostery cases (very limited, illustrative)*

Protein	term used	structure/ligand	allosteric coupling	function	ref
Hemoglobin	cooperativity	Dimer of heterodimers/oxygen	T-R transition	oxygen transport	[19]
Lactate dehydrogenase	allostery	homotetramer/fructose 1,6 bisphosphate	T-R transition	energy metabolism	[18]
Phosphofructokinase	allostery	hexamer/ATP	T-R transition	glycolysis	[29]
PSD95 PDZ domain	coupling of dynamics	monomeric/peptide ligand	dynamic allostery	signal transduction	[8]
Trp-tRNA synthetase	allostery	multidomain protein/tRNA anticodon	conformational transition	Trp-tRNA synthesis	[30]
Glycogen phosphorylase	allostery	homodimer/phosphorylation	T-R transition	energy metabolism	[20]
Catabolite activator protein (CAP)	allsotery	homodimer/cAMP	change in dynamics	DNA binding, transcription	[25]
Epidermal growth factor receptor (EGFR)	allostery	homodimer/EGF	conformational coupling	signal transduction	[31]
Flegaller switch	cooperativity	circular oligomer/response regulator CheY-Protein	stochastic conformational coupling	control of flagellar motor	[27]
aminoglycoside-N-acytyltransferase	allostery (KNF) plus binding-induced unfolding	homodimer/AcCoA	conformational coupling/partial unfolding	antibiotic resistance	[32]
Tetracyclin repressor (TetR)	folding cooperativity, allostery	homodimer/tetracycline	induced folding	DNA binding/ transcription	[33]
Porphobilinogen synthase	morpheein allostery	octamer-hexamer-dimer/	change in oligomeric state	tetrapyrorle	[34]

The essential readout of allostery is non-linearity of activity, in which the biological response to a ligand/substrate shows a switch-like behavior as a function of concentration of the given or other (effector/regulatory) ligand. Because of the stringency of structural connectivity through the structure of a folded protein, these classic allosteric mechanisms have their limitations: they usually respond to only one regulatory ligand (or substrate), and the direction of the switch is unimodal (either positive or negative, on or off). These classic switches (Table 1) are not capable of integrating multiple signals, nor can they make Boolean-type logic decisions (and/or). As we will show next, such regulatory phenomena emerged with the advent of complex multidomian proteins harboring disordered linkers, having complicated energy landscapes, which enables complex patterns of domain-domain and domain-motif interactions via subtle alterations in their population of conformations. Because the crux of this regulatory phenomenon is the presence of structural disorder, we will next discuss its essential features in some detail.

3. Structural disorder and its role in regulatory functions

3.1. Structural disorder

As detailed in several other pieces of this volume, a substantial body of structural evidence suggests that many proteins termed IDPs/IDRs[35-38] are devoid of stable tertiary structure under native, functional conditions. Such proteins are deposited into the DisProt database,[39] which contains more than 1500 cases of curated IDPs/IDRs. These observations brought about a radical change in the structure-function paradigm of proteins and launched a new field addressing how these proteins evolve, exist and function, as phrased in a call for establishing "unstructural" biology.[36]

Indirect evidence suggests that the full complement of disordered proteins (the disorderome or unfoldme) is order(s) of magnitude larger than this collection based on limited experimental evidence. Based on sequences found in DisProt, sophisticated bioinformatic prediction algorithms have been developed,[40,41] which suggest that structural disorder is abundant in all organisms, in particular in eukaryotes that rely on a complex integration of their cellular functions.[42] The consensus in the field is that about 10-35% of residues in prokaryotic proteins and 15-45% of residues in eukaryotic proteomes fall into locally disordered regions, about 50% of eukaryotic proteins contain long continuous stretches of disordered residues (≥30 residues), and up to 10-15% of eukaryotic proteins are fully disordered.[43-45]

Following the rapid conceptual transition provoked by expanding evidence protein disorder, one major recent challenge in the field is to characterize the structure (structural ensemble) of IDPs/IDRs by a combination of highly advanced structural techniques,[46,47] and interpret their function in terms of their heterogeneous and highly dynamic structural state. Current efforts also aim to provide insight into their structure either *in vivo*[48,49] or at the single-molecule level.[50,51] As outlined next, their basic functional modes are entirely relevant with the novel type of allostery we aim to describe.

3.2. Basic functions and functional modes of disordered proteins

As suggested, the recent drive in the field is to understand and describe in detail the functions and functional modes of IDPs/IDRs, both from the "biological process" (BP) and "molecular function" (MF) perspective, as defined in the Gene Ontology classification.[52]

The BP aspect of structural disorder was addressed in several large-scale bioinformatics studies. Early on, IDPs were predicted to be involved in signaling and regulatory processes of the cell.[43,44,53] A whole-scale correlation analysis of Swiss-Prot functional keywords suggested that 238 functions are positively correlated and 302 functions negatively correlated with structural disorder in 710 SwissProt functional keywords.[45] Functions associated with transcription and translation is enriched in long disordered regions, whereas functional categories associated with enzymatic catalysis are depleted

in disorder. To name a few categories, structural disorder is unusually abundant in transcription factors,[54,55] scaffold proteins,[56] chromatin organizing proteins,[57,58] chaperones,[59] intracellular regions of transmembrane receptors[54,60] and signaling proteins regulating the cell cycle.[53,61,62]

In a structural sense, IDPs/IDRs are best approached as a structural ensemble, in which a very large number of conformations exist and interconvert on the timescale of ps-ns locally and μs-ms globally.[38] Structural disorder should not be thought of as a homogeneous and easily definable structural state, rather it should be considered as a broad continuum from very disordered (random coil-like) to compact and almost folded (molten globule-like) states. In general, at none of these levels should they be considered as random.

Most relevant with regards to the allosteric behavior of complex multidomian proteins are the MF functional modes of IDPs/IDRs. Broadly speaking, disordered proteins can function either directly via their disorder (entropic chain functions) or via molecular recognition, when they bind to a partner and undergo induced folding (disorder-to-order) transition to a more structured state.[63-66] These binding modes entail multiple functional advantages, such as adaptability in binding, high functional density, separation of binding strength from specificity and ease of regulation by post-translational modifications. Linker regions connecting domains often function as entropic chains, which impart practically unrestricted conformational freedom on their flanking domains.[67-69]

Molecular recognition functions of IDPs/IDRs are most often fulfilled via short peptide-like recognition motifs termed short linear motifs (SLiMs) or eukaryotic linear motifs (ELMs),[70] preformed structural elements (PSEs)[71] or molecular recognition features (MoRFs).[72] Whereas these different manifestations of short recognition elements are not necessarily fully interchangeable, all of them are thought to correlate with local structural disorder,[73,74] and to mediate interactions with cognate recognition domains of the partner molecule (e.g. the binding of PxxP motifs to SH3 domains). Interactions of motifs are usually weak, transient, and context dependent.[70,75] There are also longer binding regions of IDPs/IDRs, which satisfy the definition of domains in a structural, functional and evolutionary sense[76] (e.g. the WASP-homology 2 (WH2) domain, a disordered G actin-binding sequence in several proteins that regulate actin cytoskeleton, such as WASP, Cordon-bleu and Spire).[77]

A further relevant feature of IDPs is molecular recognition without concomitant induced folding to a (fully) folded state. In the case of such "fuzzy" interactions,[78] the function of the protein, even in the bound state, stems from an ensemble of conformations, which demands the extension of the concept of structural disorder to the bound state.

3.3. Structural characterization of disorder

The structure of IDPs/IDRs can be best described by combining a range of techniques that characterize various and complementary features of their structural ensembles.[47,79] Many of these techniques, such as small-angle X-ray scattering (SAXS), CD, fluorescence energy transfer (FRET), electron spin resonance (ESR) and others, provide insight into

the structural and topological organization of their polypeptide chain. The uncontested champion of the structural characterization of IDPs is NMR, which, when usually combined with SAXS data and molecular dynamics (MD)-based data integration, can provide residue-level description of their structural ensemble,[46,80-82] as already demonstrated for several dozen IDPs (e.g. tau protein,[83] and α synuclein[84,85]). The usual approach is to achieve full assignment of resonances which enables to determine a range of parameters sensitive to local structure, such as chemical shift of various nuclei, hetNOE, relaxation parameters, and residual dipolar coupling (RDC) values for all residues. To interpret these observed parameters in terms of a structural ensemble, a very large number of disordered conformers are calculated (e.g. with Flexible Meccano[86]), and then a limited number of them are selected (e.g. with ASTEROIDS[87]) to fit observations. Often, additional structural data obtained by low-resolution techniques, such as SAXS, FRET or NMR paramagnetic resonance enhancement (PRE) are added to the structural constraints, to reach a description that reasonably approaches both short-range and long-range features of the structural ensemble.

These approaches can also be combined by structural elements (e.g. domains) solved and modeled by X-ray, NMR and electron microscopy (EM) experiments, enabling the characterization of complex systems (e.g. a multidomain protein such as p53,[88] an IDP bound to a multi-subunit complex, such as Sic-1 in complex with SCF[Cdc4] ref.[89] or a viral phosphoprotein bound to the viral capsid[90]). Of particular note, these techniques not only unveil the static structure, but also the dynamics of the protein: apparently, detailed interpretation of all these parameters will ultimately lead to the correct and predictive structure-function models of complex multidomain regulatory proteins, as anticipated in the novel concept of their higher-level "supertertiary" structural organization.[91]

4. Novel type of allostery in multidomain proteins

Structural-functional studies of proteins involved in signal transduction suggest that regulatory proteins are often more complex than those assumed in traditional allostery models, containing several globular domains,[92] short recognition elements,[73,93] and disordered linker regions[66,94,95] in various combinations. The combinatorial arrangements of such elements enable the construction of modular proteins of almost unlimited variety[96,97] in which dynamic interactions of binding elements enabled by structural disorder give rise to extremely complex conformational energy landscapes. Whereas these proteins are sometimes mentioned in conjunction with allosteric regulation, they no longer fit into the models of allostery based on the classical view of a unique 3-D structure of proteins.

In realization that their mechanisms of regulation cannot be interpreted in terms of classical allostery, their complex signal-integrating behavior enabled by disorder is sometimes called allostery[98-101] or interdomain allostery (Hsp-70,[13] NHERF1,[102] Dlg[103]), but often something else, e.g. intrasteric regulation,[104] autoinhibition,[105,106] intramolecular regulation,[107] active-site directed regulation,[108] signaling conduit function,[109] motif exposure,[110] or neither of these terms[111-113] (cf. Table 2). It will be shown that the underlying molecular mechanisms invariably rely on signal propagation in the ensemble of structures, as already suggested by a "new view",[22,100] "dynamic

view",[14] or "ensemble view"[10-12] of allostery. The essence of these mechanisms is not necessarily the energetic coupling between domains optimized by ligand-induced folding,[10-12] rather a population shift (remodeling) in the conformational

Table 2 Proteins of novel type of allostery

Protein, functional class (UniProt ID)	term in literature	input signal[a]	output	distance[b]	disorder (exp.)[c]	disorder (%VSL2)[d]	activation[e]	ref.
Bacterial antitoxin PhD (Q06259)	conditional cooperativity	binding of toxin Doc	DNA binding/ transcription	adjacent (two domains)	yes	70.0	D → O	114
DnaK (Hsp70, P0A6Y8)	(interdomain) allostery	ATP binding	client binding	7 (ATP-binding domain, client-binding domain)	yes	50	D → O	13
Hck, Tyr-kinase (P08631)	autoinhibited	dePh	kinase activity	SH2-kinaseD: 22	yes	95.5	O → D	26
Ets-1, transcription factor (P14921)	allosteric	multiple Phosphorylation	DNA-binding	Ser^{251}-helix1: 50	yes	100.0	D → D	99
p53, transcription factor (P04637)	allosteric	Ph, Ac	DNA-binding	DBD-regulatoryD: 77	yes	100.0	D → D	115,116
Vav, cytoskeleton (P15498)	autoinhibited	Phosphorylation	Rho binding	helix-DH-D: 10	yes	100.0	O → D	106
WASP, cytoskeleton (P42768)	allosteric	Cdc42, PIP2, Ph	Arp2/3 binding	CRIB-WH2: 179	yes	78.3	O → D	98
p27^{Kip1}, cell cycle (Q9NYG6)	signaling conduit	Phosphorylation	Ph, degr.	Tyr^{88}-Thr^{187}: 99	yes	100.0	D → D	109,117
Smurf2, E3 ligase (Q9HAU4)	autoinhibited	Smad7 binding	E2 binding	C2-HECT: 316	yes	53.2	O → D	105
importin-α, transport receptor (P52293)	intrasteric	importin β binding	NLS binding	IARS-receptorD: 18	no	100.0	O → D	118
CFTR, transmembrane channel (P13569)	none	multiple Phosphorylation	nucleotide binding	NBD1-Ser^{813}: 161	yes	95.0	D → D	111
Sic1, cell cycle (P38634)	none	multiple Phosphorylation	ubiquitination	Thr^2-Ser^{80}: 78	yes	100.0	D → D	112
Phd, bacteriophage antitoxin (Q06253)	allosteric	toxin binding	DNA binding	DBD-toxin-BD: 17	yes	47.1	D → O	114
Calcineurin, protein phosphatase 2B (CaN, Q08209)	autoinhibited	CaM binding	enzymatic activity	CaMBD-inhibitoryD: 52	yes	100.0	D → O, O → D	119
Murine double minute 2 (MDM2, Q00987)	intrasteric	phosphorylation, L5, L11, L23	ubiquitin transfer	SWIB-RING: 329	yes	93.9	D → D	120,121

			binding					
MAGUK Sap97 (Dlg, Q12959)	intramolecular regulation (interdomian allostery)	partner binding	GKAP binding	SG97N–SH3/GUK: 358	yes	45.8	D → D	[107]
p53 (P04637)	motif exposure	DNA binding	Lys^{373} (and other sites) acetylation	DNA-binding D-Lys^{373}: 81	yes	100	O → D	[110]
Ultrabitorax (P83949)	none	DNA and partner binding	transcription activity	motif 1 – I2 region: 225	yes	69.3	D → D	[113]

[a]nature of the input signal (Ph: phosphorylation, dePh: dephosphorylation, Ac: acetylation, CaM: calmodulin).
[b]the number of residues in the primary sequence (preferably of the human protein) between the effector and output sites. Because these proteins can integrate multiple signals, this is often not straightforward to define, and the value shown is rather arbitrary: usually the regulatory site farthest from the effector site is shown. "D" denotes "domain".
[c]if experimental evidence is available for the disorder of the protein (linker region).
[d]predicted disorder of the linker region between the two sites (% of residues disordered by the PONDR VSL2 algorithm [122]).
[e]it is indicated if the primary transition in the conformational ensemble upon allosteric activation is disorder-to-order (D → O), order-to-disorder (O → D), disorder-to-disorder (D → D) or a combination of both (D → O, O → D). It is to be noted that structural data at the level of the whole protein is often lacking, thus these statements are sometimes somewhat arbitrary.

ensemble (even without an appreciable change in the level of disorder) to entropically favor a particular downstream readout event, binding of a ligand or chemical conversion of a substrate.

Despite a great variety in molecular details, the unifying theme of the mechanisms that will be discussed in detail (Table 2) is the transduction of information from "another" site (or sites) to the ligand-binding or catalytic site, i.e. by this criterion they can - and ought to be – called allosteric. Binding of the effector (a small ligand, a macromolecular partner or post-translational modification) remodels the energy landscape and shifts the distribution of conformations in the ensemble. We show that this phenomenon is widespread (Table 2) and it does not even necessarily invoke folding transitions, as previously thought.[10-12] The complexity of the underlying energy landscapes enables that the activated state be either more ordered,[114] or more disordered[98,105,106] than the ground sate, and in certain cases, such as p27[Kip1] ref.[109,117] and Ets-1,[99] no folding transition occurs at all. This novel, generalized type of allostery underlies the operation of some of the most intriguing and important modular regulatory proteins (Table 2, Figure 2, Figure 3), which are sometimes termed "allosteric", but often something completely different: it will be shown that they all fit into a general, extended concept of allostery that deserves the distinguishing term "multistery".

4.1. Multidomain allosteric regulators

Certain multidomain proteins are regulated by several signals and are termed "allosteric", although the underlying molecular events cannot be accommodated in the classical mechanisms relying on either static or dynamic structural models. Two distinct modes of the optimization of allosteric activation by structural disorder appear. In a limiting case, allosteric coupling stems from domain-domain interactions optimized by ligand-induced folding (disorder-to-order transition). This "enthalpic" principle (Figure 4A) has been modeled[10-12] and appears to occur in the bacterial toxin/antitoxin system PhD/Doc[114] and nuclear hormone receptors,[123-125] for example (cf. table 2). In the PhD/Doc system, allostery results in a complex regulatory phenomenon, known as conditional cooperativity, due to which toxin Doc acts as either a corepressor or derepressor of the transcription factor PhD antitoxin, depending on their molar ratio.[114] PhD has an intrinsically disordered C-terminal domain, which becomes structured upon binding to Doc. Due to its structural coupling with the N-terminal DNA-binding domain, this initial induced folding process is allosterically coupled to DNA binding by PhD. At a different stoichiometry of the two proteins, the readout of binding might differ, coupling Doc neutralization to the conditional regulation of transcription.

A more general scenario applies in other "allosteric" cases, when a signal remodels the structural ensemble, and, perhaps even without a folding transition, changes the accessibility of certain output regions (Table 2). Two extreme cases exemplify the diversity of possible mechanisms: WASP undergoes order-to-disorder transition upon activation (Figure 2A), whereas Ets-1 is disordered in both its inhibited and activated states (Figure 2B).

The effect of Cdc42, a Rho family small GTPase, on cytoskeletal dynamics is mediated by allosteric[98,126] WASP, a modular protein of a WH1 domain, a basic region (BR), a GTPase-binding domain (GBD or CRIB), a Pro-rich linker region and a C-terminal VCA region.[98,127,128] In the autoinhibited closed state, GBD binds to the remote VCA region and prevents its interaction with other proteins (Figure 2A). Activation results from the interplay of several signals, such as binding of Cdc42 to GBD and PIP_2 to BR, which opens the structure and results in a largely disordered state in which VCA becomes exposed. In this activated state, VCA can bind Arp2/3, and thereby promote actin polymerization. Further stimulatory signals cooperate with Cdc42, such as Tyr phosphorylation within GBD, and binding of SH3 domain-containing effectors to the Pro-rich linker region. Regulatory signals propagate some 200 amino acids (Table 2) through fully disordered regions, i.e. apparently without the involvement of any identifiable pathway or interconnected amino acid network.

The transcription factor Ets-1 is involved in T-cell signaling and development.[99] The protein contains a largely helical DNA-binding module, in which an autoinhibitory region binds to the ETS domain (actually, a binding helix), thus reducing its affinity for DNA (Figure 2B). DNA-binding is allosterically regulated by a disordered serine-rich region (SRR), which is in loose association with the autoregulatory region. When SSR is phosphorylated at multiple sites by calcium-dependent kinases, it dampens the flexibility of the DNA-binding module and inhibits DNA binding. Gradual dephosphorylation releases SRR, increases flexibility of the ETS helix and promotes its DNA binding. Intriguingly, NMR experiments show that SRR remains disordered in both

the bound and unbound states,[99] i.e. allosteric information spreads through the protein via a distributed array of transient interactions between the regulatory and effectors sites. Variable phosphorylation serves as a "rheostat"[99] for cell signaling to fine-tune transcription at the level of DNA binding.

4.2. Interdomain allostery

Classical cooperativity of oligomeric proteins (Table 1) assumes structural communication between adjacent subunits. A similar mechanism can be envisaged if the two subunits are part of the same polypeptide chain, i.e. the protein has two domains connected by a disordered linker, which enables regulated contacts between the domains, such as in Hsp70 (Table 2, Figure 1D). This mechanism is termed "interdomain allostery",[13] and is also described in other cases, such as Na^+/H^+ exchanger regulatory factor (NHERF1),[102] and Membrane-associated guanylate kinase (MAGUK) proteins Sap97 ref.[107] and Dlg,[103] for example. This mechanism can be considered as a combination of classical and novel type of allostery, because the regulatory entity is brought together by the conformational freedom of the linker region, whereas allosteric communication across the two domains is realized in the traditional way, via an interdomain pathway.[13]

The ubiquitous chaperone Hsp70s fulfills basic functions in the cell[129] by binding misfolded partner proteins and assisting their (re)folding in an energy-dependent manner.[130,131] The protein has an N-terminal ATPase domain and a C-terminal substrate binding domain, connected by a disordered linker (Figure 1D). In the ADP-bound state the two domains tumble independently and the substrate-binding groove on the C-terminal domain is exposed. Exchange of ADP to ATP shifts the ensemble to a state in which the two domains predominantly form a closed unit. The two domains have a sparse but structurally contiguous group of co-evolving residues, which is termed a "sector", with reference to the conserved path of energetic coupling in single-domain allosteric proteins.[23] Here, this multidomain sector spanning the interdomain interface mediates the allosteric coupling between the ATP- and substrate-binding sites.[13]

4.3. Autoinhibition

Many proteins have a functional domain to which a disordered autoregulatory "tail" is bound in an intramolecular fashion, and keeps it in an "autoinhibited" state. A regulatory input – usually post-translational modification – releases the intramlocelular interaction and creates an activated ensemble. It should be noted that there is a fine distinction between autoinhibition and allosteric activation: the allosteric signal propagates from the site of modification to the active site of the domain, whereas autoinhibition occurs at the active site, i.e. it constitutes an important element of the overall activation mechanism. This distinction is even more important for the analogous "intrasteric" regulation outlined in the next chapter, which explains why "autoinhibited" and "intrasteric" should actually be considered allosteric. Autoinhibition is a common theme in regulation.[132]

A prime example of such allosteric activation is Vav proto-oncogene product, which relays signals from T-cell, B-cell and Fcγ receptors to the cytoskeleton and transcription regulation.[106] The protein is an autoinhibited, multidomain switch, which has a Dbl homology (DH) domain that functions as a guanine nucleotide exchange factor (GEF) for small GTPases in the Rho family (Figure 2C). GEF activity of Vav is kept at basal level by an adjacent acidic region (Ac) that folds back, binds to and inhibits the DH domain. The protein is activated by phosphorylation of Tyr174, which causes Ac to be displaced from DH. NMR studies have shown the presence of a sparsely populated excited state, where the helix is released and unfolded [106] in the autoinhibited state of the protein, which is a prerequisite for Tyr174 phosphorylation to take place. Therefore, internal dynamics controls the basal activity and activation of the autoinhibited DH domain, and allosteric regulatory information propagates from the Ac helix to the DH domain by order-to-disorder transition (Table 2).

A more complex case is the autoinhibition of The Src-family Tyr kinase Hck, which is involved in various signaling functions in hemopoietic cells, and is also implicated in cancer.[26] Like all kinases in the Src-family, Hck is a multidomain allosteric enzyme with and N-terminal SH4 domain, a unique domain, SH3 and SH2 domains, and a catalytic domain, all connected by flexible linkers. Inactive Hck is dominated by a closed (assembled) conformation maintained by intramolecular interactions between a C-terminal pTyr residue and the SH2 domain, and the Pro-rich linker and the SH3 domain (Figure 2D). Activation of the enzyme is induced by Tyr dephosphorylation and other modifications, which release intramolecular contacts and lead to a largely disassembled and more disordered active state. Recent SAXS experiments[26] showed an ensemble of distinct closed and open structures in both the inactive and active states (Figure 2D), therefore signaling due to dephosphorylation is best described as a shift in the conformational ensemble.[26] In a way similar to the case of interdomain allostery, Hck also represents a combination of traditional and novel types of allostery, because shifts in the population are converted to structure-based signal propagation in the catalytic domain (Table 2).

Autoinhibition is also the basic theme of the function of Smurf2, an E3 ubiquitin ligase involved in regulating transforming growth factor-β (TGF-β) signaling by targeting the adaptor protein Smad7 and TGF-β receptor kinase for degradation.[105] The protein also ubiquitinates and down-regulates itself. It has five domains (C2, WW1, WW2, WW3 and homologous to E6-AP carboxyl terminus (HECT)), which are separated by disordered linkers (Table 2). The intramolecular interaction between remote C2 and HECT domains keeps the protein in an inactive, more structured state.[105] Activation is initiated by binding of Smad7 to the HECT domain, which releases bound C2 and opens up the structure. In this more disordered, active state the protein can target nuclear substrates for ubiquitin-dependent degradation, and it can also translocate to the cytoplasm and target TGF-β receptor complex by anchoring to the plasma membrane due to PtdInsP binding by its exposed C2 domain. Again, the signal propagates a very long distance, more than 300 amino acids, without being mediated by an apparent pathway.

4.4. Intrasteric (active-site directed) regulation

Practically the same mechanism operates in many other regulatory proteins, such as kinases, phosphatases and metabolic enzymes, described as "intrasteric"[104,120,133] or "active-site directed" regulation.[108] These proteins have a pseudo-substrate domain (also termed intrasteric autoregulatory sequence, IARS) that folds back, and binds to the catalytic domain (Table 2). IARS regions are actually short binding motifs,[70,73,134] which reside in an IDR of the protein. The logic of activation is the same as autoinhibition above, because disruption of this interaction exposes the active site of the enzyme[120] as is often also observed in zymogen activation of proteolytic enzymes.[135-137]

This mechanism was considered distinct from allostery, because it was originally thought that these proteins combine allostery and intrasteric regulation by separating the mechanistic step of binding of an allosteric regulator to IARS from its dissociation from the active site (intrasteric activation, or, rather, de-inhibition). The essence of the mechanism of activation, however, is not (only) propagation of the signal from the "allosteric" site to IARS, but rather from both the allosteric site and IARS to the active site.

One of the best studied cases is importin-α (Table 2), a nuclear import factor that recognizes nuclear localization signals (NLSs). The protein has a large C-terminal domain composed of structured armadillo repeats, and a less structured N-terminal importin-β-binding domain containing an internal NLS region. This motif is bound to the NLS-binding site of the C-terminal receptor domain,[118] i.e., serves as an IARS. The closed structure represents the nuclear, low-affinity form, whereas in the cytoplasm it is converted to a high-affinity form through binding of the activator protein, importin-β, to the IARS, which exposes the NLS-binding site. Thereby, the regulatory input of importin-β binding at IARS is transduced into activation of the receptor domain by exposing it for productive interaction with NLS-containing cargo molecules.

Another classic case is "autoinhibition" of the Ser/Thr phsophatase calcineurin (PP2B, CaN), 91 residues of which are missing from its PDB structure due to local disorder.[119] This disordered region harbors its regulatory calmodulin (CaM) binding site. Structural disorder is critical for the activation of the enzyme, allowing CaM to wrap around this motif and to displace an 18-residue autoinhibitory sequence from the substrate-binding cleft of the A subunit of the enzyme. The activated enzyme is involved in myriads of processes ranging from hormone secretion to apoptosis,[138] and the direct link between its structural disorder (obligatory for CaM binding) and function (regulation of enzyme activity) was instrumental in suggesting that structural disorder is an intrinsic functional state of proteins.[65,139,140]

4.5. Motif exposure

Practically the same mechanism as autoinhibition and intrasteric regulation operates in proteins in which the functional readout of allosteric activation is the opposite, not the activation of a domain but rather functional exposure of a binding motif, due to which they are termed "motif exposure".[110] The basic theme in these proteins is also an autoinhibitory interaction between a binding domain and an adjacent motif embedded in an IDR, which is released upon regulatory input at the domain. The functional readout of activation, however, is mediated by the motif becoming available for other interaction

partners. This molecular logic underlies a suite of identical activation mechanisms, also termed variously "motif hiding",[141] "motif masking",[142] "cryptic disorder",[143] or "regulated unfolding",[144] as described for regulated nucleophosmin phosphorylation,[143] p53 acetylation,[110] exposure of p53 nuclear export signal (NES),[142] p53 acetylation site[110] and MapKap kinase 2 (MapKapK2) NLS,[145] for example.

The most intriguing aspect of MapKapK2 NLS and p53 NES is that they regulate the same cellular process (nuclear export/import) by the same basic mechanism (regulated binding of a motif to an adjacent domain) but by the opposite readout (sensing the motif not the domain) than autoinhibition previously mentioned in the case of importin-α. MapKapK2 plays a central role in inflammatory response, in which it is activated by p38 MAPK phosphorylation, due to which it translocates to the nucleus and phosphorylates its nuclear targets. MapKapK2 masks its NLS by folding back on the adjacent kinase domain. Upon phosphorylation, the NLS gets "unmasked" by dissociation from the domain and carries the MapKapK2-p38 complex into the nucleus.[145] A similar activation mechanism operates in p53, in which a leucine-rich NES is localized in the tetramerization domain.[142] The tetramerization domain occludes its NES, due to which the protein is retained in the nucleus. Activation occurs by converting tetramers into dimers and monomers, in which NES becomes exposed and p53 is translocated to the cytoplasm.

Although much less well characterized, similar mechanisms might operate in many IDPs in which post-translational modifications (PTMs) cannot occur in random succession. The open, exposed, and close to random structure of IDPs is amenable for modification and regulation by PTMs,[62,146-148] and the default expectation is that their distinct modifications do not interfere with each other. This is often not the case, however. For example, an Alzheimer-specific epitope of antibody AT100 in tau protein is generated by a particular sequence of phosphorylation events, at three sites Ser199, Ser202 and Thr205, followed at Thr212 by GSK-3 β, finished at Ser214 by PKA.[149] However, if Ser214 is phosphorylated first, Thr212 and adjacent sites become protected from phosphorylation, and the AT100 epitope is not formed. The most plausible explanation for this observation is the communication of phosphorylation sites through local conformational changes, which conforms to the logic of allostery. It should be noted that although antibody binding might not be considered a functional readout of the state of a protein (e.g. tau), given the importance of the exact structural state of tau in function (microtubule binding[150,151]) and disease (transition to paired helical filament in Alzheimer's disease[152-154]), the noted propagation of allosteric information in its ensemble may be of (patho)physiological significance. Whereas the underlying structural events have not yet been described, it is hardly questionable that signal propagation from one site to another in the structural ensemble of an IDP has taken place.

4.6. Signaling conduits

Allosteric signal propagation over much longer distances takes place in the completely disordered structural ensemble of the well-characterized cell-cycle dependent kinase (Cdk) inhibitor p27[Kip1] (Table 2), which is termed and described as a "signaling conduit".[109,155]

p27[Kip1] regulates cell division through interactions with the complexes Cdk2/CycA and Cdk2/CycE[156] at the G1-S phase boundary. Degradation of the inhibitor is initiated by two consecutive phosphorylation steps, first of Tyr88 by non-receptor Tyr kinases (NRTKs), followed by Thr187, catalyzed by the activated Cdk2/CycA complex itself.[117] In the inactive state, Tyr88 is locked at the ATP binding site of the Cdk2, but due to disorder of the inhibitor, it fluctuates in and out of the ATP binding pocket, making it accessible for phosphorylation by NRTKs.[117] Phospho-Tyr88 cannot bind to the active site of the enzyme, which has thus access to the remote Thr187 of p27[Kip1] still bound to the same Cdk2/CycA complex (Figure 3A). Due to its disorder and large scale structural fluctuations,[117] the inhibitor folds back with a certain probability and gets phosphorylated on Thr187, which signals for its ubiquitination and degradation. A key element of this signaling conduit is the propagation of a regulatory signal from Tyr88 to Thr187 through 100 disordered residues, due to fluctuations resulting in the remodeling of its structural ensemble. Intriguingly, this allosteric cross-talk between the two remote sites in the ensemble results from the interplay of three proteins in the complex, i.e. it is no longer the property of a single protein.

4.7. Signal propagation in the disordered ensemble

There are several other related examples of IDPs in which signal propagation occurs by conformational remodeling of their ensemble, resulting in an altered downstream allosteric readout (Table 2). These cases truly challenge our concepts of allostery, because often signal(s) propagate from multiple effector sites to multiple output sites without appreciable folding transitions in the ensemble. Some of the previous cases termed allosteric (e.g. Ets-1) have closely related mechanisms, the examples discussed here have never been explicitly termed as such.

Long-range signal propagation occurs in the Drosophila Hox transcription factor Ultrabithorax.[113] The protein has a 280-residue long, largely disordered region (non-homeodomain) and a DNA-binding C-terminal homeodomain (Table 2). The non-homeodomain region contains several short regulatory elements, an N-terminal 174 R (AER plus AC regions) region of about 170 amino acids in length, followed by a conserved YPWM motif and variable microexons encoding for b, I1 and I2 (inhibitory) regions. Deletion analysis has suggested a complex interplay between various regions, the YPWM region and the microexon I1 directly inhibiting DNA binding (approximately 2-fold), whereas I2 region inhibiting DNA binding (approximately 40-fold). Full-length Ultrabithorax, however, binds DNA only 2.5-fold weaker than its homeodomain, suggesting that the mostly disordered N-terminal R region neutralizes inhibitory interactions in a gradual manner, restoring DNA binding activity, either due to direct interaction with the DNA binding domain or the microexon regions, or both. This long-range energetic coupling can be allosterically regulated by the interaction of homedomain with DNA, and the non-homeodomain region with protein partners.[113]

An even more subtle intramolecular information transfer occurs in yeast, where DNA replication starts upon activation of the Clb-Cdc28 cyclin-Cdk complex, following the degradation of its fully disordered inhibitor, Sic1 (Table 2). Phosphorylation of Sic1 at several short sequence motifs termed Cdc4 phosphodegrons (CPDs) promotes its

interaction with Cdc4, the F-box subunit of an SCF ubiquitination ligase complex, which targets the protein for degradation by ubiquitination.[157] Sic1 contains 9 CPDs, phosphorylation of at least 6 of which acts in concert in promoting high-affinity cooperative binding to the single binding pocket on Cdc4 (Figue 3B). Multiply-phosphorylated Sic1 is engaged in a highly dynamic equilibrium with its single binding site, in which each phospho-CPD increases the probability of binding of all the others. This unique binding mode results in "ultrasensitive" binding,[112,157,158] in which modification of several CPD sites collectively modifies the behavior of all the others due to signal propagation and electrostatic cross-talk in the ensemble of structures.

A similar mechanism operates in the cystic fibrosis transmembrane conductance regulator (CFTR) chloride channel, which is best known for its mutation(s) in cystic fibrosis.[159]. The protein belongs to the ATP-binding cassette (ABC) superfamily, which typically contain two membrane-spanning domains and two nucleotide-binding domains. On the cytoplasmic side, CFTR has a unique regulatory (R) domain, which is about 200 amino acids in length and is largely disordered[160]. Full activation of the channel requires multiple motifs within R, without the involvement of any specific site. NMR studies have shown that these motifs, which mediate interactions with one of the nucleotide binding domains (NBD1), have a fractional helical propensity. This local helical propensity is reduced upon phosphorylation by PKA, which antagonizes binding at the single R-binding site of NBD1, and subsequent inhibition of the channel.[111] Phosphorylation at several sites cooperate in influencing the dynamic interaction that transiently engages different sites of the R region, which assumes the propagation of regulatory signal(s) between distinct sites and ultimately to the nucleotide-binding region of NBD1.

5. Extending the concept of allostery

All the examples cited (Table 2) suggest that structural disorder support various modes of allosteric regulation, which can hardly be reconciled with traditional models of allostery and/or traditional views of protein structure and function. In reflection of this shift in structure-function paradigm, the EAM based on domain coupling optimized by induced folding has been suggested.[10,11] Whereas this model emphasizes population shifts in the ensemble, we will show here that its central theme of enthalpic coupling between domains induced to fold is not essential for most of the examples discussed here. These multidomain regulatory proteins require a more extended model based on entropic coupling emerging from ensemble remodeling without invoking ligand-induced folding events. The transition is usually initiated by competition between signals and intermolecular recognition within the ensemble (Figure 4). Evidently, the mathematical formalism of this mechanism will need to be worked out.

5.1. Domain coupling optimized by induced folding

The mechanistic model suggested for the role of structural disorder in optimizing intramolecular site-to-site communication[10,11] rests on the assumption that ligand binding and concomitant induced folding promotes folding of adjacent domain(s), thus optimizing their binding interactions (Figure 4A). The model predicts that site-to-site

coupling is maximized when intrinsic disorder is present, which challenges a key mechanistic element of classical allostery, i.e. that two sites are coupled through a network of interactions that constitute an energetic pathway between them. This ensemble allosteric model (EAM) emphasizes that binding-induced reconfiguration (actually, induced folding) in the ensemble shifts preferences of the next binding event and actually increases the magnitude of coupling. This mechanism applies in several cases (e.g. Phd/Doc[114] and glucocorticoid receptor[12]), but the formalism developed is rather limiting, and will have to be extended to describe the complex allosteric proteins described here.

The large extent of coupling in EAM comes from assuming that the free energy of interaction between two domains is larger for the folded states than if any of them is disordered (unfolded). If we reduce the ensemble to the minimalistic and classical allosteric model of two states (R and T), or relate it simply to conformational fluctuations around an equilibrium state,[6,21,25] we might as well assume coupling of the same magnitude. Of course, the free energy of binding of folded states is intuitively much larger than that of unfolded states (based on the implicit assumption that the disordered state(s) cannot even be involved in specific binding). Binding, however, may occur even without folding, as captured by the concept of fuzziness.[78] In addition, the EAM formalism assumes that binding of the effector ligand is followed by a folding transition of the entire domain, which is rather demanding in terms of structure and is seldom encountered in the disorder literature in case of small ligands.

In addition, the key mechanistic element of the model is an enthalpic coupling (with a possible entropic component) of two binding regions, with positive coupling being its most natural readout.[10,12] The formalism needs to be extended to handle complex cases discussed here, in which disorder contributes to complex interaction patterns between different binding regions, domains and motifs, as suggested.[14,22]

5.2. Allostery by remodeling the ensemble

The key mechanistic invention in creating the modular allosteric proteins discussed (Table 2) is that their regulatory binding elements (domains and motifs) are separated by disordered linkers, which enable remodeling of their complex energy landscapes without the necessity of induced folding and coupled binding of the disordered region(s). This relieves demanding structural constraints of molecular recognition and enables the transfer of regulatory information between multiple distant elements. Internal binding competition and concomitant entropic freedom of the chain enables to create switches of practically unlimited variety and complexity: from all these, a novel type of allostery emerges (Table 2, Figure 4B, Figure 4C).

This is clearly the case for the proteins already termed allosteric (e.g. WASP, p53, Ets-1), even though their activation mechanism does not conform to the classical models. Interdomain allostery (e.g. HSP-70, ref.[13]) and intramolecular regulation (e.g. Sap97 ref.[107]) are clear extensions of the classical mechanism, in which structural transitions enabled by disorder are part of generating the allosteric entity. The "conduit-type" spread of regulatory information (e.g. p27[Kip1]) within the structural ensemble also clearly falls here because modification of one site (phosphorylation of Tyr88) affects the

modification of the other, Thr187, by internal competition, altering the entropic cost of modification.

A somewhat different logic applies to proteins that integrate multiple signals and are not yet termed allosteric (e.g. Sic1). Here multiple sites act cooperatively and modification of any of them alters behavior of all the others due to the dynamic binding equilibrium of the protein with a single binding pocket.[112] Regulatory information does not spread from one particular site to another, but among many sites in the disordered ensemble; other cases such as Ets-1,[99] which has actually been termed allosteric, and CFTR,[111] follow a very similar structural logic.

It is easy to formally show that autoinhibited and intrasteric regulatory proteins (e.g. Vav, importin-α) are also allosteric, although their IARS segment binds and inhibits the active site, suggesting that regulatory input and output occurs at the same site. As outlined above, however, the regulatory input (e.g. phosphorylation of Vav, partner binding of importin-α) at IARS occurs in the open, disordered conformation, which prevents re-binding of IARS at the active site. The input at the IARS site (which itself can be allosteric) affects the activity of the remote active site, making this relation allosteric, with the actual molecular mechanism being inhibition and de-inhibition. A similar logic applies in many other cases, such as autoinhibition and motif exposure, and more complex intramolecular regulation, such as of MAGUK proteins[103,107] and Smurf2.[105] In the case of Sap97, for example, remote regions (N-terminal S97N sequence, SH3/GUK domains, Table 2) interact in a U-shaped closed conformation, in which the inhibitory GUK-U5 interaction is relieved by the N-terminal region that promotes GKAP binding.[107] Similar, very complex intramolecular interaction patterns occur in other MAGUK proteins.[103]

The novelty of this general model (Figure 4) is that it separates binding regions and linkers, thus it does not set a limit on the enthalpic/entropic contribution of allosteric changes, as long as the altered state favors a particular downstream effect. Remodeling of the conformational ensemble may involve a transition to a more ordered state (disorder-to-order transition), a more disordered state (order-to-disorder transition), a combination of both, or, in certain cases, may proceed entirely within the disordered ensemble (Table 2). An interesting caveat to many of the above mechanisms is that often we cannot call an isolated protein allosteric, because energetic coupling between distinct sites only occurs within a complex (as apparent in the case of p27[Kip1] and Sic1, for example), i.e. allostery is associated with a particular topological arrangement of subunits within a complex.

6. Toward a quantitative description of novel allostery

6.1. Structural characterization

Describing the structural correlates of the novel allostery cases (Table 2) represents a special challenge, perhaps surpassing in complexity even that of describing the structural ensemble of "simple" IDPs.[46] These proteins often combine several folded domains and disordered regions, therefore traditional and novel structural approaches

have to be combined. Because the essence of their allosteric activation is a shift in the pattern of their intramolecular interactions (Figure 4), with a substantial and functionally important level of disorder, their structures are intermediary between tertiary and quaternary structures of ordinary proteins, termed recently "supertertiary".[91] Further, they have at least two distinct states, one naïve and one activated, both of which can be structural ensembles. Some complex signal integrators may even have multiple states, therefore we actually face the problem of solving an "ensemble of ensembles".

Of course, the adequate structural description of these signaling switches demands not only to solve a series of "static" structures, rather to ascribe dynamics and energetics to these states and their transitions. It is rather common that even in the naïve state, besides the inactive conformation (e.g. the autoinhibited closed conformation of Hck,[26] and Vav[106]) the activated state must also be present, as demanded by local accessibility to the input signal (e.g. a post-translational modification or an incoming binding partner). Determining the energetics separating the different states and dynamics of their transitions is therefore necessary for their full description.

These complexities demand that the structural characterization of these allosteric switches be approached by a combination of techniques[47,79] (cf. also section *Structural characterization of disorder*). Characterization of isolated domains and domain-domain or domain-motif interactions is best approached by traditional crystallography and NMR. Their disordered linkers can be characterized by NMR, complemented by low resolution techniques reporting on hydrodynamics behavior. This is even more important at the level of the whole protein, which can only be modeled by combining high resolution structural data on domains and interacting elements (crystallography, NMR), overall hydrodynamics behavior (SAXS, analytical ultracentrifugation (AU), size-exclusion chromatography (SEC)), structural ensemble description of disordered regions (NMR) and long-range topological information on the relative orientation of their segments (paramagnetic resonance enhancement (PRE), FRET, SAXS), integrated by advanced computational approaches possibly encompassing MD simulations[161-163] and/or statistical-mechanical descriptions.

The measurements can be collectively interpreted by generating a large number of random conformers, calculating the parameters for each conformer, and applying a genetic optimization procedure to select a limited number of conformers that collectively provide an adequate description of the ensemble. This type of approach was successfully applied in the case of Sic1,[89] p27^{Kip1} ref.[155], Hck,[26] CFTR,[111] tau protein[83] and measles virus nucleoprotein,[90] for example. It is to be noted, that a full kinetic and thermodynamic description of allosteric regulation has not been attempted in any of the cases.

Insight from these approaches may be further advanced by the application of a range of single-molecule approaches deployed recently for the characterization of IDPs. Single-molecule fluorescence resonance energy transfer (smFRET), for example, demonstrated multiple interactions in p53, between its N-terminal domain and DNA binding domains,[51] and it even enabled the characterization of the dynamics of soluble NSF attachment protein receptor (SNARE) complexes in live cells.[48] It also enabled to address the "supertertiary" structure of the complex regulatory protein PSD95.[164]

Atomic-force microscopy (AFM) has also been recently applied for studying the structure and structural changes of the complex motor protein myosin V with disordered linkers[165] and the structural ensemble and unfolding patterns of α-synuclein.[50]

6.2. Basic model(s) of novel allostery

Due to these high demands, there are only a few cases where structural description already enables to elucidate the underlying allosteric mechanism, such as in the case of Hck (both the closed and open states are characterized[26,166]), Sic1 (the ensemble in the bound state is described[89]), and p27[Kip1] (the conduit has been modeled based on high-resolution and low-resolution data[117,155,167]). In several other cases, we only have a partial picture of their structural and dynamic behavior: Smurf2 (the structure of the closed state characterized by the interaction of two domains, the long linker is not described and the open conformations are not characterized[105]), Ets-1 (the cross-talk of allosteric sites has been mapped and structural disorder in both states established, the ensemble itself not approached[99]), and WASP (the structure of two binding regions in complex with each other,[98,127] and parts of the activated structure[127,168] have been solved, without resolving the intervening long disordered linker region). These structural studies, nevertheless, represent the first steps toward achieving a quantitative structural description of this complex allosteric phenomenon.

Whereas the examples shown represent almost unlimited structural varieties, they can all be simplified to the combination of simple switch units, composed of an input/effector region and an output/decoding region, connected by a disordered linker (Figure 4B). There is experimental evidence for structural disorder of the linker in most of the cases (cf. Table 2), and often the protein is found in the DisProt database.[39] Bioinformatic predictions also underscore that their disorder usually falls between the functional elements of the protein (Table 2), the length of this IDR can range from 18 (importin-α) to up to 300 (Smurf2) amino acids with a very high (usually >50%) predicted disorder.

The essence of the ensuing mechanism is the interaction of two regions mediated by a disordered linker, which obviates induced folding upon ligand binding (Figure 4B). The regulatory input (post-translational modification, binding of a small ligand or another protein), competes (but may also promote) intramolecular interaction, thus shifts the population of conformations in favor of a particular downstream binding or catalytic event. The key element of the mechanism is a change in the entropic penalty of this readout. Due to the underlying modularity, very complex switches can be easily constructed, as illustrated by a schematic switch that gets inhibited by the coincidence of the signal of a small molecule and the removal of a PTM, which cause the binding of an autoinhibitory segment (Figure 4C).

As clearly illustrated by the foregoing examples, the three elements must not even be clearly separated, i.e. the disordered linker might contain binding elements, motifs (such as Vav) or domains (such as WASP) which might undergo induced folding upon binding, or it may harbor modification sites (e.g. Sic1) and the whole mechanism may occur in

the disordered state, without a discernible folding transition (e.g. Sic1, Ets-1, p27[Kip1], Ultrabitorax).

6.3. Quantitative thermodynamic description

As suggested, the novel allosteric mechanism rests on the reorganization of the ensemble, which basically alters the entropy of subsequent readout mechanisms, primarily the binding of a novel partner. As suggested for all the examples and clearly illustrated on Figure 2D and Figure 4, the naïve (inactive) form of the protein is an ensemble itself, which contains the active form in a certain proportion (otherwise it could not be activated by an incoming signal). By internal competition, the signal shifts the distribution of states in the ensemble toward the activated state, in which the binding region is more accessible to its partner. The enthalpy of this interaction does not increase, but its entropic penalty decreases significantly. We need detailed thermodynamic (enthalpy and entropy of binding of different states) to elaborate on this plausible scenario. For example, in the case of p27[Kip1], a very large ΔS of folding-upon-binding of the inhibitor to CycA/Cdk2 has been reported ($-T\Delta S = +28.6$ kcal mol^{-1} ref.[156]). Allosteric activation occurs by phosphorylation of Tyr88 of p27[Kip1], which disrupts its favorable interactions with the ATP binding pocket of Cdk2 (Figure 3A), and the entropy of the C-terminus, given its flat energy landscape, enables Cdk2 to phosphorylate Thr187 and, later, it permits its ubiquitination by the cognate E3 ligase.[117] Whereas entropy in the case of phosphorylated p27[Kip1] has not been reported, biochemical data shows that Tyr88 is accessible for phosphorylation by Abl, which implies local dynamics in the bound state, and NMR/SAXS data on full-length p27Kip1 shows that the C-terminus is highly dynamic in the complex.[155]

7. Multistery: a whole new world of regulation

These general principles and the actual examples (Table 2) show that regulatory proteins built as a combination of domains, motifs and linkers can display complex regulatory behavior, perfectly fit into the signaling networks of higher eukaryotes. Whereas detailed and accurate structural and thermodynamic models of this novel type of allostery are needed, several general conclusions can already been drawn on the ensuing regulatory features and advantages.

(i) Due to the combination of modularity and structural disorder, a constellation of binding elements, linker regions and input signals of practically any complexity, enabling the integration of multiple signals, can be envisioned. In the actual cases, for example, WASP is activated by Cdc42, PIP$_2$, binding partners and phosphorylation,[98,169] Ets-1,[99] Sic-1 ref.[105,112] and CFTR[105,112] are activated by multi-site phosphorylation, Mdm2 is activated by phosphorylation and binding partners L5, L11, and L23,[121] Hck is activated by dephosphorylation and autophosphorylation events.[26] Signaling input from these multiple effectors converges on an output site, but they may even act on multiple sites.

(ii) These examples also show that the regulatory signal may come in many forms, such as a small ligand (e.g. PIP$_2$ for WASP), another protein (e.g. importin-β for importin-α or

Smad7 for Smurf2), or post-translational modification (e.g. phosphorylation of p27[Kip1]), in many combinations (Table 2). The individual signals might only partially activate the switch, whereas their synergy is required to bring about full and lasting activation of the protein. This makes these allosteric effectors function as integrators of multiple input signals, shifting the emphasis of allostery from unimodal regulation to signal integration.

(iii) Classical allosteric proteins rely on energetic coupling propagating through the structure mediated by a co-evolving cluster of residues,[13,23] the evolutionary generation of which is rather slow and demanding, and cannot be exchanged between different proteins. On the contrary, multidomain allosteric regulatory switches based on disorder are modular, i.e. they may arise in evolution by simple module exchange, as suggested by the appearance of many common protein modules in these cases (e.g. SH2-, SH3-, kinase-, DNA-binding-, C2- WH2- and DH-domains, for example).

(iv) The principle of modularity also lends itself for the effective design of allosteric switches of altered or even novel activities. Modularity and disorder permitted the activity of WASP to be completely redesigned.[169] Various combinations of signaling domains (PDZ and/or SH3) and their cognate peptides have been inserted into the protein, to tune the activity of the output domain (VCA) to signals unrelated to the original function. p53 could also be rationally redesigned by inserting recognition segments into its intrinsically disordered domains to create variants activated up to 100-fold by effectors (proteases or antibodies) unrelated to the original function of the protein.[116]

(v) As opposed to the "on-off" logic of traditional allosteric proteins, this novel type of allostery enables graded (rheostat-type) responses to multiple signals in regulation[99] and it may also give rise to switch-like ultrasensitive responses.[112,158]

(vi) These manifold regulatory advantages pertain to a wide variety and diversity of cellular processes and pathways in which these switches have already been described (Table 2). The examples cited appear in transcription (Ets-1), signal transduction (Hck, CaN), apoptosis (p53), cytoskeleton reorganization (WASP, Vav), cell cycle regulation (Sic1, p27[Kip1]), transport processes (importin-α), protein degradation (Smurf2), and probably many others. Due to the apparent advantages, it is not unwarranted to suggest that related mechanisms are widespread in eukaryotes.

(vii) As best illustrated by the example of p27[Kip1], this allosteric mechanism may only actually manifest itself in a multisubunit complex, not in an isolated protein. A high level of structural disorder has been observed in scaffold proteins,[56] which function by bringing subsequent enzymes in signaling cascades to proximity, often even activating them in a classical allosteric.[170] Due to the structural disorder involved, different proteins in the complex can interact and modify each other's activity, very much like different regions of the multidomain proteins described here. Allostery in this case belongs to the entire complex, and it may be a rather general consequence of the prominent role of IDPs in assembling complexes.[171,172] Such complex allostery may simply explain the activation of transmembrane receptors (e.g. receptor tyrosine kinases), in which ligand driven homodimerization brings cytoplasmic regions in proximity, and the subsequent transphosphorylation enabled by a high level of structural disorder in the cytoplasmic tails[173] is the key mechanistic element of signal

propagation from the extracellular ligand binding domain to the intracellular signaling region. In the case of such multisubunit allosteric switches, another level of regulation, stoichiometry, may also come into play, due to which they can switch between agonistic and antagonistic regulatory modes.[101,174]

8. Conclusion

An important aspect of the novel allosteric mechanism is that it suggests a structural and functional rationale for the co-existence of a variety of elements, domains, motifs and linkers, in complex regulatory proteins. It has direct implications on the role of structural disorder, which is usually viewed only in terms of isolated functional elements, such as linkers and recognition elements.[42,66,94,95,175] Given that about 50-60% of eukaryotic proteins have at least one long disordered region,[176] there are at least several thousand[70,134] short motifs in proteomes and about two third of eukaryotic proteins have multiple domains,[92] it is very likely that allostery based on modularity and disorder is a widespread and general regulatory phenomenon in eukaryotes. In accord, this novel concept may offer a new way to rationalize the function of large modular proteins,[177] and help elucidate how structural disorder is evolutionarily tuned to orchestrate complex cellular regulatory events. Because these proteins are often involved in diseases (primarily in cancer[178,179]), their full structural-functional understanding may also lead to the development of better drugs and therapies.

Legends to figures

Figure 1. *Illustrative examples of traditional allosteric models*

This figure illustrates the evolution of the concept of allostery from static descriptions of cooperativity in homo-oligomeric proteins to models incorporating dynamics and structural disorder. **(A)** Initial phenomenological descriptions emphasized the existence of two states (tense, blue circle and relaxed, pink square) linked through subunit interactions in oligomeric proteins, such as hemoglobin, the classical allosteric (cooperative) oxygen-binding protein. The basic assumption of the model is that the ligand (oxygen) binds to the R state stronger then to the T state, therefore it initiates a transition in the complex. **(B)** These low-resolution models were later filled with structural detail, when structures corresponding to the two states were solved. Here the two conformations of catabolite activator protein (CAP), apo (orange) and cAMP-bound holo (light blue). The allosteric conformational change promotes DNA-binding (adapted with permission from ref. 7). **(C)** Allosteric transduction of regulatory information might even occur without a change in structure, mediated by altered dynamics of the protein. Here the switch domain of NtrC shown, phosphorylation of which significantly alters the dynamics of the structure, altering to the entropic component of substrate binding. Dynamics were directly characterized by NMR (shown by chemical shift perturbation, illustrated by a continuous color scale. Adapted with permission from ref. 180. Copyright (2001) The American Association for the Advancement of Science). **(D)** Allosteric communication may also occur between two separate domains of a protein that only come together by virtue of a disordered linker region, as seen in the case of Hsp-70. In the open, more disordered conformation the two domains are separated,

whereas in the closed state shown the two domains create a continuous "interdomain sector" that mediates allosteric coupling. Because here a classical allosteric entity forms through the involvement of a disordered linker, this model can be considered as a mixture of classical and novel mechanistic elements (PDB: 3d2f, adapted with permission from ref. 13. Copyright (2010) Nature Publishing Group).

Figure 2. *Allostery enabled by structural disorder of multidomain proteins*

Long-range flow of regulatory information can occur due to remodeling of the conformational ensemble of complex proteins. These cases are termed allosteric in the literature but they do not obey the classical allosteric mechanisms. **(A)** In the inactive state of WASP, the G-protein binding domain (GBD, cyan) binds the VCA region (red). Upon Cdc42 binding to GBD, reinforced by other signals such as PIP2 and phosphorylation, VCA is released and it can bind and activate Arp2/3 (light blue oval) to promote actin polymerization (images are based on the PDB structure 1ej5). **(B)** DNA-binding region (grey) of Ets-1 is regulated by a loosely associated regulatory region (SRR, green/red line) that harbors multiple phosphorylation sites (red dots). Gradual dephosphorylation of these sites loosens the interaction with the DNA-binding region, which makes it more dynamic and better fit with DNA binding (adapted with permission from ref. 181. Copyright (2010) National Academy of Sciences, U.S.A.). **(C)** Vav proto-oncogene has a Dbl homology (DH) domain (blue), the activity of which is kept at basal level by the adjacent acidic region (Ac, red). When the protein is phosphorylated at Ac, it dissociates from DH which then functions as a GEF for Rho GTPases (adapted with permission from ref. 106. Copyright (2008) Nature Publishing Group). **(D)** The inactive state of the Tyr-kinase Hck is dominated by a closed conformation maintained by intramolecular domain-motif interactions. Upon activation initiated by pTyr dephosphorylation, Hck undergoes a transition to a heterogeneous and more disordered active state (in which structures 1 through 9 are all present, adapted with permission from ref. 26). All these cases are termed allosteric in the literature, because a regulatory signal propagates from (multiple) effector sites (red arrows) to their output site (green arrow) via remodeling the structural ensemble.

Figure 3. *Signal propagation through a disordered structural ensemble*

Signal propagation via conformational remodeling of the disordered ensemble is often not termed allostery, although the essence of the mechanism is an energetic coupling between remotes sites in the protein, or protein complex. **(A)** p27[Kip1] regulates cell cycle by inhibiting the Cdk2/CycA complex. Bound p27[Kip1] can be phosphorylated at Tyr88 by NRTKs, due to which it is ejected from the active site of Cdk2, so it can "fold back" and get phosphorylated on Thr187, while still bound, by the same Cdk2/CycA complex. The state of Thr187 is thus decided by that of Tyr88, i.e. allosteric signal propagates between these two sites through the ensemble of conformations of Cdk2/CycA-bound p27[Kip1] (Adapted with permission from ref. 109. Copyright (2008) American Chemical Society). **(B)** Phosphorylation and subsequent degradation of the inhibitor Sic1 is initiated by the Clb-Cdc28 cyclin-Cdk complex. Phosphorylation of any 6 of its specific motifs, phosphodegrons (CPDs), promotes its interaction with the Cdc4 subunit of SCF ubiquitination ligase complex.[157] Multiply-phosphorylated Sic1 is in a dynamic equilibrium with its binding site, i.e. modification of any of the sites is communicated to all the others through the ensemble (adapted with permission from ref. 112. Copyright

(2008) National Academy of Sciences, U.S.A.). On both panels, regulatory signal propagates from (multiple) effector sites (red arrows) to their output site (green arrow) via remodeling the structural ensemble.

Figure 4. *Generalized models of allostery based on disorder*

(A) In the ensemble allosteric model (EAM)[10,12]), binding of a ligand (blue circle) promotes folding of a cognate domain (purple rectangle). Its structural coupling to the other domain (yellow rectangle) causes the ensemble to redistribute to prefer states in which both domains are folded, facilitating binding of the other ligand. The primary mechanism is conformational coupling between the two domains, which can cause positive (but also negative) coupling. Induced folding of disordered domain is thought to increase the magnitude of (enthalpic) coupling. **(B)** In the novel "entropic" model the elementary allosteric unit is composed of an effector/input region (yellow oval) connected to an output region (purple object) by a disordered linker (green line), in which the two regulatory modules may also be disordered prior to partner binding. The signal (blue circle, a PTM, small ligand, another protein or macromolecular binding partner) competes with binding, thus redistributes the ensemble so that the output region becomes accessible for partners. The key thermodynamic element of activation is a decreased entropic penalty of binding in the altered ensemble. **(C)** This model can easily accommodate classical allosteric coupling between interacting domains, and it does not require induced folding (disorder-to-order transition) to occur. The ensuing allosteric coupling can manifest itself also in multisubunit complexes, and, due to the lack of structural constraints, a switch of any complexity can be easily built. Here a simple example is shown in which activity of the domain (purple object) is switched off by binding of a ligand to an autoinhibitory region (yellow oval). Switches that integrate multiple signals can be easily built on this principle.

Acknowledgements

References

(1) Bohr, C.; Hasselbalch, K.; Krogh, A. *Skand. Arch. Physiol.* **1904**, *16,* 402.
(2) Adelberg, E. A.; Umbarger, H. E. *J. Biol. Chem.* **1953**, *205*, 475.
(3) Koshland, D. E., Jr.; Nemethy, G.; Filmer, D. *Biochemistry* **1966**, *5*, 365.
(4) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88.
(5) Fenton, A. W. *Trends Biochem. Sci.* **2008**, *33*, 420-5.
(6) Kern, D.; Zuiderweg, E. R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748.
(7) Popovych, N.; Tzeng, S. R.; Tonelli, M.; Ebright, R. H.; Kalodimos, C. G. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6927.
(8) Petit, C. M.; Zhang, J.; Sapienza, P. J.; Fuentes, E. J.; Lee, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18249.
(9) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins* **2004**, *57*, 433.
(10) Hilser, V. J.; Thompson, E. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8311.
(11) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. *Annu. Rev. Biophys.* **2012**, *41*, 585.
(12) Motlagh, H. N.; Li, J.; Thompson, E. B.; Hilser, V. J. *Biochem. Soc. Trans.* **2012**, *40*, 975.
(13) Smock, R. G.; Rivoire, O.; Russ, W. P.; Swain, J. F.; Leibler, S.; Ranganathan, R.; Gierasch, L. M. *Mol. Syst. Biol.* **2010**, *6*, 414.
(14) Fuxreiter, M. *Mol. Biosyst.* **2011**, *8*, 168.
(15) Smith, J. L.; Freebern, W. J.; Collins, I.; De Siervi, A.; Montano, I.; Haggerty, C. M.; McNutt, M. C.; Butscher, W. G.; Dzekunova, I.; Petersen, D. W.; Kawasaki, E.; Merchant, J. L.; Gardner, K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11554.
(16) Changeux, J. P.; Edelstein, S. J. *Science* **2005**, *308*, 1424.
(17) Baldwin, J.; Chothia, C. *J. Mol. Biol.* **1979**, *129*, 175.
(18) Iwata, S.; Kamata, K.; Yoshida, S.; Minowa, T.; Ohta, T. *Nat. Struct. Biol.* **1994**, *1*, 176.
(19) Otsuka, J.; Kunisawa, T. *Adv. Biophys.* **1978**, *11*, 53.
(20) Barford, D.; Johnson, L. N. *Nature* **1989**, *340*, 609.
(21) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831.
(22) Swain, J. F.; Gierasch, L. M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 102.
(23) Lockless, S. W.; Ranganathan, R. *Science* **1999**, *286*, 295.
(24) Pan, H.; Lee, J. C.; Hilser, V. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12020.
(25) Tzeng, S. R.; Kalodimos, C. G. *Nature* **2009**, *462*, 368.
(26) Yang, S.; Blachowicz, L.; Makowski, L.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15757.
(27) Bai, F.; Branch, R. W.; Nicolau, D. V., Jr.; Pilizota, T.; Steel, B. C.; Maini, P. K.; Berry, R. M. *Science* **2010**, *327*, 685.
(28) Selwood, T.; Jaffe, E. K. *Arch. Biochem. Biophys.* **2012**, *519*, 131.
(29) Lee, L. M.; Krupka, R. M.; Cook, R. A. *Biochemistry* **1973**, *12*, 3503.
(30) Bhattacharyya, M.; Ghosh, A.; Hansia, P.; Vishveshwara, S. *Proteins* **2010**, *78*, 506.
(31) Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J. *Cell* **2006**, *125*, 1137.
(32) Freiburger, L. A.; Baettig, O. M.; Sprules, T.; Berghuis, A. M.; Auclair, K.; Mittermaier, A. K. *Nat. Struct. Mol. Biol.* **2011**, *18*, 288.
(33) Reichheld, S. E.; Yu, Z.; Davidson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 22263.

(34)   Selwood, T.; Tang, L.; Lawrence, S. H.; Anokhina, Y.; Jaffe, E. K. *Biochemistry* **2008**, *47*, 3245.

(35)   Dyson, H. J. *Q. Rev. Biophys.* **2011**, *44*, 467.

(36)   Tompa, P. *Curr. Opin. Struct. Biol.* **2011**, *21*, 419.

(37)   Uversky, V. N.; Dunker, A. K. *Biochim. Biophys. Acta* **2010**, *1804*, 1231.

(38)   Tompa, P. *Structure and function of intrinsically disordered proteins*; CRC Press (Taylor and Francis Group): Boca Raton, FL, 2009.

(39)   Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Research.* **2007**, *35*, D786.

(40)   He, B.; Wang, K.; Liu, Y.; Xue, B.; Uversky, V. N.; Dunker, A. K. *Cell Research.* **2009**, *19*, 929.

(41)   Monastyrskyy, B.; Fidelis, K.; Moult, J.; Tramontano, A.; Kryshtafovych, A. *Proteins* **2011**, *79 Suppl 10*, 107.

(42)   Pancsa, R.; Tompa, P. *PLoS One* **2012**, *7*, e34687.

(43)   Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337*, 635.

(44)   Tompa, P.; Dosztanyi, Z.; Simon, I. *J. Proteome Res.* **2006**, *5*, 1996.

(45)   Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Obradovic, Z. *J. Proteome Res.* **2007**, *6*, 1882.

(46)   Fisher, C. K.; Stultz, C. M. *Curr. Opin. Struct. Biol.* **2011**, *21*, 426.

(47)   Eliezer, D. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23.

(48)   Sakon, J. J.; Weninger, K. R. *Nat. Methods* **2010**, *7*, 203.

(49)   Bodart, J. F.; Wieruszeski, J. M.; Amniai, L.; Leroy, A.; Landrieu, I.; Rousseau-Lescuyer, A.; Vilain, J. P.; Lippens, G. *J. Magn. Reson.* **2008**, *192*, 252.

(50)   Sandal, M.; Valle, F.; Tessari, I.; Mammi, S.; Bergantino, E.; Musiani, F.; Brucale, M.; Bubacco, L.; Samori, B. *PLoS Biol.* **2008**, *6*, e6.

(51)   Huang, F.; Rajagopalan, S.; Settanni, G.; Marsh, R. J.; Armoogum, D. A.; Nicolaou, N.; Bain, A. J.; Lerner, E.; Haas, E.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106,* 20758.

(52)   Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. *Nat. Genet.* **2000**, *25*, 25.

(53)   Iakoucheva, L.; Brown, C.; Lawson, J.; Obradovic, Z.; Dunker, A. *J. Mol. Biol.* **2002**, *323*, 573.

(54)   Minezaki, Y.; Homma, K.; Kinjo, A. R.; Nishikawa, K. *J. Mol. Biol.* **2006**, *359*, 1137.

(55)   Tantos, A.; Han, K. H.; Tompa, P. *Mol. Cell. Endocrinol.* **2012**, *348*, 457.

(56)   Balazs, A.; Csizmok, V.; Buday, L.; Rakacs, M.; Kiss, R.; Bokor, M.; Udupa, R.; Tompa, K.; Tompa, P. *FEBS J.* **2009**, *276*, 3744.

(57)   Szollosi, E.; Bokor, M.; Bodor, A.; Perczel, A.; Klement, E.; Medzihradszky, K. F.; Tompa, K.; Tompa, P. *J. Proteome Res.* **2008**, *7*, 2291.

(58)   Lu, X.; Hamkalo, B.; Parseghian, M. H.; Hansen, J. C. *Biochemistry* **2009**, *48*, 164.

(59)   Tompa, P.; Csermely, P. *FASEB J.* **2004**, *18*, 1169.

(60)   Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2006**, *45*, 6873.

(61)   Kriwacki, R. W.; Hengst, L.; Tennant, L.; Reed, S. I.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11504.

(62)   Cox, C. J.; Dutta, K.; Petri, E. T.; Hwang, W. C.; Lin, Y.; Pascal, S. M.; Basavappa, R. *FEBS Lett.* **2002**, *527*, 303.

(63)   Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54.

(64)   Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 1.

(65)   Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graphics Modelling* **2001**, *19*, 26.

(66)   Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527.

(67)   von Ossowski, I.; Eaton, J. T.; Czjzek, M.; Perkins, S. J.; Frandsen, T. P.; Schulein, M.; Panine, P.; Henrissat, B.; Receveur-Brechot, V. *Biophys J.* **2005**, *88*, 2823.

(68)   Daughdrill, G. W.; Narayanaswami, P.; Gilmore, S. H.; Belczyk, A.; Brown, C. J. *J. Mol. Evol.* **2007**, *65*, 277.

(69)   Rosenblum, G.; Van den Steen, P. E.; Cohen, S. R.; Grossmann, J. G.; Frenkel, J.; Sertchook, R.; Slack, N.; Strange, R. W.; Opdenakker, G.; Sagi, I. *Structure* **2007**, *15*, 1227.

(70)   Gould, C. M.; Diella, F.; Via, A.; Puntervoll, P.; Gemund, C.; Chabanis-Davidson, S.; Michael, S.; Sayadi, A.; Bryne, J. C.; Chica, C.; Seiler, M.; Davey, N. E.; Haslam, N.; Weatheritt, R. J.; Budd, A.; Hughes, T.; Pas, J.; Rychlewski, L.; Trave, G.; Aasland, R.; Helmer-Citterich, M.; Linding, R.; Gibson, T. J. *Nucleic Acids Res.* **2009**, *38*, D167.

(71)   Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J. Mol. Biol.* **2004**, *338*, 1015.

(72)   Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.; Uversky, V. N. *J. Mol. Biol.* **2006**, *362*, 1043.

(73)   Fuxreiter, M.; Tompa, P.; Simon, I. *Bioinformatics* **2007**, *23*, 950.

(74)   Buljan, M.; Chalancon, G.; Eustermann, S.; Wagner, G. P.; Fuxreiter, M.; Bateman, A.; Babu, M. M. *Mol. Cell* **2012**, *46*, 871.

(75)   Stein, A.; Aloy, P. *PLoS ONE* **2008**, *3*, e2524.

(76)   Tompa, P.; Fuxreiter, M.; Oldfield, C. J.; Simon, I.; Dunker, A. K.; Uversky, V. N. *Bioessays* **2009**, *31*, 328.

(77)   Renault, L.; Bugyi, B.; Carlier, M. F. *Trends Cell Biol.* **2008**, *18*, 494.

(78)   Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2.

(79)   Bernadó, P.; Svergun, D. I., *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation*, Uversky, V. N.; Longhi, S., Eds., John Wiley & Sons, Inc.: Hoboken, New Jersey, **2010,** 451.

(80)   Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 1270.

(81)   Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.

(82)   Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359.

(83)   Mukrasch, M. D.; Bibow, S.; Korukottu, J.; Jeganathan, S.; Biernat, J.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. *PLoS Biol* **2009**, *7*, e34.

(84)   Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476.

(85)   Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 18314.

(86)   Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J. R.; Jensen, M. R.; Segard, S.; Bernado, P.; Charavay, C.; Blackledge, M. *Bioinformatics* **2012**, *28*, 1463.

(87)   Schneider, R.; Huang, J. R.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon, L.; Jensen, M. R.; Blackledge, M. *Mol. Biosyst.* **2012**, *8*, 58.

(88) Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5762.

(89) Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D. *Structure* **2010**, *18*, 494.

(90) Jensen, M. R.; Communie, G.; Ribeiro, E. A., Jr.; Martinez, N.; Desfosses, A.; Salmon, L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; Ruigrok, R. W.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 9839.

(91) Tompa, P. *Nature Chem. Biol.* **2012**, *8*, 597.

(92) Ekman, D.; Bjorklund, A. K.; Frey-Skott, J.; Elofsson, A. *J. Mol. Biol.* **2005**, *348*, 231.

(93) Neduva, V.; Russell, R. B. *FEBS Lett.* **2005**, *579*, 3342.

(94) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197.

(95) Tompa, P. *FEBS Lett.* **2005**, *579*, 3346.

(96) Pawson, T.; Nash, P. *Science* **2003**, *300*, 445.

(97) Seet, B. T.; Dikic, I.; Zhou, M. M.; Pawson, T. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 473.

(98) Leung, D. W.; Rosen, M. K. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5685.

(99) Pufall, M. A.; Lee, G. M.; Nelson, M. L.; Kang, H. S.; Velyvis, A.; Kay, L. E.; McIntosh, L. P.; Graves, B. J. *Science* **2005**, *309*, 142.

(100) Smock, R. G.; Gierasch, L. M. *Science* **2009**, *324*, 198.

(101) Ferreon, A. C.; Ferreon, J. C.; Wright, P. E.; Deniz, A. A. *Nature* **2013**, *498*, 390.

(102) Li, J.; Callaway, D. J.; Bu, Z. *J. Mol. Biol.* **2009**, *392*, 166.

(103) Qian, Y.; Prehoda, K. E. *J. Biol. Chem.* **2006**, *281*, 35757.

(104) Kemp, B. E.; Pearson, R. B. *Biochim. Biophys. Acta* **1991**, *1094*, 67.

(105) Wiesner, S.; Ogunjimi, A. A.; Wang, H. R.; Rotin, D.; Sicheri, F.; Wrana, J. L.; Forman-Kay, J. D. *Cell* **2007**, *130*, 651.

(106) Li, P.; Martins, I. R.; Amarasinghe, G. K.; Rosen, M. K. *Nat. Struct. Mol. Biol.* **2008**, *15*, 613.

(107) Wu, H.; Reissner, C.; Kuhlendahl, S.; Coblentz, B.; Reuver, S.; Kindler, S.; Gundelfinger, E. D.; Garner, C. C. *EMBO J.* **2000**, *19*, 5740.

(108) Kobe, B.; Kemp, B. E. *Nature* **1999**, *402*, 373.

(109) Galea, C. A.; Wang, Y.; Sivakolundu, S. G.; Kriwacki, R. W. *Biochemistry* **2008**, *47*, 7598.

(110) Ceskova, P.; Chichger, H.; Wallace, M.; Vojtesek, B.; Hupp, T. R. *J. Mol. Biol.* **2006**, *357*, 442.

(111) Baker, J. M.; Hudson, R. P.; Kanelis, V.; Choy, W. Y.; Thibodeau, P. H.; Thomas, P. J.; Forman-Kay, J. D. *Nat. Struct. Mol. Biol.* **2007**, *14*, 738.

(112) Mittag, T.; Orlicky, S.; Choy, W. Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17772.

(113) Liu, Y.; Matthews, K. S.; Bondos, S. E. *J. Biol. Chem.* **2008**, *283*, 20874.

(114) Garcia-Pino, A.; Balasubramanian, S.; Wyns, L.; Gazit, E.; De Greve, H.; Magnuson, R. D.; Charlier, D.; van Nuland, N. A.; Loris, R. *Cell* **2010**, *142*, 101.

(115) Ayed, A.; Mulder, F. A.; Yi, G. S.; Lu, Y.; Kay, L. E.; Arrowsmith, C. H. *Nat. Struct. Biol.* **2001**, *8*, 756.

(116) Geddie, M. L.; O'Loughlin, T. L.; Woods, K. K.; Matsumura, I. *J. Biol. Chem.* **2005**, *280*, 35641.

(117) Grimmler, M.; Wang, Y.; Mund, T.; Cilensek, Z.; Keidel, E. M.; Waddell, M. B.; Jakel, H.; Kullmann, M.; Kriwacki, R. W.; Hengst, L. *Cell* **2007**, *128*, 269.

(118) Kobe, B. *Nat. Struct. Biol.* **1999**, *6*, 388.

(119) Kissinger, C. R.; Parge, H. E.; Knighton, D. R.; Lewis, C. T.; Pelletier, L. A.; Tempczyk, A.; Kalish, V. J.; Tucker, K. D.; Showalter, R. E.; Moomaw, E. W.; Gastinel, L. N.; Habuka, N.; Chen, X.; Maldonado, F.; Barker, J. E.; Bacquet, R.; Villafranca, J. E. *Nature* **1995**, *378*, 641.

(120) Shimizu, H.; Hupp, T. R. *Trends Biochem. Sci.* **2003**, *28*, 346.

(121) Dai, M. S.; Lu, H. *J. Biol. Chem.* **2004**, *279*, 44475.

(122) Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Dunker, A. K. *Proteins* **2005**, *61 Suppl 7*, 176.

(123) Hilser, V. J.; Thompson, E. B. *J. Biol. Chem.* **2011**, *286*, 39675.

(124) Garza, A. S.; Khan, S. H.; Moure, C. M.; Edwards, D. P.; Kumar, R. *PLoS One* **2011**, *6*, e25875.

(125) Brodie, J.; McEwan, I. J. *J. Mol. Endocrinol.* **2005**, *34*, 603.

(126) Peterson, J. R.; Bickford, L. C.; Morgan, D.; Kim, A. S.; Ouerfelli, O.; Kirschner, M. W.; Rosen, M. K. *Nat. Struct. Mol. Biol.* **2004**, *11*, 747.

(127) Abdul-Manan, N.; Aghazadeh, B.; Liu, G. A.; Majumdar, A.; Ouerfelli, O.; Siminovitch, K. A.; Rosen, M. K. *Nature* **1999**, *399*, 379.

(128) Kim, A. S.; Kakalis, L. T.; Abdul-Manan, N.; Liu, G. A.; Rosen, M. K. *Nature* **2000**, *404*, 151.

(129) Tavaria, M.; Gabriele, T.; Kola, I.; Anderson, R. L. *Cell Stress Chaperones* **1996**, *1*, 23.

(130) Hartl, F. U.; Bracher, A.; Hayer-Hartl, M. *Nature* **2011**, *475*, 324.

(131) Hartl, F. U.; Hayer-Hartl, M. *Nat. Struct. Mol. Biol.* **2009**, *16*, 574.

(132) Pufall, M. A.; Graves, B. J. *Annu. Rev. Cell. Dev. Biol.* **2002**, *18*, 421.

(133) Senetar, M. A.; Foster, S. J.; McCann, R. O. *Biochemistry* **2004**, *43*, 15418.

(134) Diella, F.; Haslam, N.; Chica, C.; Budd, A.; Michael, S.; Brown, N. P.; Trave, G.; Gibson, T. J. *Front. Biosci.* **2008**, *13*, 6580.

(135) Neurath, H.; Walsh, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 3825.

(136) Stroud, R. M.; Kossiakoff, A. A.; Chambers, J. L. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 177.

(137) Khan, A. R.; James, M. N. *Protein Sci.* **1998**, *7*, 815.

(138) Rusnak, F.; Mertz, P. *Physiol. Rev.* **2000**, *80*, 1483.

(139) Dunker, A. K.; Garner, E.; Guillot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z. *Pac. Symp. Biocomput.* **1998**, *3*, 473.

(140) Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guillot, S.; Dunker, A. K. *Pac. Symp. Biocomput.* **1998**, *3*, 437.

(141) Van Roey, K.; Gibson, T. J.; Davey, N. E. *Curr. Opin. Struct. Biol.* **2012**, *22*, 378.

(142) Stommel, J. M.; Marchenko, N. D.; Jimenez, G. S.; Moll, U. M.; Hope, T. J.; Wahl, G. M. *EMBO J.* **1999**, *18*, 1660.

(143) Mitrea, D. M.; Kriwacki, R. W. *Pac. Symp. Biocomput.* **2012**, 152.

(144) Follis, A. V.; Chipuk, J. E.; Fisher, J. C.; Yun, M. K.; Grace, C. R.; Nourse, A.; Baran, K.; Ou, L.; Min, L.; White, S. W.; Green, D. R.; Kriwacki, R. W. *Nat. Chem. Biol.* **2013**, *9*, 163.

(145) Meng, W.; Swenson, L. L.; Fitzgibbon, M. J.; Hayakawa, K.; Ter Haar, E.; Behrens, A. E.; Fulghum, J. R.; Lippke, J. A. *J. Biol. Chem.* **2002**, *277*, 37401.

(146) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* **2004**, *32*, 1037.

(147) Khan, A. N.; Lewis, P. N. *J. Biol. Chem.* **2005**, *280*, 36073.

(148) Hagai, T.; Azia, A.; Toth-Petroczy, A.; Levy, Y. *J. Mol. Biol.* **2011**, *412*, 319.

(149) Zheng-Fischhofer, Q.; Biernat, J.; Mandelkow, E. M.; Illenberger, S.; Godemann, R.; Mandelkow, E. *Eur. J. Biochem.* **1998**, *252*, 542.

(150) Mandelkow, E. M.; Schweers, O.; Drewes, G.; Biernat, J.; Gustke, N.; Trinczek, B.; Mandelkow, E. *Ann. N. Y. Acad. Sci.* **1996**, *777*, 96.

(151) Butner, K. A.; Kirschner, M. W. *J. Cell Biol.* **1991**, *115*, 717.

(152) Friedhoff, P.; von Bergen, M.; Mandelkow, E. M.; Mandelkow, E. *Biochim. Biophys. Acta* **2000**, *1502*, 122.

(153) von Bergen, M.; Friedhoff, P.; Biernat, J.; Heberle, J.; Mandelkow, E. M.; Mandelkow, E. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5129.

(154) Schweers, O.; Schonbrunn-Hanebeck, E.; Marx, A.; Mandelkow, E. *J. Biol. Chem.* **1994**, *269*, 24290.

(155) Galea, C. A.; Nourse, A.; Wang, Y.; Sivakolundu, S. G.; Heller, W. T.; Kriwacki, R. W. *J. Mol. Biol.* **2008**, *376*, 827.

(156) Lacy, E. R.; Filippov, I.; Lewis, W. S.; Otieno, S.; Xiao, L.; Weiss, S.; Hengst, L.; Kriwacki, R. W. *Nat. Struct. Mol. Biol.* **2004**, *11*, 358.

(157) Nash, P.; Tang, X.; Orlicky, S.; Chen, Q.; Gertler, F. B.; Mendenhall, M. D.; Sicheri, F.; Pawson, T.; Tyers, M. *Nature* **2001**, *414*, 514.

(158) Borg, M.; Mittag, T.; Pawson, T.; Tyers, M.; Forman-Kay, J. D.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9650.

(159) Riordan, J. R.; Rommens, J. M.; Kerem, B.; Alon, N.; Rozmahel, R.; Grzelczak, Z.; Zielenski, J.; Lok, S.; Plavsic, N.; Chou, J. L.; Drumm, M. L.; Iannuzzi, M. C.; Collins, F. S.; Tsui, L.-C. *Science* **1989**, *245*, 1066.

(160) Ostedgaard, L. S.; Baldursson, O.; Vermeer, D. W.; Welsh, M. J.; Robertson, A. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5657.

(161) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5148.

(162) Vuzman, D.; Azia, A.; Levy, Y. *J. Mol. Biol.* **2010**, *396*, 674.

(163) Sivakolundu, S. G.; Bashford, D.; Kriwacki, R. W. *J. Mol. Biol.* **2005**, *353*, 1118.

(164) McCann, J. J.; Zheng, L.; Rohrbeck, D.; Felekyan, S.; Kühnemuth, R.; Sutton, R. B.; Seidel, C. A. M.; Bowen, M. E. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 15775.

(165) Kodera, N.; Yamamoto, D.; Ishikawa, R.; Ando, T. *Nature* **2010**, *468*, 72.

(166) Sicheri, F.; Moarefi, I.; Kuriyan, J. *Nature* **1997**, *385*, 602.

(167) Russo, A. A.; Jeffrey, P. D.; Patten, A. K.; Massague, J.; Pavletich, N. P. *Nature* **1996**, *382*, 325.

(168) Panchal, S. C.; Kaiser, D. A.; Torres, E.; Pollard, T. D.; Rosen, M. K. *Nat. Struct. Biol.* **2003**, *10*, 591.

(169) Dueber, J. E.; Yeh, B. J.; Chak, K.; Lim, W. A. *Science* **2003**, *301*, 1904.

(170) Bhattacharyya, R. P.; Remenyi, A.; Good, M. C.; Bashor, C. J.; Falick, A. M.; Lim, W. A. *Science* **2006**, *311*, 822.

(171) Hegyi, H.; Schad, E.; Tompa, P. *BMC Struct. Biol.* **2007**, *7*, 65.

(172) Toth-Petroczy, A.; Oldfield, C. J.; Simon, I.; Takagi, Y.; Dunker, A. K.; Uversky, V. N.; Fuxreiter, M. *PLoS Comput. Biol.* **2008**, *4*, e1000243.

(173) Minezaki, Y.; Homma, K.; Nishikawa, K. *J. Mol. Biol.* **2007**, *368*, 902.

(174) Motlagh, H. N.; Hilser, V. J. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 4134.

(175) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *J. Mol. Recognit.* **2005**, *18*, 343.

(176) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. *Genome Inform. .Ser. Workshop Genome Inform.* **2000**, *11*, 161.

(177) Vogel, C.; Bashton, M.; Kerrison, N. D.; Chothia, C.; Teichmann, S. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 208.

(178) Hegyi, H.; Buday, L.; Tompa, P. *PLoS Comput. Biol.* **2009**, *5*, e1000552.

(179) Mark, W. Y.; Liao, J. C.; Lu, Y.; Ayed, A.; Laister, R.; Szymczyna, B.; Chakrabartty, A.; Arrowsmith, C. H. *J. Mol. Biol.* **2005**, *345*, 275.

(180) Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. *Science* **2001**, *291*, 2429.

(181) Mittag, T.; Kay, L. E.; Forman-Kay, J. D. *J. Mol. Recognit.* **2010,** 23, 105.

Abbreviations

ABC   ATP-binding cassette
Ac   acidic region
AFM   atomic-force microscopy
Ar   androgen receptor
AU   analytical ultracentrifugation
BP   biological process
BR   basic region
CaM   calmodulin
CAP   catabolite activator protein
CD   circular dichroism
Cdk   cell-cycle dependent kinase
CFTR   cystic fibrosis transmembrane conductance regulator
CPD   Cdc4 phosphodegron
Cyc A   cyclin A
DH   Dbl homology
EAM   ensemble allosteric model
EGFR   epidermal growth factor receptor
EM   electron microscopy
ELM   eukaryotic linear motifs
ESR   electron spin resonance
Ets-1   E26 transformation-specific-1
FRET   fluorescence energy transfer
GBD   GTPase-binding domain
GEF   guanine nucleotide exchange factor
Hck   hemopoietic cell kinase
HECT   homologous to E6-AP carboxyl terminus
HSP-70   heat shock protein 70
IARS   intrasteric autoregulatory sequence
IDP   intrinsically disordered protein
IDR   intrinsically disordered region
KNF   Koshland-Nemethy-Filmer
MAGUK   membrane-associated guanylate kinase
MapKapK2   MapKap kinase
MD   molecular dynamics
MF   molecular function
MoRF   molecular recognition feature
MWC   Monod-Wyman-Changeux
NBD1   nucleotide binding domain
NES   nuclear export signal
NHERF1   Na+/H+ exchanger regulatory factor
NLS   nuclear localization signals
NRTK   non-receptor Tyr kinase
PRE   paramagnetic resonance enhancement
PSE   preformed structural elements
PTM   post-translational modifications
SAXS   small-angle X-ray scattering

SEC   size-exclusion chromatography
Sic1   stoichiometric inhibitor of Cdk1
SLiM   short linear motif
smFRET   single-molecule fluorescence resonance energy transfer
Smurf2   SMAD specific E3 ubiquitin protein ligase 2
SNARE   soluble NSF attachment protein receptor
SRR   serine-rich region
TetR   Tetracyclin repressor
TGF-β   transforming growth factor-β
Vav   Vav-family guanine nucleotide exchange factors
WASP   Wiskott-Aldrich syndrome protein

Author: Peter Tompa

Peter Tompa, resume

Peter Tompa (1959, Budapest) is married and has two daughters. He graduated in organic chemistry (1983, ELTE University, Budapest), and obtained his PhD in 1991 (ELTE University, Budapest) based on research carried out in the Institute of Enzymology, Hungarian Academy of Sciences (HAS), Budapest, on the interaction of glycolytic enzymes. He did his postdoc in the same institute under the supervision of Prof. Peter Friedrich on the structure-function characterization of the calcium-activated protease calpain (1992-2001), and became an independent PI in 2002. He started research on intrinsically disordered proteins (IDPs) in 2000, and played an active and decisive role in the rise of this field by suggesting a functional classification of IDPs, observing the role of structural disorder in chaperone function and developing basic concepts of the unusual modes of function of IDPs (moonlighting, fuzziness, disordered domains, supertertiary structure of proteins).

He served on the Molecular Biology Evaluation Board of the Hungarian national grant agency OTKA (2005-2008), was chair of the scientific committee Biochemistry and Molecular Biology of HAS (2008-2010), and also deputy director of the Institute of Enzymology in the period 2007-2009. Currently he is the director of VIB (Flanders Institute of Biotechnology) Department of Structural Biology, Brussels (2011-), and also a professor at VUB (Free University Brussels, 2012-). He also holds the position of professor of biochemistry and protein sciences in the Institute of Enzymology, Budapest (2006-). He has research groups both in Budapest and Brussels. He was promoter of 10 doctorates, published 121 papers, 10 book chapters, and the first monograph of the field "Structure and function of intrinsically disordered proteins" (2009) by Taylor and Francis, Inc (CRC Press). He received about 5500 citations and his H-index is 32.
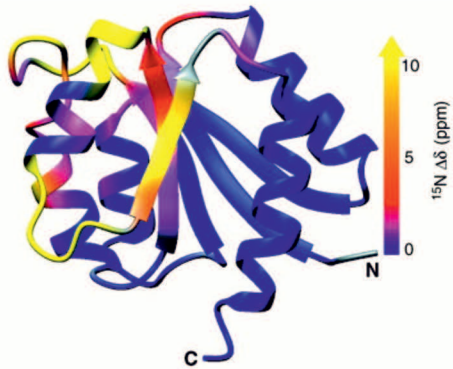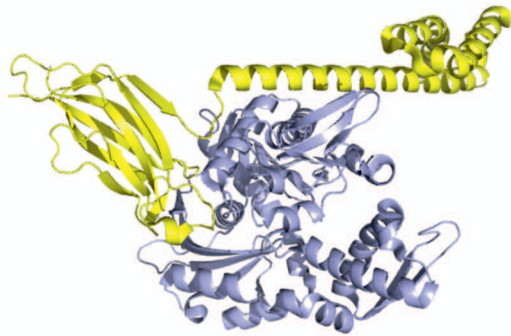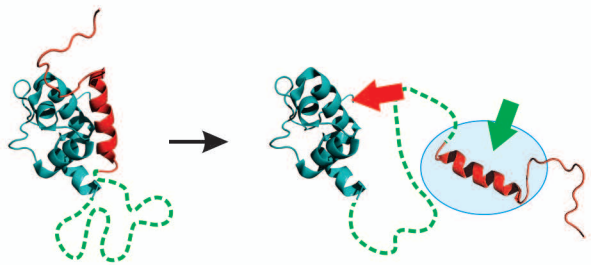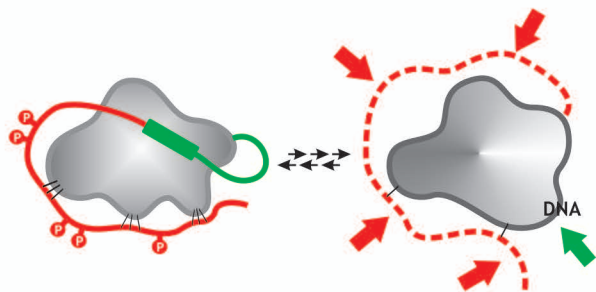
A

B

C