Chem Soc Rev

The principle of conformational signaling

Journal:	*Chemical Society Reviews*
Manuscript ID	CS-SYN-01-2016-000011.R2
Article Type:	Review Article
Date Submitted by the Author:	10-May-2016
Complete List of Authors:	Tompa, Peter; VIB, Structural Biology Research Center

The principle of conformational signaling

Peter Tompa[1,2,3]

[1]VIB Structural Biology Research Center (SBRC), Brussels, Belgium, [2]Vrije Universiteit Brussel, Brussels, Belgium, [3]Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Budapest, Hungary

email: ptompa@vub.ac.be

Abbreviations

ABC, ATP-binding cassette; AFM, Atomic force microscopy; APC, anaphase-promoting complex; APP, amyloid precursor protein; AR, androgen receptor; ATM, ataxia telangiectasia mutated; bZip, basic leucine zipper domain; CaM, camodulin; CaML, calmodulin-like domain; CAP, catabolite activator protein; CBP, CREB-binding protein; Cdk, cyclin-dependent kinase; CFTR, cystic fibrosis transmembrane conductance regulator; CPD, Cdc4 phosphodegron; CPDK, calcium-dependent protein kinase; CPEB, cytoplasmic polyadenylation element binding protein; CREB, cAMP response element binding protein; cryo-EM, cryo-electron microscopy; CSL, CBF1/suppressor of Hairless/Lag-1; DBD, DNA binding domain; DNP, dynamic nuclear polarization; EGF, epidermal growth factor; EGFR, EGF receptor; ELM, eukaryotic linear motif; EOM, Ensemble Optimization Method; EPR, electron paramagnetic resonance; ER, estrogen receptor; ESI, electrospray ionization; FGF 2, fibroblast growth factor 2; FPOP, fast photochemical oxidation of proteins; GBD, GTPase binding domain; GEF, guanine nucleotide exchange factor; GFP, green fluorescent protein; GO, gene ontology; GPCR, G-protein coupled receptor; GR, glucocorticoid receptor; HDX, H/D exchange; HEK, human embryonic kidney cell; HRE, hormone response element; HT, high-throughput; IDP, intrinsically disordered protein; IDR, intrinsically disordered region; IMC, intramolecular chaperone; KID, kinase-induced domain; KNF, Koshland-Nemethy-Filmer; LBD, ligand binding domain; LEA, late embryogenesis abundant; LiP, limited proteolysis; MAM, Mastermind; MAPK, mitogen-activated protein kinase; MAPKK, MAPK kinase; MAPKKK, MAPK kinase kinase; MD, molecular dynamics; MLCK, myosin light chain kinase; MS, mass-spectrometry; MSD, membrane-spanning domain; MWC, Monod-Wyman-Changeux; NBD, nucleotide-binding domain; NHR, nuclear hormone receptor; NICD, NOTCH intracellular domain; NTD, N-terminal domain; PDB, Protein Data Bank; PDZ, post-synaptic density-95/discs large/zonula occludens-1; PKA, protein kinase A; pKID, phosphorylated CREB KID; PLA, proximity ligation assay; PPIase, (peptidyl)prolyl isomerase; PR, progesterone receptor; Prp^{C}, cellular form of prion; PrP^{Sc}, scrapie form of prion; PCA, protein-fragment complementation assay; PTM, posttranslational modification; PUMA, p53-upregulated modulator of apoptosis; RDC, residual dipolar coupling; RyR, ryanodine receptor; SAXS, small-angle X-ray scattering; SLiM, short linear motif; smFA, single-molecule fluorescence anisotropy; smFL, single-molecule fluorescence; smFRET, single-molecule fluorescence resonance energy transfer; SNARE, soluble NSF attachment protein receptor; SOD, superoxide dismutase; SRM, selected reaction monitoring mass-spectrometry; TAD, transactivator domain; TIRFM, total internal reflection fluorescence spectroscopy; WASP, Wiskott–Aldrich syndrome protein; XFEL, X-ray free-electron laser; Y2H, yeast 2-hybrid; YFP, yellow fluorescent protein

Signal transduction is the primary device of the cell to respond to changes in its physical and chemical environment. Cellular response is initiated through a signaling protein (a receptor), which interacts with the "signal", most often a novel molecule outside or inside the cell. The mechanism of activation of the receptor is a conformational change and/or covalent modification, which then sets into motion a signaling pathway, i.e. a cascade of modification and binding events that relay and amplify the message to eventually alter the state of the cell. In reflection of this general perception, concepts such as "second messenger" and "phosphorylation cascade" dominate our views of signal transduction. The idea I advocate here is that the non-covalent change in protein conformation itself might serve as the initial or intermittent "signal" in the cascade, and it is often the primary event being recognized and interpreted by downstream receptor(s). This signaling principle is intertwined with many other cellular regulatory concepts, such as (pathway) allostery, conformational spread, induced folding/unfolding, conformational memory, hierarchical assembly of complexes, and the action of regulatory chaperones and prions. By elaborating on many examples and also recent advances in experimental methodology, I show that conformational signaling, although thus far underappreciated, is a general and robust signaling principle that most of the time operates in close interplay with covalent signals in the cell.

1. Introduction

Signaling (signal transduction) is a central principle in cellular communication and regulation, most often conceptualized as a cascade of events that starts by an extra- or intracellular signal (a first message or messenger, either a chemical entity or a physical stimulus) that activates a receptor protein through direct physical action or interaction. This initial activation often leads to the generation of a diffusible small signaling molecule (a second messenger), which then sets into motion a series of binding and modification events to instruct the cell to adapt to the change in its external or internal environment. Our general concepts of signal transduction are rooted in examples such as the synthesis or release of intracellular second messengers cAMP and Ca^{2+},[1, 2] and posttranslational modification (PTM) cascades, such as sequential phosphorylation events in the mitogen-activated protein kinase (MAPK) cascade,[3] replication start-point signaling,[4] and DNA damage pathway.[5] A signaling cascade can also include – or be built of – other types of reversible or irreversible covalent modifications, such as ubiquitination,[6] methylation,[7] or limited proteolysis,[8] and modern concepts incorporate combinations of different modifications into networks, rather than linear pathways of interconnected signaling molecules and events.[9] An overarching principle in all signaling phenomena, however, is thought to be the appearance of a new molecule inside or outside the cell, which acts as

a signal to alter the balance of transcription, metabolism, mobility, shape, excitability or some other feature of the cell. A key element of creating such a signal is that the cell spends energy to generate an "activated" species (a novel covalent chemical entity), which can then bring the system to a novel state.

Here, through multiple examples, I show that this activation principle can also operate on the protein structure as a template, when the injection of energy results not in a new molecule (a covalent entity), but the altered conformation (altered local structure) of a protein. Whereas conformational changes are central to our conceptualization of signaling as the mechanism of activation of signaling proteins, I argue here that they also conform to the definition of a signal because they represent novel chemical entities in the cell that can be recognized by appropriate receptors with specific recognition surfaces. That is, the information carried in the altered conformation (a novel spatial pattern of atoms, i.e. a binding surface) of a protein can be transduced, via molecular recognition, to downstream effects. In effect, it has always been thought that a critical element of signal transduction is alterations in the energies of protein-protein interactions, primarily by PTMs creating novel recognition surfaces. Here I suggest that the network of protein-protein interactions (the interactome) can also be remodeled by targeted (evolved) conformational changes of proteins, which can relay information[10-12] thus providing the essence of "conformational signaling". It is to be noted that the logic of a conformational signal is very close to that of a "conformational epitope" where an antibody recognizes a discontinuous segment of the antigen, brought together in space by folding of the protein structure.[13]

In principle, one can conceive an entire pathway of proteins that pass down an initial conformational signal in a "pathway allostery"[14] or "dynamic signaling pathway"[15, 16] type of fashion, in a conceptually very similar manner to pathways operating by covalent modifications. As exemplified by allostery and prions, conformational signals can be turned off, and they can synergize with PTMs, and any other signaling entity. In all, it is tempting to speculate that conformational signaling represents a simple and ancient mechanism that predated the advent of covalent signaling, which required the emergence of signaling enzymes, highly advanced catalysts. For the paucity of appropriate high-throughput (HT) and *in situ* discovery tools and approaches (but consider chapter "Novel structural biology tools for conformational signals"), however, our insight into this signaling principle is rather limited, and probably many conformational signal transduction pathways - and elements of pathways - are yet to be discovered. If this will in fact be the case, this finding will also challenge us to expand our mechanistic view of diseases, by developing novel conformational models to complement the more simplistic but prevailing genetic ones.[17]

2. Conformational changes in cellular signaling

2.1. A conformational change can constitute a signal

The concept of signal transduction emerged by recognizing that hormonal stimulation of the cell induces the appearance of a small molecule that can serve as a diffusible signal (second messenger) within the cell (Figure 1). This second messenger can modify the physiological state of the cell by way of altering the function of receptor protein(s).[2] The archetypical second messenger is cAMP, which arises from ATP by the action of adenylyl cyclase. cAMP targets enzymes (e.g. protein kinase A, PKA), transcription factors (catabolite activator protein, CAP), ion channels (HCN channel) and other types of proteins. Of similar wide-spread utility are cGMP, IP_3, Ca^{2+} and diacylglycerol, which are also diffusible and can easily reach remote protein targets. Ca^{2+}, for example, is released from intracellular stores (endoplasmic/sarcoplasmic reticulum), its level transiently increases from the resting state of about 0.1 μM to 10 μM,[18] so that it can bind a dedicated receptor, calmodulin (CaM), which can then bind and modify the activity of several hundred partner proteins.[19]

The primary stimulus (first message/messenger) can take on countless forms (Table 1), i.e., it can be a hormone (e.g. a protein (epidermal growth factor, EGF), peptide (oxytocin), or an organic molecule (testosterone)), a small metabolite or signaling molecule (e.g. glucose, NO), an environmental factor (e.g. temperature, light, pH, physical stress) or an intracellular effect (e.g. viral infection, DNA double-strand break, particular state of the cell cycle, etc…). The signal acts on an effector protein (a sensor or receptor), which becomes activated by a conformational change (or rather, the remodeling of its conformational ensemble) or covalent modification. The activated receptor protein then transduces the signal by generating a second messenger or by directly changing the enzymatic activity, cellular localization and/or binding activity of a downstream partner protein. Most often, multiple steps are linked into signaling cascades, as best exemplified by the MAPK cascade.[20] Here (cf. Scheme 1), an extracellular protein hormone (e.g. EGF), activates a transmembrane receptor (EGF receptor, EGFR), inducing the phosphorylation of its cytoplasmic domain. This way the “signal” internalizes, and is “perceived” by proteins sensitive to the presence of phosphorylated Tyr residues of the receptor. A signaling complex is assembled at the membrane, and triggers a cascade of events, in which MAPK kinase kinase (MAPKKK, e.g. RAF) phosphorylates MAPK kinase (MAPKK, e.g. MEK), which in turn phosphorylates a MAPK (e.g. ERK). This enzyme translocates to the nucleus, and phosphorylates and activates specific transcription factors, such as Elk1. Such a phosphorylation pathway can be switched off by phosphatases and it can be interlinked with many other types of modification, such as

acetylation, methylation, ubiquitination, sumoylation, which often act as the "terminal" signal in the readout of the pathway.[21, 22]

Signal transduction is one of the most central concepts of cell biology that helps rationalize the operation of the cell under physiological and pathological conditions, and thus it has been the subject of countless studies. The number of human proteins related to signaling is thousands,[33, 34] and there are hundreds of signal transduction pathways described.[35] From the underlying staggering variety of molecular details, certain unifying themes emerge (Scheme 2), as also outlined next.

1) The signal can be a physical effect, or, in most of the cases, a novel molecule (a small molecule or a (modified) protein), i.e. a novel chemical entity with a unique and distinctive pattern of atoms and chemical groups. It is to be emphasized that in this regard the primary "signal" (Table 1) and second message are not strictly different, and their distinction only reflects the experimental setup used for studying signaling. With the exception of external signals (cf. Table 1), most signals are highly intertwined and interdependent, and the primary signal for one cell is the product of signaling response of another.

2) To trigger a cellular respons, the change in concentration of the signaling molecule has to be significant (creating the false perception that it increases from 0 to a high value). For example, intracellular Ca^{2+} varies between 0.1 μM and 10 μM,[36] whereas cAMP increases from 0.15 μM to 15 μM.[37] This change is sufficient to bring the receptor of this signal to a new state. The two states of the receptor or signaling protein are usually denoted as "inactive" and "active",[38] but alternative terms, such as "relaxed" and "tense", "right" and "left", "minus" and "plus", "naïve" and "sensitized" or "signaling", "off" and "on" are also used in the literature.[39, 40]

3) The generation of a signal of any kind (a novel molecule or the PTM of a protein) takes energy, which brings the system to the activated state. This state has to be of sufficient stability to exist long enough to elicit a cellular response. One should not forget, however, that signaling might occur on different timescales, i.e. less stable signals might appropriately serve transient effects.

4) Further, with the exception of external signals, the cell does not wait for the signal to decay spontaneously, but turns it off in an active process, e.g. cAMP is decomposed by phosphodiesterase, Ca^{2+} is removed from the cytoplasm by Ca^{2+} transporters and phosphate groups are removed from proteins by phosphatases (Table 2). The signals, therefore, do not work at the temporal limit set by their chemical or physical stability.

5) Most importantly, a molecule can only signal in the cell if it has a dedicated receptor protein that is "primed" to respond to its appearance (cf. Table 2), i.e., it has a cognate binding surface, it can change activity upon recognition of the signal and it is integrated into a pathway that can elicit an appropriate cellular response.

6) A further recurring theme in signaling – as outlined in great detail throughout this review – is that the activation of a receptor or downstream signaling protein often occurs by a conformational change.

To give some detail, cAMP generated by adenylyl cyclase (Figure 1A), binds and activates PKA, which is a heterotetramer A_2C_2 of two catalytic (C) and two regulatory (A) subunits. Binding of cAMP to the regulatory subunits causes their conformational change, thereby they dissociate from the tetramer and each other, exposing active catalytic subunits.[41, 42] Kinases of the above mentioned MAPK cascade[20] are also activated through global and local conformational changes, this time elicited by covalent modifications.[43, 44] As shown for receptor Tyr kinases,[45] MAPKs,[46] ataxia telangiectasia mutated (ATM) kinase[47] and cyclin-dependent kinases (Cdks),[48] for example, they are kept inactive by the activation loop (T-loop), which covers the active site and prevents entry of the substrate, and the αC helix, which prevents nucleotide binding. Phosphorylation of the activation loop causes its conformational change to clear the way for the substrate, and also the concerted movement of αC helix to restore full nucleotide binding activity. These local changes of activation, however, are connected to more global structural reorganization, as exemplified by cyclin binding in Cdk activation,[48] or regulatory domain SH2/SH3 interactions in Src kinase activation.[49] It is appropriate to say that the covalent (post-translational) modification of the protein changes its activity and/or it can be recognized by a downstream receptor protein, altering its behavior (Figure 2A).[3, 50]

In many other cases, however, the conformational change of the protein behaves much more like a signal, because it creates a novel surface that is recognized by a downstream partner (receptor) (Figure 2B). For example, Ca^{2+} binds to CaM, which has two Ca^{2+}-binding domains connected by a flexible linker. In the Ca^{2+}-free (apo) form, the linker is in a helical conformation, whereas in the Ca^{2+}-bound (holo) state it is more flexible, enabling it to adapt to a large number of partner molecules.[57-59] In the holo state there are also hydrophobic pockets on the surface of the two globular domains which are not present in the apo state. Thus, the essence of Ca^{2+} signaling through CaM is the generation of a variety of conformational states by Ca^{2+} binding; all these serve signaling purposes,

without covalent modification.[19] CaM is thought to have some 600 partners, which recognize it through several distinct interfaces and binding modes.

Therefore, a conformational change represents an activated state of the protein that can serve as a signal in the cell. Does it satisfy all the above criteria (cf. Scheme 2)? I will provide the definite answer by the many examples discussed in the paper, here I will make three general statements. 1) By definition, an altered conformation corresponds to an altered constellation (pattern) of atoms, i.e. a novel chemical entity which did not before exist in the cell that can be recognized by a receptor protein. In several cases, the generation of a conformational signal is the primary event, which sets into motion a PTM cascade, and cascades relying entirely on conformational signaling can also be conceived. 2) The generation of a conformational signal takes activation energy, just like covalent modification of any sort. 3) A conformational signal can be of sufficient stability, because protein structure is inherently stable, and it can assume alternative local or global conformations that last long enough to serve as a signal (Scheme 2). Prions and allosteric enzymes are prime examples, and the phenomenon of protein memory also suggests the existence of stable alternative conformations of functional proteins.[60, 61]

2.2. Conformational change vs. remodeling of dynamic conformational ensembles

One has to be aware, however, that describing the conformation and conformational change(s) of a protein as a switch between stable and well-defined structural states is a definite oversimplification. Although our traditional view of protein structure is based on the notion that the native and functional state of the protein corresponds to the global minimum in the conformational free energy space,[62] a recent surge of theoretical and experimental studies have shown a complex conformational free energy landscape, arguing that protein structure is better described as an ensemble of – rather than one or a few stable – conformations. Whereas this view received its ultimate verification from the general appreciation of the phenomenon of protein disorder,[63-65] this is also true for many (strictly speaking, all) structured proteins.[66-68] As described in the chapter “Novel structural biology tools for conformational signals”, recent rapid advance in structural biology lays at our disposal exciting novel techniques that can characterize in detail the heterogeneity and dynamics of protein structures. In this scenario, a “conformational change” is better viewed as the remodeling of the conformational landscape that results in a shift in the ensemble of conformations. The importance of this description has a recognized role in protein-protein interactions[69] and has been implicated in the classical protein regulatory phenomenon, allostery.[70-73]

The central element of this view is that all proteins fluctuate between an ensemble of alternative conformations on a biologically relevant timescale.[74] For the different sub-states, distinct functions have evolved, i.e. the importance in signaling may come from that protein conformer(s) that are most complementary to particular partners, is preferentially bound. Binding stabilizes the cognate conformation, therefore it disturbs the equilibrium between different conformations and the population shifts and a new equilibrium is established. Therefore, the dynamic ensemble of conformations encodes for functional preferences in signaling.

For example, it was demonstrated in detailed structural analysis that the free ubiquitin samples distinct conformations globally similar to those of the protein in the bound state.[75] The conformational ensemble and dynamics of the free state was delineated by NMR relying on residual dipolar coupling (RDC) measurements, and it was compared to 46 X-ray structures of ubiquitin present in the Protein Data Bank (PDB) determined mostly in complex with other partners. Whereas the result can be interpreted in favor of conformational selection, rather than induced-fit, as the primary mechanism of molecular recognition by ubiquitin, it should rather be perceived as a key argument for the importance of alterations in the dynamic conformational ensemble of proteins in their partner binding and downstream signaling.

In accord, for establishing the signaling role of alternative conformations (conformational signaling[76]), the ensemble view of protein structure is more appropriate than a switch-like model, because it helps rationalize many cases when the simple two-state behavior and complementarity in selecting a binding partner is not sufficient to account for observations. As outlined later in detail, it helps explain the existence of multi-state signaling switches, such as G-protein coupled receptors (GPCRs)[77] or CaM,[19] or the seeming contradiction of specific binding of multiple ligand partners by the same protein binding site.[78] In all such examples, pre-existing multiple conformations, conformational selection and induced folding by the partner, and shifts in the ensemble, have to be invoked to explain the observed complex signaling behavior.[79] The inherent conformational heterogeneity of proteins is also instrumental in the evolution of novel functions. It has been suggested[80] that conformational states present in the structural ensemble of a protein, not yet utilized for function, represent a pool of potential functional (signaling) entities on which evolutionary selection can operate. Just like modularity (to be discussed later), this may significantly facilitate evolutionary innovation in signaling.

It must also not escape our attention that structural communication may also occur without a change in the equilibrium structure of the protein, as a result of altered dynamics that can affect the entropy

associated with partner binding. As illustrated by allostery (cf. Chapter "Allostery"),[71] this mechanism may be considered as intramolecular conformational signaling, because at the extreme, a conformational signal can be the altered dynamics of protein structure, without an appreciable change in equilibrium conformation. For example, ligand recognition by post-synaptic density-95/discs large/zonula occludens-1 (PDZ) domains has been studied to understand allosteric communication between the ligand-binding site and distal functional regions of the domain.[81] By NMR measurements, long-range effects on side-chain methyl dynamics were detected that correspond to previously observed pair-wise allosteric energetic couplings. It was concluded that ps-ns timescale dynamic fluctuations, rather than definite structural changes, contribute to allosteric signal transduction. This behavior has been observed in many other cases.[71, 82]

In the context of our concept, all these results imply that conformational signaling should not be strictly interpreted in terms of the traditional lock-and-key mechanism, i.e. by assuming that a single and well defined conformational signals develop upon signaling that then tightly fit to the binding pockets of receptor proteins.[83] Rather, we may speak about the remodeling (shift) in the population of conformations of a given protein, which alters preferences for signaling binding partners, i.e., the probability of binding a particular partner or selection between different partners. Conformational selection, induced fit, and population shifts all play then a role in conveying the message to the rest of the cell.

2.3. Allostery

As follows from the foregoing considerations, allostery is a prime example of the transduction of information represented in a protein conformation, also illustrating the interplay of covalent and conformational signals.[84] In the traditional mechanism of allostery, it is assumed that the allosteric signal (a ligand or a PTM) induces a conformational change in one subunit of an oligomeric protein, which then propagates into adjacent subunit(s).[85] The classical case is hemoglobin, a homotetrameric oxygen-binding protein, which shows positive cooperativity in oxygen binding: binding of the first oxygen molecule to one subunit changes its conformation, influencing the other subunits so that they bind oxygen more tightly,[86, 87] i.e., the conformational change spreads from one subunit to the others (Figure 3A). Whether the mechanism is induction of the conformational change (concerted Monod-Wyman-Changeux (MWC) model[86]) or selection from a preexisting ensemble of conformations (sequential Koshland-Nemethy-Filmer (KNF) model,[87]) the conformational change spreads between subunits like a signal. There are countless varieties on this theme. For example, cAMP-activated CAP[88] is a homodimeric transcription factor responsible for catabolite repression.[89]

When the amount of glucose is low, the level of cAMP increases, it binds to CAP and elicits a coil-to-helix transition that causes a rotation and translation of the DNA-binding helical regions (Figure 3B). Subsequent DNA binding at the *lac* operon induces three genes involved in lactose transport and metabolism.

The underlying classical notion of the spread of conformational information in oligomeric proteins as the basic mechanism of allosteric activation has been refined but not basically altered by later extensions of the concept, such as allostery within single-subunit proteins, and/or its operation by a change in dynamics without appreciable change in equilibrium structure.[71, 82]

If such allosteric signaling occurs between different subunits of heterologous complexes, one might speak about signal transduction via the propagation of a conformational signal from one protein to another. If diffusible proteins involved in such a mechanism interact with each other transiently, the cascade of events may actually be considered as a pathway constituted of conformational signals, as captured recently in the "pathway allostery"[14] and "dynamic signaling pathway"[15] models.

This actually is the essence of "conformational spread", a phenomenon described in the case of bacterial chemotaxis receptors. These receptors mediate the detection of attractants and repellants in bacteria, and through histidine kinase CheA-mediated phosphorylation of response-regulator CheY, they regulate rotor rotation and bacterial move.[90] Quantitative modeling of the short pathway revealed that the actual receptor occupancy by attractants would elicit CheY phosphorylation that would be insufficient to account for the observed frequency of the change in the direction of motor rotation. The contradiction can be resolved by assuming a crosstalk between receptors that results in "front end" amplification,[91] meaning that the conformational change of one receptor upon ligand binding triggers similar conformational changes in the neighboring receptors. Close packing of receptors in a cluster ensures the efficient spread of conformational information resulting in a huge degree of signal amplification.[92, 93] Similar cooperative conformational signal propagation and amplification occurs between other channels and receptors, such as transmitter-gated cation channels[94] and ryanodine receptors (RyRs).[95] In the latter, coupled gating in the closely packed array of receptors is instrumental in massive calcium release in muscle contraction. Conformational cross-talk in a lattice of receptors was also suggested in T-cell receptor signaling, where negative and positive conformational crosstalk among neighboring receptors appears to be instrumental in enhancing T-cell specificity.[96] Conformational spread has also been described in the C-ring of 34 copies of FliM proteins in regulating the rotation of bacterial flagella.[97]

More linear and signaling-like relations and signal amplification has been suggested in other cases, such as in actin-myosin coupling, where conformational changes of individual myosin motors is coupled through conformational strain, which ensures that the conformational change of may myosin molecules associated with actin change almost simultaneously in a process termed "mechanochemical coupling".[98]

2.4. Transitions between folded and disordered states

As outlined above, the "ensemble view" of conformational transitions has already had a basic impact on interpreting allosteric phenomena and it also received a big impetus by the recent breakthrough of recognizing that many proteins (intrinsically disordered proteins/regions, IDPs/IDRs) are devoid of a stable structure under native, functional conditions.[64, 65, 99] Structural disorder, which can be most appropriately described as an ensemble of conformations,[67, 68, 100] is particularly abundant in proteins of signaling and regulatory function,[101, 102] where it is often involved in protein-protein interactions. In these, regions of IDPs/IDRs undergo induced folding[103] or induced unfolding,[104, 105] which represent strong conformational signaling principles.

2.4.1. Induced folding

When binding occurs by induced folding, experimental evidence often points to the pre-formation of the bound conformation (or conformational element) in the unbound ensemble.[106, 107] It follows that if a nearby signaling input (binding of a partner or a PTM) shifts the ensemble towards - or away from - the binding competent conformation, it will have a profound effect on binding of the partner. In either direction, the energy provided by the PTM or partner binding brings the signaling protein into an "active" or "activated" state, with a functional signaling readout different from the "inactive" state.

This feature has been characterized in detail for the transcription factor cAMP response element-binding protein (CREB). The kinase-induced domain (KID) of CREB is an IDR with two short segments, helix A and helix B, sampling local helical conformations.[108] Phosphorylation of Ser133 in between the two regions increases the helicity of helix B, pushing the ensemble toward the state competent for binding to the KIX domain of the transcription coactivator CREB-binding protein (CBP). The action of basic leucine zipper domain (bZip) transcription factors (such as Fos, Jun, CREB, ATF), can also be rationalized along similar lines. These transcription factors are fully disordered in the monomeric state, and can only bind DNA as homo- or heterodimers. Dimerization occurs by the formation of an extended coiled-coil interface. Helical propensity already exists in the monomeric state,[109] but it is

not sufficient to engage in a functional interaction with DNA. The DNA-binding interface only fully forms upon folding induced by dimerization, i.e., the conformational signal for the assembly of the transcription complex is generated by binding of the partner (Figure 4A).

A different PTM-induced conformational change occurs in proline-containing sequences of proteins, where their cis-trans isomerization creates structural isomers, which can have different interaction propensity with partners.[110] The isomerization, which is somewhere in between a covalent and a conformational change, can be catalyzed by specific enzymes (peptidyl)prolyl isomerases (PPIase, such as Pin1), which may act in a highly regulated phosphorylation-dependent manner in creating a conformational switch.[111]

Such signaling conformational changes upon intra- or intermolecular protein-protein interactions also enable the emergence of complex signal-integrating switches, as exemplified by nuclear hormone receptors (NHRs). NHRs constitute a family of 48 transcription factors in humans, mediating cellular response to a variety of vitamins and steroid hormones, as studied in detail for estrogen receptor (ER), glucocorticoid receptor (GR), androgen receptor (AR) and progesterone receptor (PR),[112] for example. All NHRs contain an intrinsically disordered N-terminal domain (NTD), a conserved DNA binding domain (DBD) and a ligand binding domain (LBD). Upon ligand binding, the activated receptor translocates to the nucleus, binds to a specific DNA sequence (hormone response element, HRE), and activates transcription through interaction with a plethora of co-activator proteins and general transcription factors. The interactions are accompanied by the induced folding of NTD to a more helical state(s).[113-115]

The formation of the exact binding-competent NTD structure is regulated by external signals, which may generate multiple structural states, and thus regulate the choice between partners.[113, 116] Due to interdomain allosteric communication, specific binding-competent local helical structures may arise in the NTD upon interaction of the adjacent DBD with a specific HRE sequence, as seen with GR,[116, 117] and/or with a co-activator protein (e.g. JDP2), as seen with PR.[118] Allosteric communication between distinct co-activator partners (e.g. TFIIF and SRC-1) might also be mediated by local conformational signals, as demonstrated for AR.[119] Post-translational modifications, such as phosphorylation of ER[120] and GR,[121] may also promote binding-competent local conformations in the NTD. In all, due to the possible synergy between all the different allosteric signals, conformational signaling by NTDs may underlie the complex signal integration capacity of NHRs.

2.4.2. Induced unfolding

Induced unfolding, when a binding region becomes activated by its local unfolding into a more disordered and accessible state, is also frequently encountered in signaling. For example, Wiskott–Aldrich syndrome protein (WASP), which regulates cytoskeletal actin polymerization, has a closed autoinhibited state, held by an intramolecular interaction between its N-terminal GTPase binding domain (GBD) and C-terminal VCA region.[122] WASP is activated by an elevated level of the Rho-family GTPase Cdc42, which binds to GBD and induces the exposure of the VCA region. VCA then recruits the Arp2/3 complex and promotes actin polymerization; i.e., the essence of activation is the increased accessibility of the VCA motif due to the unfolding of the protein (Figure 4B).

Signaling by induced unfolding also occurs in the case of p53-upregulated modulator of apoptosis (PUMA) protein, a BH3-only protein that is induced upon DNA damage by p53. PUMA is part of a large family of BCL-2-related pro- and anti-apoptotic proteins,[123, 124] it binds and inhibits anti-apoptotic BCL-2 proteins, such as BCL-xL,[125] disrupting the interaction between cytosolic p53 and BCL-xL. This allows p53 to interact with BCL-2 effector molecules BAX and BAK in promoting apoptosis. When PUMA binds to BCL-xL, it induces a partial unfolding of two of its α-helices, which disrupts its interaction with cytosolic p53, i.e., the disordered state of the helices of BCL-xL acts as a signal that instructs p53 to serve in the cytosol as an apoptotic activator. It might be argued that the signal is the unfolded state, i.e. the folded state represses the signal.

A conceptually similar, but structurally much less understood case of induced unfolding explicitly termed “conformational signaling” occurs in integrin receptors.[76] Integrins are dimeric transmembrane receptors mediating adhesion of cells to the extracellular matrix. They contain two non-covalently associated α and β subunits, with highly non-symmetric structures. Their “bent” conformation represents a physiological low-affinity state, whereas ligand binding induces large conformational changes (extension) in the extracellular region, which moves the transmembrane regions apart and separates the two short cytoplasmic domains. This brings about a shift in the conformational ensemble of the cytoplasmic helices, and also makes the region more accessible for signaling partners talin and filamin by changing their structural population and membrane proximity, which together constitute the “signal”.[126]

There are many other examples of signaling by regulated unfolding,[104, 105] as also described in proteins subject to autoinhibition,[127] or active-site directed regulation.[128] Due to the nature of the structural transition in these proteins that results in the exposure of a potential signaling region, it is

also termed “cryptic” and “conditional” disorder.[104, 105] Together, induced folding and induced unfolding can be accommodated into complex signal integrating proteins, in which conformational rearrangements mediate and aggravate communication between remote regions.[72, 73]

2.5. Hierarchic assembly of complexes

The hierarchic assembly of complexes[129] may also be interpreted in terms of the operation of conformational signals. Frequently, protein-protein or protein-RNA complexes are assembled in a non-random order in evolutionarily conserved assembly pathways.[130-132] The order of binding might depend on the free energy of successive interactions, and/or on the appearance of novel binding surfaces, which are constituted by more than one subunit and/or are generated by structural transitions of subunits during the assembly process (Scheme 3). A novel surface might be considered as a conformational signal that is recognized by the next subunit incorporated into the complex. For example, as suggested in the previous chapter for bZip transcription factors, their dimerization and ensuing induced folding creates a novel recognition interface with DNA, which is instrumental in the subsequent assembly of a transcription preinitiation complex (Figure 4A). The assembly of the transcriptionally active complex composed of NOTCH intracellular domain (NICD), Mastermind (MAM) and CBF1/suppressor of Hairless/Lag-1 (CSL) transcription factor in NOTCH signaling[133] also follows a similar logic. Here, the addition of MAM can only occur after the association of NICD with CSL, which creates a novel binding groove for an elongated and kinked helical region of MAM.

The complex regulation of induced folding in NHRs, discussed in the previous chapter, is also instrumental in the hierarchic assembly of the transcription pre-initiation complex, which contains the NHR transcription factor, DNA, co-activators, members of the general transcription machinery and RNA polymerase II holoenzyme.[114]

This signaling principle has also been articulated in the assembly of ribonucleoprotein complexes. Induced fit of the components is prevalent in protein-RNA recognition:[134] it has been found that either the RNA molecule (e.g. ribosomal S15 – rRNA interaction[135]) or the protein (e.g. ribosomal L11 – rRNA interaction[136]) or both (e.g. U1A – mRNA 3’ UTR interaction[137] or ribosomal protein L5 – 5S rRNA interaction[138]) undergo induced fit (or in the latter case, mutual induced fit) upon complex formation. The generality of this phenomenon is perfectly in line with the observed very high incidence of structural disorder in RNA-binding proteins,[139] such as ribosomal proteins[140, 141] or splicing factors.[142] It has been explicitly suggested that the major role of this (mutual) induced folding in ribonucleprotein complexes is to promote a “function” that is not manifest in the isolated

components.[134] The "signal" in this case is the novel binding surface which ensures that there is no inappropriate signaling, i.e. binding of new subunit(s), without complex formation (cf. Scheme 3).

The important role of induced folding in creating novel interaction surfaces in the hierarchic assembly of complexes might infer that the larger the protein complex (the more subunits it has), the more prevalent this mechanism is. In accord, bioinformatics predictions have shown that there is a statistically significant correlation between structural disorder and the number of proteins assembled into complexes,[143] as has been analyzed in detail for the transcription-regulatory Mediator complex.[144]

3. Conformational signaling: a general theme in signal transduction

3.1. Signaling pathways

A key feature of cellular signal transduction is that signaling entities (messengers and proteins) are linked up into pathways, with significant amplification along the way. The switch-like behavior of elements endows the pathway with complex signal-integration capacity. To cite traditional signaling, a growth-factor signaling cascade can typically involve 9 (e.g. EGF, EGFR, Grb2, SOS, Ras, Raf, MEK, ERK, Elk1 (Scheme 1)[3]), a caspase cascade 7 (Death Factor, FasR, FADD, Casp8, Casp3, Casp6, Lamin),[8] and the Wnt signaling pathway 7 (Wnt, Frizzled, Disheveled, GSK-3β APC, β-catenin, TCF)[145] proteins. This multiplicity of signaling steps is probably important for providing spatial distance coverage (e.g. from the plasma membrane to the nucleus), amplification of the signal (an increase of insulin from 50 to 250 pM, causing a decrease in blood sugar level from 10 to 5 mM,[18]) and also the introduction of additional control into reaching a regulatory decision.

Do we see entire signaling cascades, relying purely on the conformational principle, of comparable length? Well, such conformational signaling pathways have not yet been described but this mechanism often represents elements or segments of well-established pathways. For example, several conformational steps are connected in GPCR signaling. GPCRs are a diverse family of seven-transmembrane domain receptors, which act through heterotrimeric GTP-binding proteins (G proteins) in mediating the majority of cellular responses to hormones and neurotransmitters as well as the senses of light, smell and taste.[146, 147] They also attract much attention because they (alongside kinases) are the primary targets in drug development efforts.[148] Binding of the ligand to the receptor causes a conformational change in the GPCR (such as adrenaline binding to β2 adrenergic receptor (Figure 5)), which thus recruits a heterotrimeric Gs protein (Gsαβγ) and acts as a guanine nucleotide

exchange factor (GEF) for its alpha subunit (Gα). Gα undergoes a major conformational change, exchanges its bound GDP for GTP,[149, 150] and dissociates from the βγ heterodimer. Both Gα and Gβγ have signaling roles through interactions with downstream signaling proteins. For example, Gsα binds to adenylyl cyclase, a 12 transmembrane-helix protein with a cytoplasmic catalytic domain. Gsα binds to this domain, which then undergoes a major conformational transition[151] that activates it to produce cAMP that induces smooth muscle relaxation.

Interestingly, the conformational change – and the ensuing signaling mechanism – of Gsα seem to be evolutionarily conserved, because the α subunit of heterotrimeric G proteins is homologous to small GTPases, monomeric proteins homologous to (and sometimes named after) Ras. These proteins are divided into subfamilies such as Ras, Rho, Rab, Arf and Ran, and they are regulatory switches in several cellular phenomena, such as growth, differentiation, motility and vesicle transport.[40, 152] Their basic mechanism of activation is also an exchange of GDP to GTP, which brings about a characteristic conformational change that activates them; for example Ras activation enables its binding to and activation of the MAPKKK Raf.[40] Actually, the MAPK cascade in which Ras is embedded in represents another signaling pathway which transduces information by the extended use of conformational signals (Scheme 1). Whereas the pathway is classically referred to as a cascade of phosphorylation events of successive kinases, several of its steps (such as that of the EGFR receptor, SOS, RAS, RAF kinase and the transcription factor Elk1), undergo characteristic conformational changes that signal to the next member of the pathway.

3.2. Conformational signals as decision-making switches

The GPCR pathway shows how a chemical signal (adrenaline) can be transduced to a series of conformational signals (that of the receptor, Gsα and eventually adenylyl cyclase) before switching back to a chemical signal (cAMP). Actually, it also demonstrates that the plasticity of a protein structure might enable switches of more than two states that may signal to multiple downstream pathways, or have graded responses (cf. chapter ”Conformational change vs. remodeling of dynamic conformational ensembles”). The highly dynamic structure of a GPCR can modulate the activity of more than one second messenger system in a ligand-specific manner,[153] which suggests that its structure can assume multiple distinct conformations depending on the bound ligand and associated signaling proteins (Figure 5). The presence of multiple distinct conformational states can be demonstrated by biophysical techniques (e.g. NMR), but solving them separately by X-ray crystallography presents a significant challenge,[77, 154] as demonstrated by β2-adrenergic receptor,

which activates more than one G proteins and signals through at least one G-protein independent pathway, via arrestin (Figure 5).[77]

Signaling to multiple downstream targets also appears in the case of ephrin signaling, where the conformational signal has the capacity to support signaling decisions and branching.[14, 155] There is a family of 10 EphA and 6 EphB receptors controlling signal transduction between cells, by interacting with 6 ephrinA and 3 ephrinB ligands, maximizing the number of Eph-ephrin interactions. The transient binding of different ephrin ligands to receptors results in a slightly different preferred conformation or dynamics of the second site of the receptor, which selects for different partners, initiating different pathways with different functional outcomes. Conformational selection and population shift events presage dynamic conformational changes during Eph-ephrin recognition,[155] i.e. specificity of particular pathways, to a significant extent, is controlled by population shifts of the receptor and ligand.

As described in the chapter "Transitions between folded and disordered states", multi-state switches have also been described in the case of NHRs, the NTD of which can assume distinct signaling conformations upon the combined action of diverse signals, such as PTMs, the binding of HRE DNA sequences and co-activator partners.[113, 116] The multiple modes of CaM signaling to a very large number of partners[19] also demonstrate this feature of protein structure. Prions also display switch-like behavior, because the same prion molecule can acquire numerous different transmissible prion states (termed prion strains[156]), corresponding to different conformations.

3.3. Signal amplification

Signal amplification is also a critical feature of signal transduction that can bridge a difference of orders of magnitude between the concentrations of the signal and output, enabling a massive cellular response to a small but critical change in environmental conditions.[157] For example, the level of insulin upon glucose challenge increases from 50 to about 250 pM, causing an eventual reduction in blood sugar level from 10 to 5 mM.[18] Such amplification can be conceived if one signaling event generates multiple signals at the next level, as apparent in the case of enzymes (generating multiple molecules of a second messenger or modifying multiple substrate proteins), channels (releasing multiple Ca^{2+} ions) and transcription factors (causing many rounds of transcription augmented by multiple rounds of translation). As clearly shown by the above examples, a conformational signal interlaced by a chemical signal can also result in such amplification. For example, the assembly of a bZip transcription factor dimer causes many rounds of transcription, or the association of activated

Gsα with adenylyl cyclase causes the generation of many cAMP molecules.[158] If the interaction mediated by the conformational signal is transient, it is conceivable that even the primary conformational signal can modify the action of multiple receptors. This is the case of classical allostery (e.g., binding of one oxygen molecule to hemoglobin sensitizes three further subunits), and to a larger degree in the case of conformational spread, as discussed above. In the GPCR-G-protein conformational activation switch, activated GPCR can also activate multiple G proteins,[147, 158] i.e. its conformational signal can be amplified in a manner similar to enzymes catalyzing multiple rounds of chemical modification. The conformational signal in the activation of RAS in the MAPK cascade (Scheme 1) also entails a significant extent of amplification, because its association with, and activation of, RAF kinase brings about many rounds of phosphorylation of the downstream kinase MEK. In fact, oncogenic mutations of RAS tend keep the protein in an activated conformational state, which causes oncogenic transformation due to the tremendous amplification of the embedded conformational signal through the pathway.[159]

Prions (see Chapter "Prions as pathological and physiological signaling molecules") are probably the best examples of signal amplification, because they can transduce the initial signal (which may be the altered conformation arising spontaneously, or due to a mutation or environmental effect[160] of a few molecules sufficient to assemble into a "seed") across many proteins, eventually converting the entire pool of molecules to the prion state.

Closely related to signal amplification is that many signaling pathways display oscillatory behavior,[157, 161] which seems entirely compatible with the action of conformational signals. The most critical elements of oscillating pathways are sensitivity (signal amplification) and regulatory feedback loops, with remote inhibitory and activatory relations, which can be readily constructed from conformational signaling elements. Of further relevance, (conformational) transitions of protein switches can be very fast (down to the μs timescale), and result in stable alternative states (as exemplified by protein memory and alternative allosteric states). In accord, it might not be fully by chance that two pathways repeatedly mentioned here for conformational signaling elements (that is, the NOTCH pathway[161] and the MAPK pathway[157]) are also prime examples of oscillatory behavior.

3.4. Thermodynamics of conformational signaling

The existence of conformational signals and the comparison of conformational and covalent signaling raise two further important questions, the thermodynamic stability of signals and their regulated generation. A lot remains to be discovered here because our insight into both is rather limited.

The generation of stable alternative conformational states is compatible with two possible thermodynamic scenarios (Figure 6), both of which seems operational in different cases. In one, the active signaling conformation is more stable than the inactive one, but it is kinetically inaccessible due to the high energy barrier separating the two (Figure 6). To generate the signal, the energy barrier has to be reduced, which accelerates the transition to the signaling state. This can be achieved by a PTM, an interacting partner or the action of a chaperone-like protein, which raises the possibility of the existence of a special class of signaling chaperones. Actually, such a function has been suggested for intramolecular chaperone (IMC) sequences of subtilisin, which can drive the formation of alternative active conformations of the enzyme.[61] Also in prion propagation, one model (template assistance[162]) assumes that the prion state is inherently more stable than its cellular counterpart, with their interconversion being extremely slow due to a high energy barrier separating them. Interaction with a scrapie seed lowers the barrier and accelerates conformational conversion, i.e. the prion acts as a conformational catalyst.

In the other scenario, we may assume that the active state is less stable than the inactive one, and the conformational energy landscape of the protein has to be remodeled to make this state favorable (Figure 6). This may be achieved by binding of a small signaling molecule or by PTM (as often observed in allostery), or the action of a protein binding partner (as occurs in the hierarchic assembly of complexes). At the extreme, even a different type of signaling chaperone action is also feasible here, which folds a protein partner into a less stable, but kinetically trapped conformational state (Figure 6). Such as “steric chaperone” effect has been described for the intermolecular action of a lipase-specific foldase and the pilus subunit specific chaperone.[163, 164]

In all, our view on proteins supports the notion that alternative protein structures can be stable enough to support signaling roles. Allostery is an eminent case for this,[85] and prions also demonstrate stable alternative conformational states.[160] The phenomenon of protein memory adds a further dimension to this view.[60, 61]

3.5. Modularity in conformational signaling

3.5.1. Modules in covalent signaling

The general utility of signal transduction is strongly supported by its modularity, i.e. that the same system, or its elements, can be used in a different biological context, effectively moving both horizontally (within the same cell or organism) and vertically (between species) in evolution. The high degree of modularity makes signaling systems (proteins, pathways, networks) extremely evolvable and adaptable, ensuring the fast emergence of highly innovative regulatory solutions by simple genetic events such as recombination, deletion, or insertion.[165, 166] Certain signaling molecules are used in dozens, if not hundreds, of pathways. For example, calcium signals to hundreds of partners,[19] cAMP has at least a dozen targets,[2, 167] and there are about 500 homologous protein kinases encoded in the human genome[168, 169] incorporated into countless pathways (Figure 7A). This attests that evolution by gene duplication followed by sequence divergence and functional specialization[170] is incomparably faster than by the *de novo* generation of novel proteins in the creation of novel signaling solutions.

Modularity also applies within proteins, where internal duplications and shuffling of domains promotes functional innovation (Figure 7B). Domains are autonomous folding units[171, 172] often used in different genomic context, i.e. they constitute evolutionary modules.[173, 174] They can mediate distinct functions, such as catalysis (e.g. kinase, phosphatase), or interaction, with peptide motifs (e.g. phospho-Tyr motif binding domains), other domains (e.g. interaction between SAM or CARD domains), RNA/DNA (e.g. helix-turn-helix, RRM, KH domains) or membranes/phospholipids (e.g. PH and C2 domains).[175-177] Shuffling of domains into novel combinations is the primary mechanism of generating complex signaling proteins.[165, 166]

For example, a recurring theme in the assembly of signaling complexes is the binding of Pro-rich segments by SH3 domains, binding of regions of phosphorylated Tyr residues by SH2 domains, and phosphorylation of Tyr residues by Tyr-kinase domains (Figure 7B). The extreme success of these domains in promoting diversity in signal transduction as modules is illustrated by their use in many different contexts: SH3 domains appear in 300 different human proteins,[178] SH2 domains have 120 copies in the genome,[179] and kinase domains appear in more than 500 different human proteins (the kinome).[169] The effective modular use of signaling domains is also apparent in their frequent appearance in tandem repeats within the same protein, that can go as high as dozens of domains (such as zinc fingers, Ig domains, EGF domains, etc…) or supersecondary structural elements making up signaling domains (such as leucine-rich repetas, armadillo repetas, ankyrin repeats, TIM barrel, porin, β-trefoil structure, etc…).[180-182]

Signaling domains also often appear in characteristic combinations termed supradomains, when their combination results in a module that is used repeatedly in different proteins.[181, 183] Particularly successful such domain combinations are the signal transducing histidine kinase domain and dimerisation ATPase domain,[184] the GroES-like domain and NAD(P)-binding Rossmann-fold domain[185] and P-loop containing nucleotide triphosphate hydrolase domain and translation proteins domain;[186, 187] such combinations of domains usually carry out a particular function that distinguishes them from the isolated domains.

Modularity also applies for shorter elements of signaling proteins, termed motifs. As suggested, domains most often function by recognizing short linear segments (peptide motifs) of their signaling partner, which often fall into intrinsically disordered regions of proteins.[188] They are typically 3–8 amino acids in length, and the majority of their interaction specificity resides in only 3–4 core positions.[189, 190] Peptide motifs are most often termed short linear motifs (SLiMs) or eukaryotic linear motifs (ELMs).[190] There are about 200 different motif classes (types) with more than 2400 experimentally validated instances,[191, 192] but indirect evidence posits that their number of instances can be as high as a million in the human proteome.[193] Because the binding surface of motifs is limited, their binding usually results in low-affinity, transient, and modulatable interactions, which is of significant benefit in signaling. They belong to two broad and not perfectly separable functional classes, those that are recognized by post-translational modification enzymes (modification sites), and those that mediate more stable interactions with a domain (ligands motifs).

The same type of motif can be used in many different contexts, i.e. motifs are also evolutionary modules, but due to their short length and limited information content, their evolutionary history is much less certain than that of domains. In certain (longer) cases, such as BH3-motifs in BH3-only BCL-2 family pro-apoptotic proteins[194] or K-segments in class 2 (dehydrins) of late embryogenesis abundant (LEA) plant stress proteins,[195] there is some evidence that they might be homologous, i.e. they have descended in evolution by divergence. Most of the time, however, motifs, such as the Pro-Ser segment of cyclin-dependent kinases or the Pro-XX-Pro binding segment of SH3-containing signaling proteins, are short and their different instances are more likely to have arisen by random mutations, i.e. by evolutionary convergence.[193] For their limited information content, the presence of motifs in the same proteins is harder to assess, but functional evidence suggests that they can also be used in tandem arrays and/or combinations in the same protein, for example, Tyr kinase receptors (e.g. EGFR) contains 5 pTyr motifs,[196] and the Sic1 Cdk inhibitor contains 9 copies of Cdc4 phosphodegrons (CPDs).[197]

3.5.2. Evolution of conformational signaling by module exchange

At the moment, it is significantly harder to answer if conformational signaling elements of proteins are also used as modules in evolution, although it should not escape our attention that several modules outlined in the previous chapter recognize conformational, rather than covalent, signals. On the side of the signals, as we think of them as alternative conformations of a native protein, it can be conceived that the protein segment (domain and motif) capable of undergoing regulated conformational change can be used in different contexts.

A notable example is the calcium signaling protein CaM (Figure 8A). As suggested, it can bind almost 600 different partners in several different conformations, binding modes,[19] to regulate very different functions in the cell. CaM, however, does not only occur as a stand-alone protein, it can also be the regulatory element of multidomain signaling proteins, in which it acts as a calcium-sensor domain, relaying regulatory information to other domains of the protein (Figure 8A). A prime example of the modular use of CaM as a conformational signaling element is provided by calpains, the intracellular calcium-dependent cysteine proteases.[198] Calpains play a variety of roles in the cell, such as remodeling of actin cytoskeleton in cell motility,[199] and have also been implicated in disease, such as limb-girdle muscular dystrophy type 2A.[200] Traditional calpains have two subunits, an 80 kDa catalytic subunit and a 30 kDa regulatory subunit both having a CaM-like domain (CaML), which are the primary calcium-sensors of the protein. In the absence of calcium, the catalytic domain of the protein is in an inactive, distorted conformation.[201] Calcium binding causes a conformational change in the two CaMLs that propagates to the papain-like domain, which assumes a catalytically competent, active conformation. That is, CaMLs of calpain should be considered as signaling modules based on the conformational principle, recruited for this use by evolutionary domain shuffling and divergence. Similar CaM-like regulatory domains also occur in plant calcium-dependent protein kinases (CDPKs).[202]

As discussed in detail in chapter “Prions as pathological and physiological signaling molecules”, self-propagating structural switches, prions also operate on the conformational signaling principle. Prions have been described originally as infectious pathogenic entities, but recent results provide strong evidence that many prions (such as Sup35 and Ure2 in yeast[204] and CPEB in *Aplysia californica*[205]) have physiological functions (Table 3). Intriguingly, the characteristic conformational transition of these proteins rests on the presence of a domain of unique amino acid composition, often enriched locally in Gln and Asn residues (Q/N rich domains)[206, 207]). These “prion domains” carry the capacity to undergo autocatalytic amyloid formation on their own,[208-210] and they are sufficient determinants of

the transition to the prion state.[206, 207] The principle of modularity is strongly supported by the transferability of the prion domain, enabling to turn reporter proteins to prions by attaching prion domains to them (creating "synthetic" prions[207, 211]). For example, when the prion domain of Sup35p was fused to the glucocorticoid receptor (GR) in cells in which β-galactosidase synthesis is regulated by the glucocorticoid receptor, this protein chimera behaved like a prion, practically eliminating GR activity[211] (Figure 8B). This modularity also enabled to predict novel prions in animal proteins based on their distinctive compositional and sequence features.[212]

As suggested, it is difficult to reconstitute the evolutionary history of ligand motifs, and to ascertain that two similar (or identical) motifs are descendants of the same ancestral entity. As suggested above through the examples of repeated pTyr motifs in Tyr kinase receptors and Cdc4 phophosdegrons in Sic1, however, internal duplications may indicate that this is in fact the case, i.e., we may be assured that multiple copies represent evolutionary divergence. Based on this mechanism, the regulatory (R) region of the cystic fibrosis transmembrane conductance regulator (CFTR) protein is a prime example of the modularity of a motif operating by the conformational signaling principle.[213] This channel belongs to the ATP-binding cassette (ABC) superfamily of proteins.[214, 215] Its mutation in cystic fibrosis causes a loss of activity due to degradation in the endoplasmic reticulum secretion pathway. CFTR has two membrane-spanning domains (MSD1 and MSD2), two nucleotide-binding domains (NBD1 and NBD2), and a region between the transmembrane segments that harbors an intrinsically disordered regulatory (R) domain of about 200 amino acids in length. R domain interacts with NBD1 predominantly via multiple short motifs assuming transient helical conformations,[213, 216] which are regulated by PKA phosphorylation. Without phosphorylation, these sites tend to have fractional helical propensity necessary for mediating interaction with NBDs. Phosphorylation of the sites by PKA shifts their conformational state toward lower helicity, which reduces their NBD1 interactions, i.e. the conformational signal in this case is a more random structural state.[213, 217] Due to internal repetition, we can be positive that the conformational signal represented by this short motif is a module reused in evolution.

4. Conformational missignaling

4.1. Deregulation of conformational signaling can cause disease

Because signal transduction is central for the cell to respond to external or internal stimuli, it is of no surprise that defects of signaling are often conducive to human diseases,[101, 218-221] due to which signaling proteins are the most prevalent drug targets.[148, 221] Many diseases are caused by inherited

or acquired modifications in protein sequence that may cause the deregulation of function by interfering with the synthesis or stability (intracellular level), transport (localization), and/or activity (as an enzyme, receptor, transcription factor, etc…) of the protein. Irrespective of the actual mechanism that upsets the operation of these signaling proteins, we may call the resulting imbalance of signaling "missignaling", involved in a wide array of disorders from cancer through inflammation to several types of neurodegenerative diseases. To give a few examples, some of the best studied heritable (familial) diseases are caused by mutations in key signaling proteins, such as p53[222, 223] and Ras[224] in cancer, EGFR in inflammation,[225] amyloid precursor protein (APP) in Alzheimer's disease[226, 227] and superoxide dismutase 1 (SOD1) in familial amyotrophic lateral sclerosis (ALS).[228, 229]

Can we extend the analogy on proteins involved in conformational signaling, i.e. can we speak about conformational missignaling? Actually, we know about many cases when the aberrant conformational change of a protein causes disease. Such "conformational diseases"[230] include all the prion and amyloid diseases (Alzheimer's and Parkinson's diseases, and all other amyloidoses, which are often caused by mutations affecting the conformational state of the protein[160, 231, 232]), but also disorders caused by other mechanisms. This latter is exemplified by proteinase inhibitors serpins (such as α-antitrypsin, antithrombin, C1-inhibitor, and the inhibitors of fibrinolysis), which control coagulation and inflammation by inhibiting target proteases by large conformational changes in their reactive loop[233] that irreversibly locks the protease and serpin together. Familial mutations in serpins inhibit this conformational change and make serpins form inactive misfolded polymers and aggregates, leading to "serpinopathies", such as emphysema, cirrhosis, thrombosis and angioedema.

Missignaling may have more subtle appearances, when a shift in the conformational ensemble of the protein changes its interaction with partner protein molecules, compromising the fidelity of signaling. It is of note that many genes are harmful when they are overexpressed, i.e. they show "dosage sensitivity".[234] The gene products appearing at a much higher concentration are thought to cause disease, because they engage in interactions with non-physiological partners. No wonder these genes are tightly regulated at the level of transcription, translation, and protein regulation, in order to prevent their missignaling. It is also closely related that the functional relevance of many experiments addressing gene function with the aid of transiently transfected cells can be questioned due to the protein in question entering erroneous interactions.[235] It has been noted that overexpression violates balanced gene dosage, it affects protein folding, complex assembly and downstream regulation, which fits with our ideas of conformational missignaling.

Such general effects may be uncovered by analyzing protein-protein interaction networks (interactomes) in normal and disease conditions (cf. Scheme 4), as outlined in detail in the chapter "Approaching conformational signals in situ". In short, interactome analysis can delineate signaling relationships, i.e. signaling pathways,[10-12] and unveil the disruption of specific protein-protein interactions in disease, as exemplified by discovering genes potentially involved in cancer, for example.[236] Compromised protein-protein interactions may also point to inappropriate conformational signaling under the conditions of disease.[237, 238] The detailed description of such relations also offers new strategies for drug discovery, by developing small-molecule chaperones and disruptors of the pathologic interactions.

4.2. Prions as pathological and physiological signaling molecules

Probably the most direct example of conformational (mis)signaling via remodeled surfaces is represented by prions (Table 3). Discovered as proteinaceous infectious entities,[160, 231, 239] they were initially thought of as pathological conformational variants (prion or scrapie form, PrP^{Sc}) of an otherwise normal cellular protein (cellular form, Prp^{C}), which can propagate by an autocatalytic conformational change that can even be transmitted to other cells and organisms. Whether the transition proceeds by induced folding or conformational selection,[103] the essence of prion propagation is that the pathological, non-native conformation instructs the conformational transition of the native state in an autocatalytic fashion, upon their interaction (Figure 9). An initial conformational seed can cause the entire population of molecules to change, making the transition to the prion state appear as a signaling cascade based on the conformational principle.

Whereas pathological prions might be dismissed as signaling molecules for the lack of a "function" associated with the prion state, more recent findings suggest that several functional proteins (termed physiological prions) propagate their altered conformational and functional state in a prion-like fashion (Table 3). The best studied and/or most interesting examples are physiological yeast prions, which enable the cell to grow on poor nitrogen sources (Ure2p), or to generate novel protein products by translational read-through (Sup35p), when in the prion state,[204, 240, 241] and the cytoplasmic polyadenylation element binding protein (CPEB) in *Aplysia californica*, a protein that can activate dormant mRNA molecules by promoting their polyadenylation, which seems to be important in memory formation.[205]

In prions, the pathological or physiological readout of the conformational signal can be the sequestration of the prion itself, the recruitment of other proteins or the alteration of the activity of

domains associated with the prion domain (the region of the protein that can undergo autocatalytic conformational transition).[203, 242, 243] Therefore, physiological prions represent both inter- and intramolecular conformational signaling.

The generality of this molecular mechanism is also shown by non-prion neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's disease), in which the spread of the pathological state within and between neurons is mediated by the prion-like propagation of the altered conformational state (amyloid) of the polypeptide involved.[244, 245]

5. Novel structural biology tools for conformational signals

A basic difference between PTMs and conformational changes potentially involved in signaling is that the latter are non-covalent, i.e., they may be very sensitive to experimental conditions and may not reveal themselves under *in vitro* conditions and/or may not survive isolation. A recent surge of technical advances, however, raises hope that we may overcome these hurdles as there is a clear progress toward increasing sensitivity that enables to detect and characterize even minor but potentially relevant conformational changes. I outline the most exciting cases next (cf. also Table 4).

5.1. X-ray crystallography

X-ray crystallography has been the workhorse of structural biology in the past, delivering more than 80% of structures deposited in the Protein Data Bank (PDB).[246] The current development of X-ray free-electron laser (XFEL) technology seems to offer a quantum leap in this technique, the magnitude of which is hard to predict at the moment. At its conception, it was thought to be able to solve structures of single protein molecules by very short, femtosecond, X-ray laser pulses generated by oscillating free electron beams in a magnetic undulator,[247] in a time much shorter than what it takes for the sample to disintegrate (in a matter of tens of femtoseconds). Whereas the solution of the structure of a single molecule remains to be seen, it has already been demonstrated that high-quality diffraction data can be obtained with a single X-ray pulse from a noncrystalline biological sample, such as a single mimivirus particle, taking advantage of its highly internal symmetry[248], enabling the full three-dimensional reconstruction of structure.[249] XFEL has also been successfully used to determine the structure of photosystem I, one of the largest membrane protein complexes, by shooting at nanocrystals of only 0.2-2 μm in size.[250]

Novel applications of XFEL entirely relevant with the concept of conformational signaling have also been demonstrated, such as the solution of the high resolution structure of the human GPCR serotonin receptor at room-temperature by shooting sub-micrometer microcrystals grown in a membrane mimetic lipidic cubic phase.[251] This approach, in contrast with traditional crystallography on much larger cryo-cooled crystals, can provide information on the distribution of thermal motions and conformations at a temperature relevant for life.

5.2. NMR

NMR, an inherently insensitive technique that has delivered about 16% of PDB structures, has also witnessed a revolutionary advance recently. New developments allow to ask questions hardly amenable even a few years ago, such as analyzing the dynamic behavior of large systems (by specific labeling),[253] characterizing proteins at concentrations an order of magnitude lower than it was possible before with dynamic nuclear polarization (DNP),[252, 261, 262] visualizing thus-far invisible activated states (populated only 1-2% of the time) by relaxation dispersion measurements,[254, 263-265] and probing structural details in the solid (aggregated) state[266] or in live cells[267]. I outline these advances next.

Large systems can now be approached by NMR using site-specific labeling. For example, although the 20S core particle of the proteasome is 670 kDa, way above the size amenable for traditional NMR technology, it was possible to study it in detail by specific isotope labelling.[253] In this, Ile, Leu and Val methyl groups have been protonated in highly deuterated background, which enabled to characterize functionally important motions and interactions in this large protein complex.

Activated states of proteins that constitute only a minor part of the population of structures (and are thus invisible to most traditional structural techniques) are essential for protein function and may be critical in representing conformational signals. A study of dihydrofolate reductase based on relaxation dispersion experiments has demonstrated the utility of NMR in delineating such high-energy conformational sub-states.[268] The study was based on the premise that dynamics leading to various sub-states are critical in various steps along the catalytic cycle of the enzyme. NMR relaxation dispersion has shown such high-energy activated states along the reaction coordinate, resembling the structure of adjacent intermediates seen by X-ray crystallography.[268-270] Similar observations have been made with the prolyl cis-trans isomerase cyclophilinA (CypA), characterizing its substrate-free state and dynamics during catalysis.[70] It was observed that characteristic enzyme motions that occur during the catalytic cycle are already represented in the free enzyme, at frequencies that correspond

to the turnover rates of the enzyme, i.e., delocalized collective motions may govern enzymatic steps.[271, 272] Relaxation dispersion has also been instrumental in studying dynamics-based allosteric communication in PDZ structure.[81, 273]

Perhaps even more relevant with regards to conformational signaling, NMR relaxation dispersion can probe the details of binding mechanism in protein-protein interactions,[254] unraveling high-energy intermediary states in the encounter complex that are of very low abundance in the population. As already mentioned, the KID domain of CREB binds to the KIX domain of CBP, and goes from a disordered state to a folded state dominated by two helices.[274] Although CREB transiently samples the helical conformations in isolation,[108] it could not be decided if the mechanism of binding is dominated by the capture of pre-formed helices (conformational selection), or by the capture of a primarily disorder state followed by folding in the partner bound state (induced folding). NMR relaxation dispersion experiments showed that phosphorylated CREB KID (pKID) forms an ensemble of structures in the transient encounter complex with CBP KIX, in which local structure is primarily stabilized by non-specific hydrophobic contacts. Helical conformations evolve in the bound state, without dissociation from KIX, i.e. the transition state of helix B (the C-terminal helix) is only partially folded in the intermediate.

Another breakthrough in NMR technology is the rapid advance of solid-state NMR approaches. The basic idea here is that anisotropic interactions in solids, which cause severe line-broadening, can be mitigated by spinning the sample at a large frequency at a particular angle relative to the magnetic field (magic angle spinning), which allows traditional NMR parameters to be determined. This approach may even have a cross-talk with XFEL, as demonstrated by characterizing the structure and backbone dynamics of a microcrystalline metalloprotein.[255] Solid-state NMR enabled to determine the high-resolution structure of fibrils associated with prions and with Alzheimer's disease, for example,[266] and enabled to study the structure and functional dynamics of a 7-helix transmembrane protein, Anabaena Sensory Rhodopsin.[275, 276]

5.3. Atomic force microscopy

Atomic force microscopy (AFM) also holds a tremendous potential in characterizing conformational changes of proteins. AFM applied in the pulling mode enables to monitor structural stability through unfolding transitions,[257] such as in the case of the IDP α-synuclein, a protein implicated in Parkinson's disease. The protein showed a significant level of structural heterogeneity, encompassing three main classes of conformations, including disordered (random-coil), "β-like" and a less characterized one in

which molecules engage in weak interactions with each other. The significance of β-like structures is that they can lay on path toward the pathological amyloid state, because their proportion increases under conditions promoting the aggregation of α-synuclein, such as the pathogenic A30P mutation and high ionic strength.

AFM can also be applied in the tapping mode, when it can visualize function-related dynamic conformational changes of the protein. For example, fast AFM capable of repetitive surface scans at times only a few ms apart was used to study the dynamic behaviour of myosin V molecules translocating along actin filaments.[256, 277] Fast AFM enabled direct visualization of this motor protein taking processive steps, providing visual evidence for previously speculated mechanistic details, such as lever-arm swing.

5.4. Single-molecule fluorescence

Single-molecule fluorescence (smFL) enables extremely sensitive detection of structural features/changes of proteins either by measuring fluorescence resonance energy transfer (smFRET) between dye pairs and/or resolution of fluorescence lifetime and decay of fluorescence anisotropy (smFA). For example, by applying a great number of donor-acceptor pairs, smFRET could be used to describe the structure and structural dynamics of HIV-1 reverse transcriptase bound to its DNA primer-template.[278] In a different study, membrane binding of α-synuclein was analyzed,[279] and multiple α-helical structures have been observed that fall into two main classes. smFRET-based conformational studies of full-length tetrameric p53 also suggested multiple conformations of the protein, which likely indicates the interactions of the disordered NTD with the DNA binding domain of each monomeric p53.[280]

In an even more complex multidomain protein, the canonical MAGuK scaffold protein PSD-95 (which contains five domains PDZ1, PDZ2, PDZ3, SH3 and GuK), multiple distance restraints derived in smFRET studies were used to describe its complex conformational landscape outlining characteristic conformational changes.[281] Its five domains were found to partition into two independent supramodules held together by transient domain-domain interactions: one including the module PDZ1-PDZ2 and the other PDZ3-SH3-GuK, which defined the dynamic “supertertiary”[282] structural organization of MAGuK. The most exciting feature of this technique, from the point of view of conformational signaling, is that it enables single-molecule conformational studies in live cells, as detailed in chapter “Approaching conformational signals in situ”.

5.5. Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) has also undergone a revolutionary advance recently, enabling to study and characterize structurally highly heterogeneous systems, such as flexible multi-domain proteins and IDPs.[258] A key ingredient of recent advance is the development of Ensemble Optimization Method (EOM), which allows for the co-existence of multiple protein conformations in solution, and fits scattering data with a compatible sub-set. Often, the method is combined with NMR to incorporate short-range NMR constraints, and X-ray crystallography to represent high-resolution structural elements, by way of advanced computational methods, as exemplified by the structural ensembles of p53,[283] α-synuclein and tau protein[100] and also of Hck Tyr kinase, which converts from a compact and well-structured inactive state to an ensemble of open, highly dynamic conformations upon activation.[284]

5.6. Cryo-electron microscopy

Perhaps the most spectacular progress in protein structure determination has been brought along by rapid advances of cryo-electron microscopy (cryo-EM) technology, relying on the reconstruction of images recorded from a large number of molecules. Over the past two decades, the resolution of cryo-EM maps has been improving steadily, primarily due to the introduction of direct electron detection, automated data collection and powerful image processing.[259, 285] Near-atomic resolution can now be reached not only for megadalton macromolecular complexes (such as the ribosome[286]), highly symmetrical assemblies (such as viruses[287-289]) or membrane proteins that form ordered two-dimensional arrays (such as the mammalian ryanodine receptor RyR1[290-292]), but also for complexes with either low or no symmetry (such as the general transcription factor TFIID,[293] the transcription co-activator Mediator complex,[294] and the RNA polymerase II transcription pre-initiation complex[295]), smaller, less symmetrical proteins,[296] and even for highly dynamic proteins of multiple structural states.[297]

A common strategy is to combine cryo-EM of large proteins or complexes with X-ray crystallographic structures of subunits/domains to obtain a full structural picture at large resolution, as exemplified by the structure of Sulfolobus turreted icosahedral virus,[298] the cell-cyle regulatory protein separase[299] and the estrogen receptor-coactivator complex on DNA.[300]

5.7. Native mass-spectrometry

Mass-spectrometry (MS) has also natured into an effective tool for characterizing protein structure, primarily for addressing the topology of protein complexes. It was discovered that under mild ionizing conditions (native MS), primarily by electrospray ionization (ESI), proteins and protein assemblies retain their solution phase structural properties in the gas phase, which, combined with the superior mass-resolution of MS enables to address a range of issues related to structural biology. The power of native MS is exemplified by studying the architecture of chaperonin complex,[301] the evolutionary conservation of protein complex assembly pathways[130] or conformational changes in the multi-drug resistance efflux pump.[302] This method is also often integrated with other approaches, for example, the structure and subunit assembly pathway of the anaphase-promoting complex (APC), was characterized by combining cryo-EM, native MS and docking of crystallographic and homology-derived coordinates.[260]

A further asset of MS is that it can be combined with a variety of labeling methods, which provide further information on the structural state of the protein. A range of recent applications exploit the performance of MS in analyzing the pattern of chemical cross-linking, H/D exchange and surface labeling.

In short, chemical cross-linking of reactive groups in proteins followed by the identification of cross-linked sites by MS can measure proximity of various sites, thus providing information about protein structure, structural changes and the topology of protein complexes.[303, 304] Chemical cross-linking also has the capacity to stabilize weak and transient interactions, thus, to capture and analyze interactions *in vivo*: a special field is *in vivo* crosslinking of protein complexes by photo-reactive amino acid analogs in cells by UV-activated diazirines.[305] The information content and structural resolution of crosslinking is usually limited by the distinct chemical reactivity of amino acids. This has recently been overcome by the application of the heterobifunctional reagent, sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA), which combines a traditional sulfo-N-hydroxysuccinimide (sulfo-NHS) ester and a UV photoactivatable diazirine group. This diazirine yields a highly reactive and promiscuous reactive carbine species, which can react with practically all amino acids, and, when combined with computational conformational space search, has the potential to unveil the entire structure of a protein; this has been demonstrated for human serum albumin crosslinked in human blood serum.[306]

Hydrogen/deuterium (H/D) exchange, also called HDX, is a mild way of labeling proteins, and is a very sensitive indicator of the solvent accessibility of peptide bonds.[307] In solution, amide hydrogens in the peptide bonds exchange protons with the solvent, and if the protein is placed into heavy water, they

exchange with deuterium. Usually, the exchange is performed at physiological pH, the reaction is quenched by lowering the pH, and then analysis is performed by NMR or MS. The rate of exchange is measured and quantitatively characterized by a "protection factor", i.e. the ratio of the "intrinsic" amide hydrogen exchange rate of a small peptide of the same sequence (assuming random coil conformation) and the peptide bond in the given conformational state of the protein. By definition, the protection factor is sensitive to the conformational state and conformational changes of the protein[308, 309] and HDX has started to be applied recently for the characterization of IDPs.[310]

Covalent labeling, or footprinting, is also sensitive to probe the accessible surface of proteins. A variety of reagents can be used for this purpose, of which the most widely used and successful is labeling with hydroxyl radicals.[311, 312] Hydroxyl radicals are generated by radiolysis of water or photolysis of H_2O_2, when it is also termed fast photochemical oxidation of proteins (FPOP). The exact location of oxidation can be determined by MS/MS (similar to PTM mapping). In principle, all side chains can react, although the reaction rates for different amino acids significantly differ. By quenching (e.g. by the addition of L-glutamine), the lifetime of the radical can be limited to about 1 µs, which is faster than most conformational changes of proteins, which makes the method very effective in mapping interaction interfaces and protein structures, also effectively identifying conformational changes. A relevant recent study is the characterization of the fast conformational switch between the open and closed states of a potassium channel.[313]

6. Approaching conformational signals in situ

Motivated by the spectacular advance in structural biology surveyed in the previous chapter, we have to ask how effectively can this technological arsenal help delineate conformational signals in live cells. This is not a trivial question and the definite answer can only be given as dedicated research unfolds in the coming years.

6.1. High-throughput approaches to post-translational modifications of proteins in the cell

To set the stage for the possible high-throughput (HT) discovery of conformational signals, we have to understand and appreciate the limitations of our insight on the identity and functional consequences of covalent protein modifications in the cell, not the least because of the strong conceptual parallels of covalent and conformational signaling. Even with PTMs, we tend to take them into consideration only if they are well understood, having conclusive evidence for their functional

importance. For example, we have no doubt that the phosphorylation of glycogen phosphorylase by phosphorylase kinase activates the enzyme and leads to glycogen mobilization,[314] or ubiquitination by Mdm2 destabilizes p53, ensuring its low cellular level prior to DNA damage.[315] This "simple" situation has significantly changed with the HT identification of PTMs by mass spectrometry:[316, 317] whereas these tools generated previously unimaginable amount of data, we lack confidence in the functional importance of most of these. For example, recent analyses suggested more than 100k phosphorylation sites in the human proteome,[318] and an unbiased assessment of the total number of the possible functional motifs suggests their number above 1 million in total,[193] for the more than 300 distinct types of enzymatic PTMs.[319] Despite the statistical rigor of HT analyses, experimental assessment of the functional importance of changes in most of these cases is lacking. Furthermore, drawing on the analogy with HT protein-protein interaction data, which has been concluded to be inherently noisy,[320] we may also infer that many modifications might be missed because of i) low abundance of proteins, ii) low stoichiometry of modifications, iii) their specificity to a particular cell type or physiological state, and iv) their confinement to rare splice variants of proteoforms.

There are clear indications that a modification that is reproducibly detectable is not necessarily functional. In a recent analysis,[321] it was found that although phosphosites are more conserved than equivalent residues that are not phosphorylated, phosphosites of known function are highly significantly more conserved than the ones without a known function. Since about two thirds of the sites detected by phosphoproteomic experiments evolve as fast as random Ser and Thr residues, it is rather imperative to conclude that their modification is probably non-functional.

By analogy, even if we identify significant conformational changes of proteins in the cell, it is a long way to prove that they are important and functional in a signaling sense. Whereas we have increasingly effective tools to address conformational changes *in vitro* and even *in vivo*, the use of any of them alone is unlikely to be conclusive, and their combined use and interpretation will only bring the synergy required for pinning down conformational changes of signaling importance.

6.2. The limits of extrapolation from *in vitro* results

Since a conformational change is a common theme in structural biology, there must be many as yet unnoticed signaling cases. Is conformational signaling more prevalent than suggested by the few instances discussed in this paper? This whole review has been motivated by that this is in fact the case. We might take the analogy of PTMs, the current discovery of which relies heavily on MS-based HT techniques, due to which we have a rather comprehensive view of all such modifications in the

cell.[193] Quantitative temporal (and spatial) resolution of such modifications has been - and still is - instrumental in discovering and describing dynamic signaling pathways.[322, 323]

In contrast, our toolkit for the unbiased description of conformational changes of proteins in situ or *in vivo*, let alone conformational changes relevant for signaling, are much more limited. Practically all the examples outlined in the paper have come from detailed structural-functional studies on isolated proteins and complexes *in vitro*. Although these provide insight into how proteins sense regulatory signals by conformational changes, they are difficult to extrapolate to conditions that prevail in the cell, for at least three reasons. For one, the cellular environment is extremely complex,[324] and it contains a very high concentration of other macromolecules, which gives rise to a viscous and highly crowded environment that alters the conformational landscape of proteins by excluded volume effects.[325] Although mimicking this intracellular environment does provide some hints at the structural behavior of proteins in the cell,[326, 327] full understanding can only be expected from techniques that can peek into live cells.[328] Second, the cellular environment presents a multitude of potential interaction partners, for example the number of identified high-quality binary macromolecular interactions in a human cell is on the order of 14,000,[329] whereas theoretical estimations suggest as many as 650,000 protein-protein interactions in the human interactome.[330] In addition, this medium contains a great variety of ions and metabolites, and the interactions of any given protein with most of these are unexplored in terms of affecting its conformational state. Current estimates of the human metabolome suggest as many as 40,000 different compounds.[331, 332] Last but not least, PTMs of proteins in the cell can be extremely complex, and represent a largely unexplored territory. Whereas most of our respective knowledge focuses on only a few modifications (phosphorylation, acetylation, ubiquitination, methylation), there are more than 300 different types of PTMs in the cell,[319] HT studies suggest more than hundred thousand actual modifications in the human proteome,[318, 333] and unbiased assessment suggest more than a million such modification sites in the cell.[193] This means that, on the average, every protein carries dozens of modifications *in vivo*, which can be in a complex interplay with protein conformation, i.e., conformational signals (Scheme 5).

Next, we will explore if our new structural biology methods are able to provide structural information on the proteome, relevant with signaling, in live cell conditions.

6.3. Projecting PTM data on the interactome to ascertain conformational signals *in vivo*

An indirect approach is provided by the vast collection of protein-protein interactions[334-337] and PTMs[318, 333] catalogued by HT methods (Scheme 4). We might agree that a direct readout of the appearance of a conformational signal is the local remodeling of the interactome. We have a highly effective array of HT techniques for studying protein-protein interactions in the cell (e.g. co-IP, TAP-tag, yeast 2-hybrid (Y2H), protein-fragment complementation assays (PCA), MAPPIT, proximity ligation assay (PLA), and many others[338-342]). As suggested, we know about 14,000 high-confidence binary interactions in the human interactome,[329] which might only be the tip of the iceberg of a total of 650,000 interactions that may be realized under some conditions between human proteins.[330] The observed interactions have been deposited alongside the interactome of other model organisms into databases, such as DIP, IntAct and MINT.[343-345] These data (and future applications of the relevant techniques) are extremely rich in information on conformational signaling, as we might intuitively agree that a change of interaction pattern upon some cellular stimulus may directly point to the generation and operation of conformational (and covalent, of course) signals. Actually, interactome analysis has been very powerful in delineating signaling pathways.[10-12] One has to be fully aware, however, that the different HT approaches for protein interactions are very different in terms of their potential to uncover binary interactions, and thus provide direct information on the possible operation of conformational signals. Some techniques (e.g. Y2H, MAPPIT, PCA) are invented for binary interactions whereas others (co-IP, TAP-tag) more for interactions within complexes, supported by partner proteins. As suggested, the first type of interaction is absolutely relevant for conformational signals, whereas the second has limited value, because we cannot unequivocally ascribe the (change of the) interaction to a particular protein or protein pair. A further caveat to these studies is that tagging proteins and/or their ectopic expression can skew regulation, and create an artificial situation in signaling.[235]

Nevertheless, HT protein-protein interaction maps and techniques might be our first take on the "conformational signalome". It should not be mistaken though, that a change in interactions may simply reflect covalent signals, i.e. PTM(s) of the given protein or changes in the level of the proteins or third partners. To exclude these possibilities, we should project HT PTM data on the interactome:[318, 333] if there is a change in the local interaction pattern but not in the PTM status of a protein, we might reasonably suspect that we deal with a conformational signal (Scheme 4). Of course, it may also be that a PTM causes a conformational change, which then promotes interaction with another protein, which we also consider a conformational signal that requires more detailed studies and much deeper mechanistic understanding (Scheme 5). Nevertheless, it might be highly inspirational to conduct targeted studies on the perturbation of the interactome[346] and to analyze

and compare current HT interactome and PTM data to compile a putative list of conformational signaling cases.

6.4. Looking for conformational signals in the cell

To nail the mechanism down, however, we should resort to more direct structural techniques. As already alluded to in the previous sections, several structural biology techniques have the capacity to probe structural changes of signaling importance in cellular extracts (in situ) and even in live cells (*in vivo*). I will survey these possible approaches next (Table 5).

For example, smFRET has the sensitivity and specificity to provide insight into structural changes of proteins in a cellular context, even in a HT format. For that, smFRET requires the delivery of labeled proteins into cells, and offers detection of conformational changes at the single-molecule level. This has been demonstrated with soluble NSF attachment protein receptor (SNARE) proteins in live cells.[348] SNARE proteins mediate vesicle fusion, for example, of synaptic vesicles with the presynaptic membrane in neurons. By microinjecting SNARE proteins site-specifically labeled by fluorescence donor-acceptor dye pairs (such as Alexa Fluor 555/Alexa Fluor 647 or Cy3/Cy5) into BS-C-1 cells, it was observed by very sensitive total internal reflection fluorescence spectroscopy (TIRFM) that they rapidly undergo a transition from a low-FRET (disordered) to a high-FRET (coiled-coil, structured) state, which indicates their rapid association with membranes and transition to a conformational state conducive to membrane fusion. Apparently, this approach enables single-molecule applications, whereas it requires *in vitro* labeling and microinjection into cells (potentially hindering HT applications). For high throughput, one should use fused protein tags (such as green fluorescent protein (GFP) and its variant, yellow fluorescent protein (YFP)[351]), which can be generated directly in the cell.

In vivo detection can also be achieved by NMR, which is rather insensitive and reports on ensemble averages; recent applications make it compatible with the subject of in situ structural characterization. For such in-cell NMR, isotopically labeled proteins are delivered into live cells, achieved by either microinjection (into Xenopus oocytes), induced overexpression (in *E. coli* cells), delivery by cell-penetrating peptides or transiently permeabilizing cell membranes by electroporation (into mammalian cells).[267, 328, 352]

Although in-cell NMR initially provided a breakthrough by solving protein structures *in vivo* (e.g. protein G B1 domain and heavy-metal binding protein TTHA1718[353, 354]), recently this approach has

been used to address the structural state and phosphorylation of tau protein microinjected into Xenopus oocytes,[355] where the structural transition of its microtubule-binding region upon microtubule binding could be observed. The significance of this observation is also corroborated by α-synuclein studies, which, for the lack of an appropriate binding partner, undergo no such structural transition when overexpressed in *E. coli,*[356, 357] and remains fully disordered in human neuronal and non-neuronal cells under crowded intracellular conditions.[328]

The power of in-cell NMR in dissecting the signaling capacity of a local conformational change in a protein *in vivo* is probably best illustrated by studies on the activator domain of the tumor suppressor 53.[347] p53, the "guardian of the genome", is a transcription factor of a low constitutive level in the cell. Upon DNA damage, it is stabilized, its level increases and it initiates changes in the cell that lead to DNA repair, cell cycle arrest or apoptosis.[358] The primary regulatory partner of p53 is the E3 ubiquitination ligase Mdm2, to which p53 binds via a short helical segment within its N-terminal disordered transactivator domain (TAD).[359] The free form of p53 TAD exists in an equilibrium between disordered and partially helical conformations,[283, 360] with its region 19–25 forming a stable amphipathic α-helix in complex with Mdm2.[359] Mutations that increase the level of residual helicity of TAD increase the binding affinity between p53 and Mdm2, and facilitate its ubiquitination, which basically alter p53 cellular dynamics and lead to impaired target gene expression and failure to induce cell cycle arrest upon DNA damage.[347] These results show that the local helix within p53 TAD is a strong conformational signal in the cell.

Conformational changes of proteins in complex biological mixtures (in situ and/or *in vivo*) can also be addressed by limited proteolysis. It is known that the location of initial cleavage sites under controlled conditions by broad-specificity proteases (such as proteinase K, thermolysin, subtilisin, papain, and elastase) is dictated by structural features of proteins,[361-363] which makes it a sensitive and sensible approach for the low-resolution mapping of conformational changes in purified proteins. A recent development combines limited proteolysis (LiP) in cellular extracts with selected reaction monitoring mass-spectrometry (LiP-SRM)[350] to achieve the same goal in complex biological mixtures. The method combines two proteolytic steps, a LiP under native conditions by a broad-specificity protease followed by exhaustive tryptinolytic digestion under denaturing conditions. Peptides are detected and quantified by SRM, and peptides specific to a given cellular state, indicative of a specific conformation (termed "conformotypic" peptides), are identified. For example, in analyzing proteome-wide conformational changes in yeast upon the metabolic transition from glucose- to ethanol-based growth, which induces substantial metabolic remodeling in yeast,[364] 587 specific peptides mapping onto 283 proteins could be observed.

Another low-resolution structural technique that has the potential to characterize structure and structural changes *in vivo* is chemical cross-linking,[303] which can provide distance constraints for structure calculation, as demonstrated first for fibroblast growth factor 2 (FGF-2).[365] Recent advances include novel cross-linking protocols (development of novel homo- and heterobifunctional cross-linkers), MS-identification and analysis software, which have provided important information on native protein structure and the topology of protein complexes,[366] as demonstrated for the ATP synthase complex[367] and the nuclear pore complex,[368] for example. It has also been used to analyze protein structure[369] *in vivo*: by using a collection of derivatives of chemoreceptor Trg, each containing a single cysteine at particular positions, the pattern of crosslinking *in vivo* and *in vitro* were compared to delineate the exact topology of protein-RNA interactions.[370] The solution of the structure of human serum albumin by extensive crosslinking in serum also suggests the power of this technique to solve the structure of a protein in situ.[306]

As a final note, it should also not be forgotten that even X-ray crystallography has recently advanced to making protein crystals and detecting diffraction in cells possible. A recent intriguing XFEL structural study[371] on cyrstals of Cry3A toxin naturally crystallized within Bt cells opens a novel avenue for authentic *in vivo* diffraction studies.

7. Conclusions and outlook: the prevalence of conformational signaling

The principal message of this paper is that there is a fine line between structural changes in signal transduction and the signaling principle based on conformational changes. Conformational changes are prevalent in signal transduction[15, 76, 372], and it is clear that the altered conformation of a protein in many cases conforms to the definition of a signal (Scheme 2). The most clear-cut cases are multi-state switches (e.g., prion, GPCR, CaM), where the signal is not represented by the protein itself, but by its distinct activated states, as embodied in particular conformations. The altered conformation(s) is (are) recognized by a downstream signaling partner, thus it can be propagated and transduced to alter the physiological state of the cell. PTMs, induced folding/unfolding, hierarchic assembly of complexes, conformational memory and regulatory chaperones may all be involved, and signaling is the very essence of allostery, conformational spread and prion propagation. It appears to pop up everywhere and there are already several well-established and instructive cases in the literature (Table 6). Its s prevalence is hard to establish, because conformational signaling elements or sections constitute part of all “traditional” signaling pathways, and our current views rely on detailed low-

throughput studies of *in vitro* reconstituted systems. As outlined, rapid methodological advance in structural biology offers alliance in this endeavor, providing ever more sensitive tools for uncovering relevant conformational changes *in vitro* (in the test tube), in situ (in cell extracts) and now also *in vivo* (in live cells) even at the level of HT coverage of the proteome (Scheme 4). We do anticipate that many such events (pathways) will be discovered, and will prove that conformational signaling is a prevalent cell-biological phenomenon. We should bear in mind that a covalent signal, such as a signaling metabolite or modified protein, is easy to identify, whereas the recognition of a local conformational change as a signal requires much deeper mechanistic insight, and requires the integration of *in vitro*, in situ and *in vivo* observations with HT *in vivo* data on the interactome and PTMs (Scheme 4).

For the emergence of conformational and covalent signaling, their evolutionary relationship appears to be rather straightforward. Simple conformational switches could already be present at a time when complex catalytic signaling enzymes have not yet emerged. As suggested, the first proteins must have been short and disordered,[139] which is sufficient to support a primitive, direct and non-modular conformational signaling mechanism. The rise of enzymatic signaling proteins/domains tipped the balance, because they can be effectively used in different contexts and they can modify multiple substrates, enabling amplification of the signal. Proteins/domains that specialize in "reading" the modification can also be repeatedly used. Based on these highly effective components, complex cellular regulatory behavior emerged. Yet, simple signaling by conformational changes has not yet vanished: it has also benefitted from modularity and it has its due place in the heart of most extant signaling pathways. Whereas the noted evolutionary sequence does not necessarily dominate in extant organisms (bacteria vs. eukaryotes), our concepts of covalent signal transduction including PTMs derives much more from studies on eukaryotes, where disordered sites preferring modifications are also more prevalent; at the same time the balance may be tipped toward signaling relying on conformational changes more in bacteria. This may have relevance for drug development, for example, when we are contemplating on developing a new breed of antibiotics.

It should also not be overlooked, though that many diseases of higher organisms, such as Alzheimer's disease, Parkinson's disease, prion diseases also rely on this principle (and are termed "conformational diseases") and are caused by the emergence of a non-physiological structural state of the protein.[160, 244, 245] As suggested, the (pathological) readout of such misfolded states may be the loss of function of the protein, but also unwanted binding to other proteins,[242, 243] that is, erroneous signaling through altering the interactome of the cell. With all exciting developments in structural

biology technology, we are now well equipped to tap into this vast area of cellular molecular biology, enabling to start exploring its far-reaching biomedical implications.

Legends of Figures

Figure 1 *Covalent and conformational signals in the cell*

Signal transduction in the cell relies on the appearance of a novel chemical entity, i.e. a novel spatial and chemical constellation of atoms perceived by a receptor protein with an appropriate binding surface. **A)** The classic example is the second messenger cAMP, which is generated form ATP by the action of the enzyme adenylyl cyclase. cAMP then binds to and activates a broad range of proteins. **B)** A messenger may also be "generated" by cellular translocation, as appears with Ca^{2+} ions, which, upon activation, are released from the endoplasmic reticulum into the cytoplasm where they can act through dedicated mediators (e.g. calmodulin). **C)** A covalent signal can also arise by the modification of a protein, as best exemplified by phosphorylation of Tyr residues by protein kinases. **D)** A novel constellation of atoms might also arise by the conformational change of a protein, as illustrated here by the appearance of a helical segment. This can bind to partners with the appropriate recognition site.

Figure 2 *Covalent modification of a protein can generate a conformational signal*

A) The modification of a protein can generate a covalent signal, when a PTM creates a novel binding surface recognized by a cognate binding domain (blue). **B)** The modification may also have its primary effect on the conformation of the protein, which either acts on the activity of the protein or serves itself as the signal, being recognized by a downstream effector.

Figure 3 *Allosteric signaling*

A) The essence of classical allostery is a conformational change that spreads between subunits of a homo-oligomeric protein, such as hemoglobin. This scheme shows that binding of the ligand (oxygen, a yellow square) to one subunit shifts the structure of the entire tetramer to favor subsequent binding events at the other subunits. The initial conformational change thus spreads between subunits as a signal. The conformational change is subtle and details are not shown. **B)** Allosteric conformational change can also spread within one subunit, as demonstrated for catabolite activator protein (CAP). CAP is a cAMP-dependent transcription factor, which has a DNA binding helical region, rendered incompetent for DNA binding by a 90-degree rotation in the absence of cAMP (yellow, pdb 2WC2[88]). When cAMP binds to both subunits (in a cooperative manner itself), it causes a significant conformational change, i.e. rotation of the helix, to assume a position relevant for DNA binding (blue, pdb 1J59[373]).

Figure 4 *Conformational signaling can be generated by induced folding and unfolding*

A) c-Myc belongs to the bZip transcription factor family, which are structurally disordered and inactive proteins in the monomeric state. Their activation is usually induced by a PTM (specific phosphorylation), which makes them assemble into a coiled coil dimeric structure. The mutual induced folding of the two subunits creates a signal (i.e. a DNA binding surface), due to which c-Myc binds the appropriate recognition element and initiates transcription. **B)** Induced unfolding can also generate a signal, as is the case with WASP protein. In the naïve state, the protein is kept inactive by the interaction between its N-terminal GBD (blue) and C-terminal VCA (green) domains. Upon activation by Cdc42 binding to GBD, the VCA region is released form GBD and is exposed for interaction with Arp2/3 to promote actin polymeriztation.

Figure 5 *Conformational switches: alternative states of G-protein coupled receptors*

Conformational signals can make switches of multiple states, which signal to distinct downstream signaling pathways. Canonical GPCR signaling proceeds through heterotrimeric G-protein partners, such as Gs$\alpha\beta\gamma$. The highly dynamic GPCR structure (blue), however, can assume multiple conformational states that can signal through distinct messenger system in a ligand-specific manner (distinct ligands indicated as L1 through L3). For example, the existence of multiple conformations has been shown for β2-adrenergic receptor by NMR,[77] but only in one state, in the presence of an agonist (green) and Gs heterotrimer (orange), could the structure of the receptor be solved by X-ray crystallography (pdb 3SN6).[149]

Figure 6 *Thermodynamics of conformational signaling*

Generation of a conformational signal can be conceived by two energetic pathways, both of which can be activated by any type of signaling input, i.e., a PTM, ligand, physical input, and interacting protein partner, or the action of specific chaperone-like proteins. **A)** If the active (act) conformation is more stable than the inactive (inact) one, but it is separated by a high energy barrier, activation can occur by bringing the barrier down. **B)** If the active conformation is less stable than the inactive one, signaling may remodel the conformational landscape (a) to make it more favorable, or (b) to bring it into a high-energy but kinetically stable state by transient interaction.

Figure 7 *Modularity in signal transduction*

A key principle of signal transduction is modularity, when the same functional unit is used in different evolutionary context(s). Distinct signaling pathways may use the same (type of) enzymatic activity, when the homologue of the same gene product is linked with different upstream and downstream signaling components. For example: **(A)** the same kinase can be recruited into distinct pathways (e.g. Ste11 MAPKKK in yeast used in the mating and osmosensing pathways[165, 374, 375] by the assistance of

different scaffold proteins (grey rectangles), or **(B)** the same domain can be used in combination with distinct other domains to constitute various signaling switches (e.g. the SH2 pTyr-binding is domain combined with other types of domains to create multidomain signaling proteins of complex signal integrating capacity[176, 376]).

Figure 8 *Modularity in conformational signaling*

Modularity is also apparent in conformational signaling, when the same module capable of generating a conformational signal is used in different contexts. **(A)** For example, calmodulin, a calcium-sensing mini-protein that changes conformation upon calcium binding is incorporated into myriads of pathways as a stand-alone protein, but it is also part of the multi-domain protein calpain, a calcium-dependent cysteine protease of two subunits, both incorporating a calmodulin-like domain (CaML). **(B)** A modular conformational signal can also be created artificially. The NM region of the physiological yeast prion Sup35p can undergo a reversible change to a prion (amyloid) state. In [psi-] cells, Sup35p is soluble and functional (causing translation termination at stop codons). When the NM region converts to the prion state, Sup35p becomes insoluble and nonfunctional, causing an increased rate of nonsense suppression (translational read-through of stop codons), termed the [PSI+] phenotype. When the NM region is fused to glucocorticoid receptor (GR), it turns GR into a prion (NMGR), which loses normal GR function when it converts to the prion state.[211] NM can thus be considered a module capable of generating a conformational signal.

Figure 9 *Conformational switches made of amyolids*

Conformational signals are dynamic, they can make switches that can build pathways. **(A)** Prions are autocatalytic conformational switches, which can exist in a naïve (cellular, inactive, PrP^{C}) and scrapie (prion, active, PrP^{Sc}) state. When the prion conformation arises, it can "signal" naïve molecules to undergo a conformational change and become prions themselves. Due to the autocatalytic nature of the conversion, the signal can affect many proteins, it can be amplified and even be transmitted between cells, very much like a signal in signaling cascades. **(B)** Formation of amyloids typically entails a transition to a cross-β structure, which forms the basis of autocatalytic transformation of the cellular form to the prion state. X-ray crystallography of a 7-residue segment of Sup35 prion domain reveals a tightly-packed β-conformational steric zipper structure.[377] Reproduced with permission from Nature Publishing Group.

Scheme 1 *MAPK cascade*

The MAPK cascade transduces information from an extracellular signal (EGF) to activating functional genes (by the transcription factor Elk1) via a cascade of phosphorylation events of successive kinases

(shown by **P** marks). Nevertheless, several proteins of the pathway (shown in dark) undergo large signaling conformational changes, which are not necessarily linked to phosphorylation.

Scheme 2 *Covalent and conformational signals have to meet the same requirements*

There are several requirements for both a covalent (a small molecule, covalently modified protein) and a conformational (a protein of altered conformational state) signal, which they have to meet to be able to play a signaling role.

Scheme 3 *Assembly pathway of a complex is determined by conformational signals*

The assembly of macromolecular complexes does not occur by the random addition of components, but usually follows a concrete pathway. The scheme depicts such as pathway, in which proper assembly (full arrows and dark objects) is defined by conformational signals that develop upon the assembly of the components (object **a** and **b** turning to **a'** and **b'** upon interaction with **b** and **c**, respectively) in a predefined order. The addition of components out of their proper context (dashed arrows) results in dead-end, non-specific interactions (grey objects) that cannot proceed toward the final fully functional state.

Scheme 4 *Synergy of structural methods can address conformational signals*

The scheme depicts that the HT identification of conformational signals takes much more than simply solving structures or describing structural changes of isolated proteins. Approaches applicable on isolated proteins *in vitro* have to synergize with studies *in situ* (on extracts) and *in vivo.* Furthermore, data generated by these approaches have then to be co-interpreted with HT interactome and PTM data, to set apart conformational signals and outline likely conformational signaling pathways in the cell.

Scheme 5 *Interplay of covalent and conformational signals*

A PTM is thought to generate a covalent signal (a modified protein), whereas other stimuli are thought primarily to act by conformational changes. As shown through many examples throughout the manuscript, the mode of action of the two stimuli overlaps, often PTMs act by way of conformational changes that act as signals.

Acknowledgements

This work was supported by the Odysseus grant G.0029.12 from Research Foundation Flanders (FWO). Thanks are due to Han Remaut, Jan Steyaert, Kris Pauwels, Mainak Guha Roy, Laszlo Buday, and Shoshana Wodak for their invaluable discussion and advice on the manuscript. I am also indebted to Simone Kosol for her help in preparing the figures

References

1. M. J. Berridge, P. Lipp and M. D. Bootman, *Nat Rev Mol Cell Biol*, 2000, **1**, 11-21.
2. E. W. Sutherland, *JAMA*, 1970, **214**, 1281-1288.
3. G. Pearson, F. Robinson, T. Beers Gibson, B. E. Xu, M. Karandikar, K. Berman and M. H. Cobb, *Endocr Rev*, 2001, **22**, 153-183.
4. C. S. Sorensen and R. G. Syljuasen, *Nucleic Acids Res*, 2012, **40**, 477-486.
5. R. Kitagawa and M. B. Kastan, *Cold Spring Harb Symp Quant Biol*, 2005, **70**, 99-109.
6. A. Ciechanover, *Nat Rev Mol Cell Biol*, 2015, **16**, 322-324.
7. D. Y. Lee, C. Teyssier, B. D. Strahl and M. R. Stallcup, *Endocr Rev*, 2005, **26**, 147-170.
8. M. Lamkanfi, N. Festjens, W. Declercq, T. Vanden Berghe and P. Vandenabeele, *Cell Death Differ*, 2007, **14**, 44-55.
9. H. A. Kestler, C. Wawra, B. Kracher and M. Kuhl, *Bioessays*, 2008, **30**, 1110-1125.
10. K. G. Guruharsha, K. Hori, R. A. Obar and S. Artavanis-Tsakonas, *Methods Mol Biol*, 2014, **1187**, 181-192.
11. I. Remy and S. W. Michnick, *Methods Mol Biol*, 2015, **1278**, 467-481.
12. P. Tieri, A. Termanini, E. Bellavista, S. Salvioli, M. Capri and C. Franceschi, *PLoS One*, 2012, **7**, e32678.
13. O. Ivanciuc, C. H. Schein, T. Garcia, N. Oezguen, S. S. Negi and W. Braun, *Regul Toxicol Pharmacol*, 2009, **54**, S11-19.
14. R. Nussinov, C. J. Tsai and B. Ma, *Annu Rev Biophys*, 2013, **42**, 169-189.
15. R. G. Smock and L. M. Gierasch, *Science*, 2009, **324**, 198-203.
16. G. Wei, W. Xi, R. Nussinov and B. Ma, *Chem Rev*, 2016.
17. Z. Ahmed, Z. Timsah, K. M. Suen, N. P. Cook, G. R. t. Lee, C. C. Lin, M. Gagea, A. A. Marti and J. E. Ladbury, *Nat Commun*, 2015, **6**, 7354.
18. J. Suckale and M. Solimena, *Front Biosci*, 2008, **13**, 7156-7171.
19. H. Tidow and P. Nissen, *FEBS J*, 2013, **280**, 5551-5565.
20. M. Cargnello and P. P. Roux, *Microbiol Mol Biol Rev*, 2011, **75**, 50-83.
21. A. Leutz, O. Pless, M. Lappe, G. Dittmar and E. Kowenz-Leutz, *Transcription*, 2011, **2**, 3-8.
22. D. L. Swaney, P. Beltrao, L. Starita, A. Guo, J. Rush, S. Fields, N. J. Krogan and J. Villen, *Nat Methods*, 2013, **10**, 676-682.
23. E. Thomas and K. Yiannis, *Progress in Retinal and Eye Research*, 2001, **20**, 49–94.
24. P. K. Sorger, *Cell*, 1991, **65**, 363–366.
25. E. A. Kravitz, *Science* 1988, **241**, 1775–1781.
26. M. R. Picciotto, M. J. Higley and Y. S. Mineur, *Neuron*, 2012, **76**, 116–129.
27. A. D. Mooradian, J. E. Morley and S. G. Korenman, *Endocrine Reviews*, 1987, **8**, 1–28.
28. S. Gadkar-Sable, C. Shah, G. Rosario, G. Sachdeva and C. Puri, *Front. Biosci.*, 2005, **10**, 2118–2130.
29. C. W. Ward and M. C. Lawrence, *BioEssays* 2009, **31**, 422–434.
30. C. L. Holmes, D. W. Landry and J. T. Granton, *Critical Care*, 2003, **7**, 427.
31. G. Carpenter and S. Cohen, *J. Biol. Chem*, 1990, **265**, 7709–7712.
32. W. Liao, J. X. Lin and W. J. Leonard, *Curr. Opin. Immunol.*, 2011, **23**, 598–604.
33. T. R. G. G. o. t. G. O. Consortium, *PLoS Comput Biol*, 2009, **5**, e1000431.
34. J. Cox and M. Mann, *Annu Rev Biochem*, 2011, **80**, 273-299.
35. M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K. K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-

Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman and M. Snyder, *Nature*, 2012, **489**, 91-100.
36. M. D. Bootman, T. J. Collins, C. M. Peppiatt, L. S. Prothero, L. MacKenzie, P. De Smet, M. Travers, S. C. Tovey, J. T. Seo, M. J. Berridge, F. Ciccolini and P. Lipp, *Semin Cell Dev Biol*, 2001, **12**, 3-10.
37. P. S. Salonikidis, A. Zeug, F. Kobe, E. Ponimaskin and D. W. Richter, *Biophys J*, 2008, **95**, 5412-5423.
38. K. Van Roey, T. J. Gibson and N. E. Davey, *Curr Opin Struct Biol*, 2012, **22**, 378-385.
39. D. Bray and T. Duke, *Annu Rev Biophys Biomol Struct*, 2004, **33**, 53-73.
40. A. Wittinghofer and I. R. Vetter, *Annu Rev Biochem*, 2011, **80**, 943-971.
41. D. Stapleton, K. I. Mitchelhill, G. Gao, J. Widmer, B. J. Michell, T. Teh, C. M. House, C. S. Fernandez, T. Cox, L. A. Witters and B. E. Kemp, *J Biol Chem*, 1996, **271**, 611-614.
42. S. S. Taylor, J. Yang, J. Wu, N. M. Haste, E. Radzio-Andzelm and G. Anand, *Biochim Biophys Acta.*, 2004, **1697**, 259-269.
43. M. Huse and J. Kuriyan, *Cell*, 2002, **109**, 275-282.
44. L. N. Johnson, M. E. Noble and D. J. Owen, *Cell*, 1996, **85**, 149-158.
45. X. Zhang, J. Gureasko, K. Shen, P. A. Cole and J. Kuriyan, *Cell*, 2006, **125**, 1137-1149.
46. D. F. Brennan, A. C. Dar, N. T. Hertz, W. C. Chao, A. L. Burlingame, K. M. Shokat and D. Barford, *Nature*, 2011, **472**, 366-369.
47. S. V. Kozlov, M. E. Graham, B. Jakob, F. Tobias, A. W. Kijas, M. Tanuji, P. Chen, P. J. Robinson, G. Taucher-Scholz, K. Suzuki, S. So, D. Chen and M. F. Lavin, *J Biol Chem*, 2011, **286**, 9107-9119.
48. P. D. Jeffrey, A. A. Russo, K. Polyak, E. Gibbs, J. Hurwitz, J. Massague and N. P. Pavletich, *Nature*, 1995, **376**, 313-320.
49. W. Xu, A. Doshi, M. Lei, M. J. Eck and S. C. Harrison, *Mol Cell*, 1999, **3**, 629-638.
50. P. Cohen, *Nat Cell Biol*, 2002, **4**, E127-130.
51. P. B. Daniel, W. H. Walker and J. F. Habener, *Annu. Rev. Biochem.*, 1998, **18**, 353-383.
52. D. E. Clapham, *Cell*, 2007, **131**, 1047-1058.
53. T. M. Lincoln and T. L. Cornwell, *FASEB J.*, 1993, **7**, 328-338.
54. S. Dimmeler, I. Fleming, B. Fisslthaler, C. Hermann, R. Busse and A. M. Zeiher, *Nature*, 1999, **399**, 601-605.
55. K. Mikoshiba, *J. Neurochem.*, 2007, **102**, 1426-1446.
56. A. Toker, *EMBO Rep.*, 2005, **6**, 310–314.
57. H. Kuboniwa, N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee and A. Bax, *Nat Struct Biol*, 1995, **2**, 768-776.
58. B. A. Seaton, J. F. Head, D. M. Engelman and F. M. Richards, *Biochemistry*, 1985, **24**, 6740-6743.
59. G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor and A. Bax, *Biochemistry*, 1992, **31**, 5269-5278.
60. A. V. Shkumatov, S. Chinnathambi, E. Mandelkow and D. I. Svergun, *Proteins*, 2011, **79**, 2122-2131.
61. U. P. Shinde, J. J. Liu and M. Inouye, *Nature*, 1997, **389**, 520-522.
62. C. B. Anfinsen, *Science*, 1973, **181**, 223-230.
63. P. Tompa, *Curr Opin Struct Biol*, 2011, **21**, 419-425.
64. P. E. Wright and H. J. Dyson, *Nat Rev Mol Cell Biol*, 2015, **16**, 18-29.
65. R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem Rev*, 2014, **114**, 6589-6631.
66. V. J. Hilser, E. B. Garcia-Moreno, T. G. Oas, G. Kapp and S. T. Whitten, *Chem Rev*, 2006, **106**, 1545-1558.
67. P. Tompa and M. Varadi, *Structure*, 2014, **22**, 177-178.

68. M. Varadi, S. Kosol, P. Lebrun, E. Valentini, M. Blackledge, A. K. Dunker, I. C. Felli, J. D. Forman-Kay, R. W. Kriwacki, R. Pierattelli, J. Sussman, D. I. Svergun, V. N. Uversky, M. Vendruscolo, D. Wishart, P. E. Wright and P. Tompa, *Nucleic Acids Res*, 2014, **42**, D326-335.
69. D. D. Boehr, R. Nussinov and P. E. Wright, *Nat Chem Biol*, 2009, **5**, 789-796.
70. E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay and D. Kern, *Nature*, 2005, **438**, 117-121.
71. D. Kern and E. R. Zuiderweg, *Curr Opin Struct Biol*, 2003, **13**, 748-757.
72. H. N. Motlagh, J. O. Wrabl, J. Li and V. J. Hilser, *Nature*, 2014, **508**, 331-339.
73. P. Tompa, *Chem Rev*, 2014, **114**, 6715-6732.
74. R. Nussinov and B. Ma, *BMC Biol*, 2012, **10**, 2.
75. O. F. Lange, N. A. Lakomek, C. Fares, G. F. Schroder, K. F. Walter, S. Becker, J. Meiler, H. Grubmuller, C. Griesinger and B. L. de Groot, *Science*, 2008, **320**, 1471-1475.
76. B. H. Luo and T. A. Springer, *Curr Opin Cell Biol*, 2006, **18**, 579-586.
77. R. Nygaard, Y. Zou, R. O. Dror, T. J. Mildorf, D. H. Arlow, A. Manglik, A. C. Pan, C. W. Liu, J. J. Fung, M. P. Bokoch, F. S. Thian, T. S. Kobilka, D. E. Shaw, L. Mueller, R. S. Prosser and B. K. Kobilka, *Cell*, 2013, **152**, 532-542.
78. B. Ma, M. Shatsky, H. J. Wolfson and R. Nussinov, *Protein Sci*, 2002, **11**, 184-197.
79. D. Tobi and I. Bahar, *Proc Natl Acad Sci U S A*, 2005, **102**, 18908-18913.
80. N. Tokuriki and D. S. Tawfik, *Science*, 2009, **324**, 203-207.
81. E. J. Fuentes, C. J. Der and A. L. Lee, *J Mol Biol*, 2004, **335**, 1105-1115.
82. S. R. Tzeng and C. G. Kalodimos, *Nature*, 2009, **462**, 368-372.
83. T. Wlodarski and B. Zagrovic, *Proc Natl Acad Sci U S A*, 2009, **106**, 19346-19351.
84. A. W. Fenton, *Trends Biochem Sci*, 2008, **33**, 420-425.
85. J. P. Changeux and S. J. Edelstein, *Science*, 2005, **308**, 1424-1428.
86. J. Monod, J. Wyman and J. P. Changeux, *J Mol Biol*, 1965, **12**, 88-118.
87. D. E. Koshland, Jr., G. Nemethy and D. Filmer, *Biochemistry*, 1966, **5**, 365-385.
88. N. Popovych, S. R. Tzeng, M. Tonelli, R. H. Ebright and C. G. Kalodimos, *Proc Natl Acad Sci U S A*, 2009, **106**, 6927-6932.
89. S. Busby and R. H. Ebright, *J Mol Biol*, 1999, **293**, 199-213.
90. A. Bren and M. Eisenbach, *J Bacteriol*, 2000, **182**, 6865-6873.
91. D. Bray, M. D. Levin and C. J. Morton-Firth, *Nature*, 1998, **393**, 85-88.
92. S. H. Kim, W. Wang and K. K. Kim, *Proc Natl Acad Sci U S A*, 2002, **99**, 11611-11615.
93. M. Skoge, Y. Meir and N. S. Wingreen, *Phys Rev Lett*, 2011, **107**, 178101.
94. B. S. Khakh, X. Zhou, J. Sydes, J. J. Galligan and H. A. Lester, *Nature*, 2000, **406**, 405-410.
95. S. O. Marx, K. Ondrias and A. R. Marks, *Science*, 1998, **281**, 818-821.
96. C. Chan, A. J. George and J. Stark, *Proc Natl Acad Sci U S A*, 2001, **98**, 5758-5763.
97. T. A. Duke, N. Le Novere and D. Bray, *J Mol Biol*, 2001, **308**, 541-553.
98. T. Duke, *Philos Trans R Soc Lond B Biol Sci*, 2000, **355**, 529-538.
99. P. Tompa, *Trends Biochem Sci*, 2012, **37**, 509-516.
100. M. Schwalbe, V. Ozenne, S. Bibow, M. Jaremko, L. Jaremko, M. Gajda, M. R. Jensen, J. Biernat, S. Becker, E. Mandelkow, M. Zweckstetter and M. Blackledge, *Structure*, 2014, **22**, 238-249.
101. L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradovic and A. K. Dunker, *J Mol Biol*, 2002, **323**, 573-584.
102. P. Tompa, Z. Dosztanyi and I. Simon, *J Proteome Res*, 2006, **5**, 1996-2000.
103. P. E. Wright and H. J. Dyson, *Curr Opin Struct Biol*, 2009, **19**, 31-38.
104. U. Jakob, R. Kriwacki and V. N. Uversky, *Chem Rev*, 2014, **114**, 6779-6805.
105. D. M. Mitrea and R. W. Kriwacki, *FEBS Lett*, 2013, **587**, 1081-1088.
106. S. H. Lee, D. H. Kim, J. J. Han, E. J. Cha, J. E. Lim, Y. J. Cho, C. Lee and K. H. Han, *Curr Protein Pept Sci*, 2012, **13**, 34-54.
107. M. Fuxreiter, I. Simon, P. Friedrich and P. Tompa, *J Mol Biol*, 2004, **338**, 1015-1026.

108. I. Radhakrishnan, G. C. Perez-Alvarado, H. J. Dyson and P. E. Wright, *FEBS Lett*, 1998, **430**, 317-322.
109. J. A. Zitzewitz, B. Ibarra-Molero, D. R. Fishel, K. L. Terry and C. R. Matthews, *J Mol Biol*, 2000, **296**, 1105-1116.
110. P. Sarkar, C. Reichman, T. Saleh, R. B. Birge and C. G. Kalodimos, *Mol Cell*, 2007, **25**, 413-426.
111. J. F. Rippmann, S. Hobbie, C. Daiber, B. Guilliard, M. Bauer, J. Birk, H. Nar, P. Garin-Chesa, W. J. Rettig and A. Schnapp, *Cell Growth Differ*, 2000, **11**, 409-416.
112. M. D. Krasowski, E. J. Reschly and S. Ekins, *J Proteome Res*, 2008.
113. R. Kumar and E. B. Thompson, *Mol Endocrinol*, 2003, **17**, 1-10.
114. D. N. Lavery and I. J. McEwan, *Biochem. J.*, 2005, **391**, 449-464.
115. I. J. McEwan, *Mol Biosyst*, 2012, **8**, 82-90.
116. R. Kumar and E. B. Thompson, *Mol Cell Endocrinol*, 2012, **348**, 450-456.
117. A. S. Garza, S. H. Khan, C. M. Moure, D. P. Edwards and R. Kumar, *PLoS One*, 2012, **6**, e25875.
118. S. E. Wardell, S. C. Kwok, L. Sherman, R. S. Hodges and D. P. Edwards, *Mol Cell Biol*, 2005, **25**, 8792-8808.
119. R. Kumar, R. Betney, J. Li, E. B. Thompson and I. J. McEwan, *Biochemistry*, 2004, **43**, 3008-3013.
120. Y. X. Chen, J. T. Du, L. X. Zhou, X. H. Liu, Y. F. Zhao, H. Nakanishi and Y. M. Li, *Chem Biol*, 2006, **13**, 937-944.
121. A. M. Garza, S. H. Khan and R. Kumar, *Mol Cell Biol*, 2010, **30**, 220-230.
122. D. W. Leung and M. K. Rosen, *Proc Natl Acad Sci U S A*, 2005, **102**, 5685-5690.
123. T. Moldoveanu, A. V. Follis, R. W. Kriwacki and D. R. Green, *Trends Biochem Sci*, 2014, **39**, 101-111.
124. A. Shamas-Din, J. Kale, B. Leber and D. W. Andrews, *Cold Spring Harb Perspect Biol*, 2013, **5**, a008714.
125. A. V. Follis, J. E. Chipuk, J. C. Fisher, M. K. Yun, C. R. Grace, A. Nourse, K. Baran, L. Ou, L. Min, S. W. White, D. R. Green and R. W. Kriwacki, *Nat Chem Biol*, 2013, **9**, 163-168.
126. O. Vinogradova, J. Vaynberg, X. Kong, T. A. Haas, E. F. Plow and J. Qin, *Proc Natl Acad Sci U S A*, 2004, **101**, 4094-4099.
127. M. A. Pufall and B. J. Graves, *Annu Rev Cell Dev Biol*, 2002, **18**, 421-462.
128. B. Kobe and B. E. Kemp, *Nature*, 1999, **402**, 373-376.
129. B. Alberts, *Cell*, 1998, **92**, 291-294.
130. J. A. Marsh, H. Hernandez, Z. Hall, S. E. Ahnert, T. Perica, C. V. Robinson and S. A. Teichmann, *Cell*, 2013, **153**, 461-470.
131. W. A. Held, B. Ballou, S. Mizushima and M. Nomura, *J Biol Chem*, 1974, **249**, 3103-3111.
132. W. A. Held, S. Mizushima and M. Nomura, *J Biol Chem*, 1973, **248**, 5720-5730.
133. A. Bertagna, D. Toptygin, L. Brand and D. Barrick, *Biochem Soc Trans*, 2008, **36**, 157-166.
134. J. R. Williamson, *Nat Struct Biol*, 2000, **7**, 834-837.
135. S. C. Agalarov, G. Sridhar Prasad, P. M. Funke, C. D. Stout and J. R. Williamson, *Science*, 2000, **288**, 107-113.
136. A. P. Hinck, M. A. Markus, S. Huang, S. Grzesiek, I. Kustonovich, D. E. Draper and D. A. Torchia, *J Mol Biol*, 1997, **274**, 101-113.
137. L. Varani, S. I. Gunderson, I. W. Mattaj, L. E. Kay, D. Neuhaus and G. Varani, *Nat Struct Biol*, 2000, **7**, 329-335.
138. J. P. DiNitto and P. W. Huber, *J Mol Biol*, 2003, **330**, 979-992.
139. P. Tompa and P. Csermely, *FASEB J.*, 2004, **18**, 1169-1175.
140. F. A. Mulder, L. Bouakaz, A. Lundell, M. Venkataramana, A. Liljas, M. Akke and S. Sanyal, *Biochemistry*, 2004, **43**, 5930-5936.
141. M. Nishimura, T. Yoshida, M. Shirouzu, T. Terada, S. Kuramitsu, S. Yokoyama, T. Ohkubo and Y. Kobayashi, *J Mol Biol*, 2004, **344**, 1369-1383.
142. C. Haynes and L. M. Iakoucheva, *Nucleic Acids Res*, 2006, **34**, 305-312.
143. H. Hegyi, E. Schad and P. Tompa, *BMC Struct Biol*, 2007, **7**, 65.

144. A. Toth-Petroczy, C. J. Oldfield, I. Simon, Y. Takagi, A. K. Dunker, V. N. Uversky and M.-. Fuxreiter, *PLoS Comp. Biol.*, 2008, **4**, e1000243.
145. H. Clevers and R. Nusse, *Cell*, 2012, **149**, 1192-1205.
146. V. Katritch, V. Cherezov and R. C. Stevens, *Annu Rev Pharmacol Toxicol*, 2013, **53**, 531-556.
147. B. K. Kobilka, *Biochim Biophys Acta*, 2007, **1768**, 794-807.
148. J. P. Overington, B. Al-Lazikani and A. L. Hopkins, *Nat Rev Drug Discov*, 2006, **5**, 993-996.
149. S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara and B. K. Kobilka, *Nature*, 2011, **477**, 549-555.
150. A. C. Kruse, A. M. Ring, A. Manglik, J. Hu, K. Hu, K. Eitel, H. Hubner, E. Pardon, C. Valant, P. M. Sexton, A. Christopoulos, C. C. Felder, P. Gmeiner, J. Steyaert, W. I. Weis, K. C. Garcia, J. Wess and B. K. Kobilka, *Nature*, 2013, **504**, 101-106.
151. J. J. Tesmer, R. K. Sunahara, A. G. Gilman and S. R. Sprang, *Science*, 1997, **278**, 1907-1916.
152. H. R. Bourne, D. A. Sanders and F. McCormick, *Nature*, 1991, **349**, 117-127.
153. K. Rajagopal, R. J. Lefkowitz and H. A. Rockman, *J Clin Invest*, 2005, **115**, 2971-2974.
154. J. J. Liu, R. Horst, V. Katritch, R. C. Stevens and K. Wuthrich, *Science*, 2012, **335**, 1106-1110.
155. D. Dai, Q. Huang, R. Nussinov and B. Ma, *Biochim Biophys Acta*, 2014, **1844**, 1729-1740.
156. E. M. Jones and W. K. Surewicz, *Cell*, 2005, **121**, 63-72.
157. B. N. Kholodenko, *Eur J Biochem*, 2000, **267**, 1583-1588.
158. W. M. Oldham and H. E. Hamm, *Nat Rev Mol Cell Biol*, 2008, **9**, 60-71.
159. S. Lu, H. Jang, R. Nussinov and J. Zhang, *Sci Rep*, 2016, **6**, 21949.
160. S. B. Prusiner, *Proc Natl Acad Sci U S A*, 1998, **95**, 13363-13383.
161. R. Kageyama, Y. Masamizu and Y. Niwa, *Dev Dyn*, 2007, **236**, 1403-1409.
162. A. L. Horwich and J. S. Weissman, *Cell*, 1997, **89**, 499-510.
163. K. Pauwels, I. Van Molle, J. Tommassen and P. Van Gelder, *Mol Microbiol*, 2007, **64**, 917-922.
164. K. Pauwels, M. M. Sanchez del Pino, G. Feller and P. Van Gelder, *PLoS One*, 2012, **7**, e36999.
165. R. P. Bhattacharyya, A. Remenyi, B. J. Yeh and W. A. Lim, *Annu Rev Biochem*, 2006, **75**, 655-680.
166. B. T. Seet, I. Dikic, M. M. Zhou and T. Pawson, *Nat Rev Mol Cell Biol*, 2006, **7**, 473-483.
167. G. M. Fimia and P. Sassone-Corsi, *J Cell Sci*, 2001, **114**, 1971-1972.
168. L. Milanesi, M. Petrillo, L. Sepe, A. Boccia, N. D'Agostino, M. Passamano, S. Di Nardo, G. Tasco, R. Casadio and G. Paolella, *BMC Bioinformatics*, 2005, **6 Suppl 4**, S20.
169. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, *Science*, 2002, **298**, 1912-1934.
170. V. E. Prince and F. B. Pickett, *Nat Rev Genet*, 2002, **3**, 827-837.
171. R. R. Copley, L. Goodstadt and C. Ponting, *Curr Opin Genet Dev*, 2003, **13**, 623-628.
172. C. P. Ponting, J. Schultz, R. R. Copley, M. A. Andrade and P. Bork, *Adv Protein Chem*, 2000, **54**, 185-244.
173. L. Patthy, *Gene*, 1999, **238**, 103-114.
174. H. Tordai, A. Nagy, K. Farkas, L. Banyai and L. Patthy, *Febs J*, 2005, **272**, 5064-5078.
175. T. Pawson and P. Nash, *Science*, 2003, **300**, 445-452.
176. T. Pawson, M. Raina and P. Nash, *FEBS Lett*, 2002, **513**, 2-10.
177. W. H. Hudson and E. A. Ortlund, *Nat Rev Mol Cell Biol*, 2014, **15**, 749-760.
178. S. Karkkainen, M. Hiipakka, J. H. Wang, I. Kleino, M. Vaha-Jaakkola, G. H. Renkema, M. Liss, R. Wagner and K. Saksela, *EMBO Rep*, 2006, **7**, 186-191.
179. H. Huang, L. Li, C. Wu, D. Schibli, K. Colwill, S. Ma, C. Li, P. Roy, K. Ho, Z. Songyang, T. Pawson, Y. Gao and S. S. Li, *Mol Cell Proteomics*, 2008, **7**, 768-784.
180. A. K. Bjorklund, D. Ekman and A. Elofsson, *PLoS Comput Biol*, 2006, **2**, e114.
181. C. Vogel, C. Berzuini, M. Bashton, J. Gough and S. A. Teichmann, *J Mol Biol*, 2004, **336**, 809-823.
182. J. Soding and A. N. Lupas, *Bioessays*, 2003, **25**, 837-846.

183. C. Vogel, M. Bashton, N. D. Kerrison, C. Chothia and S. A. Teichmann, *Curr Opin Struct Biol*, 2004, **14**, 208-216.
184. A. M. Bilwes, L. A. Alex, B. R. Crane and M. I. Simon, *Cell*, 1999, **96**, 131-141.
185. G. J. Davis, W. F. Bosron, C. L. Stone, K. Owusu-Dekyi and T. D. Hurley, *J Biol Chem*, 1996, **271**, 17057-17061.
186. A. Roll-Mecak, B. S. Shin, T. E. Dever and S. K. Burley, *Trends Biochem Sci*, 2001, **26**, 705-709.
187. H. Song, M. R. Parsons, S. Rowsell, G. Leonard and S. E. Phillips, *J Mol Biol*, 1999, **285**, 1245-1256.
188. M. Fuxreiter, P. Tompa and I. Simon, *Bioinformatics*, 2007, **23**, 950-956.
189. F. Diella, N. Haslam, C. Chica, A. Budd, S. Michael, N. P. Brown, G. Trave and T. J. Gibson, *Front Biosci*, 2008, **13**, 6580-6603.
190. K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson and N. E. Davey, *Chem Rev*, 2014, **114**, 6733-6778.
191. H. Dinkel, K. Van Roey, S. Michael, N. E. Davey, R. J. Weatheritt, D. Born, T. Speck, D. Kruger, G. Grebnev, M. Kuban, M. Strumillo, B. Uyar, A. Budd, B. Altenberg, M. Seiler, L. B. Chemes, J. Glavina, I. E. Sanchez, F. Diella and T. J. Gibson, *Nucleic Acids Res*, 2014, **42**, D259-266.
192. H. Dinkel, K. Van Roey, S. Michael, M. Kumar, B. Uyar, B. Altenberg, V. Milchevskaya, M. Schneider, H. Kuhn, A. Behrendt, S. L. Dahl, V. Damerell, S. Diebel, S. Kalman, S. Klein, A. C. Knudsen, C. Mader, S. Merrill, A. Staudt, V. Thiel, L. Welti, N. E. Davey, F. Diella and T. J. Gibson, *Nucleic Acids Res*, 2016, **44**, D294-300.
193. P. Tompa, N. E. Davey, T. J. Gibson and M. M. Babu, *Mol. Cell*, 2014, **55**, 161-169.
194. A. Aouacheria, V. Rech de Laval, C. Combet and J. M. Hardwick, *Trends Cell Biol*, 2012, **23**, 103-111.
195. M. Hundertmark and D. K. Hincha, *BMC Genomics*, 2008, **9**, 118.
196. N. F. Endres, K. Engel, R. Das, E. Kovacs and J. Kuriyan, *Curr Opin Struct Biol*, 2011, **21**, 777-784.
197. T. Mittag, S. Orlicky, W. Y. Choy, X. Tang, H. Lin, F. Sicheri, L. E. Kay, M. Tyers and J. D. Forman-Kay, *Proc Natl Acad Sci U S A*, 2008, **105**, 17772-17777.
198. Y. Ono and H. Sorimachi, *Biochim Biophys Acta*, 2012, **1824**, 224-236.
199. D. A. Potter, J. S. Tirnauer, R. Janssen, D. E. Croall, C. N. Hughes, K. A. Fiacco, J. W. Mier, M. Maki and I. M. Herman, *J Cell Biol*, 1998, **141**, 647-662.
200. I. Richard, O. Broux, V. Allamand, F. Fougerousse, N. Chiannilkulchai, N. Bourg, L. Brenguier, C. Devaud, P. Pasturaud, C. Roudaut and et al., *Cell*, 1995, **81**, 27-40.
201. R. A. Hanna, R. L. Campbell and P. L. Davies, *Nature*, 2008, **456**, 409-412.
202. J. F. Harper, M. R. Sussman, G. E. Schaller, C. Putnam-Evans, H. Charbonneau and A. C. Harmon, *Science*, 1991, **252**, 951-954.
203. F. Chiti and C. M. Dobson, *Annu Rev Biochem*, 2006, **75**, 333-366.
204. R. B. Wickner, H. K. Edskes, E. D. Ross, M. M. Pierce, F. Shewmaker, U. Baxa and A. Brachmann, *Cold Spring Harb Symp Quant Biol*, 2004, **69**, 489-496.
205. K. Si, S. Lindquist and E. R. Kandel, *Cell*, 2003, **115**, 879-891.
206. M. F. Tuite, *Cell*, 2000, **100**, 289-292.
207. R. B. Wickner, K. L. Taylor, H. K. Edskes and M. L. Maddelein, *Curr Biol*, 2000, **10**, R335-337.
208. U. Baxa, P. D. Ross, R. B. Wickner and A. C. Steven, *J Mol Biol*, 2004, **339**, 259-264.
209. A. Balguerie, S. Dos Reis, C. Ritter, S. Chaignepain, B. Coulary-Salin, V. Forge, K. Bathany, I. Lascu, J. M. Schmitter, R. Riek and S. J. Saupe, *EMBO J*, 2003, **22**, 2071-2081.
210. F. Shewmaker, R. B. Wickner and R. Tycko, *Proc Natl Acad Sci U S A*, 2006, **103**, 19754-19759.
211. L. Li and S. Lindquist, *Science*, 2000, **287**, 661-664.
212. S. Alberti, R. Halfmann, O. King, A. Kapila and S. Lindquist, *Cell*, 2009, **137**, 146-158.
213. J. M. Baker, R. P. Hudson, V. Kanelis, W. Y. Choy, P. H. Thibodeau, P. J. Thomas and J. D. Forman-Kay, *Nat Struct Mol Biol*, 2007, **14**, 738-745.
214. J. R. Riordan, *Annu Rev Biochem*, 2008, **77**, 701-726.

215. J. R. Riordan, J. M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic, J. L. Chou, M. L. Drumm, M. C. Iannuzzi, F. S. Collins and L.-C. Tsui, *Science*, 1989, **245**, 1066-1073.
216. L. S. Ostedgaard, O. Baldursson, D. W. Vermeer, M. J. Welsh and A. D. Robertson, *Proc Natl Acad Sci U S A*, 2000, **97**, 5657-5662.
217. Z. Bozoky, M. Krzeminski, P. A. Chong and J. D. Forman-Kay, *FEBS J*, 2013, **280**, 4407-4416.
218. J. Becker, *Nat Biotechnol*, 2004, **22**, 15-18.
219. A. M. Spiegel, L. S. Weinstein and A. Shenker, *J Clin Invest*, 1993, **92**, 1119-1125.
220. M. Pajkos, B. Meszaros, I. Simon and Z. Dosztanyi, *Mol Biosyst*, 2012, **8**, 296-307.
221. P. Imming, C. Sinning and A. Meyer, *Nat Rev Drug Discov*, 2006, **5**, 821-834.
222. P. A. Muller and K. H. Vousden, *Nat Cell Biol*, 2013, **15**, 2-8.
223. N. Rivlin, R. Brosh, M. Oren and V. Rotter, *Genes Cancer*, 2011, **2**, 466-474.
224. I. A. Prior, P. D. Lewis and C. Mattos, *Cancer Res*, 2012, **72**, 2457-2467.
225. P. Campbell, P. E. Morton, T. Takeichi, A. Salam, N. Roberts, L. E. Proudfoot, J. E. Mellerio, K. Aminu, C. Wellington, S. N. Patil, M. Akiyama, L. Liu, J. R. McMillan, S. Aristodemou, A. Ishida-Yamamoto, A. Abdul-Wahab, G. Petrof, K. Fong, S. Harnchoowong, K. L. Stone, J. I. Harper, W. H. McLean, M. A. Simpson, M. Parsons and J. A. McGrath, *J Invest Dermatol*, 2014, **134**, 2570-2578.
226. M. C. Chartier-Harlin, F. Crawford, H. Houlden, A. Warren, D. Hughes, L. Fidani, A. Goate, M. Rossor, P. Roques, J. Hardy and et al., *Nature*, 1991, **353**, 844-846.
227. A. Goate, M. C. Chartier-Harlin, M. Mullan, J. Brown, F. Crawford, L. Fidani, L. Giuffra, A. Haynes, N. Irving, L. James and et al., *Nature*, 1991, **349**, 704-706.
228. A. Al-Chalabi and P. N. Leigh, *Curr Opin Neurol*, 2000, **13**, 397-405.
229. R. A. Conwit, *J Neurol Sci*, 2006, **251**, 1-2.
230. R. W. Carrell and D. A. Lomas, *Lancet*, 1997, **350**, 134-138.
231. A. Aguzzi and A. M. Calella, *Physiol Rev*, 2009, **89**, 1105-1152.
232. P. Brundin, R. Melki and R. Kopito, *Nat Rev Mol Cell Biol*, 2010, **11**, 301-307.
233. P. R. Elliott, D. A. Lomas, R. W. Carrell and J. P. Abrahams, *Nat Struct Biol*, 1996, **3**, 676-681.
234. T. Vavouri, J. I. Semple, R. Garcia-Verdugo and B. Lehner, *Cell*, 2009, **138**, 198-208.
235. T. J. Gibson, M. Seiler and R. A. Veitia, *Nat Methods*, 2013, **10**, 715-721.
236. M. A. Pujana, J. D. Han, L. M. Starita, K. N. Stevens, M. Tewari, J. S. Ahn, G. Rennert, V. Moreno, T. Kirchhoff, B. Gold, V. Assmann, W. M. Elshamy, J. F. Rual, D. Levine, L. S. Rozek, R. S. Gelman, K. C. Gunsalus, R. A. Greenberg, B. Sobhian, N. Bertin, K. Venkatesan, N. Ayivi-Guedehoussou, X. Sole, P. Hernandez, C. Lazaro, K. L. Nathanson, B. L. Weber, M. E. Cusick, D. E. Hill, K. Offit, D. M. Livingston, S. B. Gruber, J. D. Parvin and M. Vidal, *Nat Genet*, 2007, **39**, 1338-1349.
237. M. Vidal, M. E. Cusick and A. L. Barabasi, *Cell*, 2011, **144**, 986-998.
238. V. N. Uversky, C. J. Oldfield and A. K. Dunker, *Annu Rev Biophys*, 2008, **37**, 215-246.
239. C. Weissmann, *Nat Rev Microbiol*, 2004, **2**, 861-871.
240. S. Lindquist, S. Krobitsch, L. Li and N. Sondheimer, *Philos Trans R Soc Lond B Biol Sci*, 2001, **356**, 169-176.
241. R. B. Wickner, F. P. Shewmaker, D. A. Bateman, H. K. Edskes, A. Gorkovskiy, Y. Dayani and E. E. Bezsonov, *Microbiol Mol Biol Rev*, 2015, **79**, 1-17.
242. N. Vishveshwara, M. E. Bradley and S. W. Liebman, *Mol Microbiol*, 2009, **73**, 1101-1114.
243. S. Treusch and S. Lindquist, *J Cell Biol*, 2012, **197**, 369-379.
244. A. Sydow and E. M. Mandelkow, *Neurodegener Dis*, 2010, **7**, 28-31.
245. N. Fernandez-Borges, H. Erana, S. R. Elezgarai, C. Harrathi, M. Gayosso and J. Castilla, *Int J Cell Biol*, 2013, **2013**, 583498.
246. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res*, 2000, **28**, 235-242.
247. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert and J. Hajdu, *Nature*, 2000, **406**, 752-757.

248. M. M. Seibert, T. Ekeberg, F. R. Maia, M. Svenda, J. Andreasson, O. Jonsson, D. Odic, B. Iwan, A. Rocker, D. Westphal, M. Hantke, D. P. DePonte, A. Barty, J. Schulz, L. Gumprecht, N. Coppola, A. Aquila, M. Liang, T. A. White, A. Martin, C. Caleman, S. Stern, C. Abergel, V. Seltzer, J. M. Claverie, C. Bostedt, J. D. Bozek, S. Boutet, A. A. Miahnahri, M. Messerschmidt, J. Krzywinski, G. Williams, K. O. Hodgson, M. J. Bogan, C. Y. Hampton, R. G. Sierra, D. Starodub, I. Andersson, S. Bajt, M. Barthelmess, J. C. Spence, P. Fromme, U. Weierstall, R. Kirian, M. Hunter, R. B. Doak, S. Marchesini, S. P. Hau-Riege, M. Frank, R. L. Shoeman, L. Lomb, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, C. Schmidt, L. Foucar, N. Kimmel, P. Holl, B. Rudek, B. Erk, A. Homke, C. Reich, D. Pietschner, G. Weidenspointner, L. Struder, G. Hauser, H. Gorke, J. Ullrich, I. Schlichting, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Kuhnel, R. Andritschke, C. D. Schroter, F. Krasniqi, M. Bott, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, H. N. Chapman and J. Hajdu, *Nature*, 2011, **470**, 78-81.

249. T. Ekeberg, M. Svenda, C. Abergel, F. R. Maia, V. Seltzer, J. M. Claverie, M. Hantke, O. Jonsson, C. Nettelblad, G. van der Schot, M. Liang, D. P. DePonte, A. Barty, M. M. Seibert, B. Iwan, I. Andersson, N. D. Loh, A. V. Martin, H. Chapman, C. Bostedt, J. D. Bozek, K. R. Ferguson, J. Krzywinski, S. W. Epp, D. Rolles, A. Rudenko, R. Hartmann, N. Kimmel and J. Hajdu, *Phys Rev Lett*, 2015, **114**, 098102.

250. H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Homke, C. Reich, D. Pietschner, L. Struder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Kuhnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jonsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C. D. Schroter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson and J. C. Spence, *Nature*, 2011, **470**, 73-77.

251. W. Liu, D. Wacker, C. Gati, G. W. Han, D. James, D. Wang, G. Nelson, U. Weierstall, V. Katritch, A. Barty, N. A. Zatsepin, D. Li, M. Messerschmidt, S. Boutet, G. J. Williams, J. E. Koglin, M. M. Seibert, C. Wang, S. T. Shah, S. Basu, R. Fromme, C. Kupitz, K. N. Rendek, I. Grotjohann, P. Fromme, R. A. Kirian, K. R. Beyerlein, T. A. White, H. N. Chapman, M. Caffrey, J. C. Spence, R. C. Stevens and V. Cherezov, *Science*, 2013, **342**, 1521-1524.

252. K. H. Sze, Q. Wu, H. S. Tse and G. Zhu, *Top Curr Chem*, 2012, **326**, 215-242.

253. R. Sprangers and L. E. Kay, *Nature*, 2007, **445**, 618-622.

254. K. Sugase, H. J. Dyson and P. E. Wright, *Nature*, 2007, **447**, 1021-1025.

255. M. J. Knight, A. J. Pell, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, T. Herrmann, L. Emsley and G. Pintacuda, *Proc Natl Acad Sci U S A*, 2012, **109**, 11095-11100.

256. N. Kodera, D. Yamamoto, R. Ishikawa and T. Ando, *Nature*, 2010, **468**, 72-76.

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

258. P. Bernado and D. I. Svergun, *Mol Biosyst*, 2012, **8**, 151-167.

259. E. Binshtein and M. D. Ohi, *Biochemistry*, 2015, **54**, 3133-3141.

260. A. Schreiber, F. Stengel, Z. Zhang, R. I. Enchev, E. H. Kong, E. P. Morris, C. V. Robinson, P. C. da Fonseca and D. Barford, *Nature*, 2011, **470**, 227-232.

261. A. Abragam and M. Goldman, *Reports on Progress in Physics*, 1976, **41**, 395–467.

262. C. Y. Cheng and S. Han, *Annu Rev Phys Chem*, 2013, **64**, 507-532.

263. D. M. Korzhnev, R. M. Vernon, T. L. Religa, A. L. Hansen, D. Baker, A. R. Fersht and L. E. Kay, *J Am Chem Soc*, 2011, **133**, 10974-10982.

264. L. E. Kay, *J Magn Reson*, 2005, **173**, 193-207.
265. D. M. Korzhnev and L. E. Kay, *Acc Chem Res*, 2008, **41**, 442-451.
266. R. Tycko, *Annu Rev Phys Chem*, 2011, **62**, 279-299.
267. Y. Ito and P. Selenko, *Curr Opin Struct Biol*, 2010, **20**, 640-648.
268. D. D. Boehr, D. McElheny, H. J. Dyson and P. E. Wright, *Science*, 2006, **313**, 1638-1642.
269. M. R. Sawaya and J. Kraut, *Biochemistry*, 1997, **36**, 586-603.
270. R. P. Venkitakrishnan, E. Zaborowski, D. McElheny, S. J. Benkovic, H. J. Dyson and P. E. Wright, *Biochemistry*, 2004, **43**, 16046-16055.
271. H. J. Berendsen and S. Hayward, *Curr Opin Struct Biol*, 2000, **10**, 165-169.
272. T. H. Rod, J. L. Radkiewicz and C. L. Brooks, 3rd, *Proc Natl Acad Sci U S A*, 2003, **100**, 6980-6985.
273. C. M. Petit, J. Zhang, P. J. Sapienza, E. J. Fuentes and A. L. Lee, *Proc Natl Acad Sci U S A*, 2009, **106**, 18249-18254.
274. I. Radhakrishnan, G. C. Perez-Alvarado, D. Parker, H. J. Dyson, M. R. Montminy and P. E. Wright, *Cell*, 1997, **91**, 741-752.
275. S. J. Opella, *J Magn Reson*, 2015, **253**, 129-137.
276. M. E. Ward, L. S. Brown and V. Ladizhansky, *J Magn Reson*, 2015, **253**, 119-128.
277. D. Yamamoto, T. Uchihashi, N. Kodera, H. Yamashita, S. Nishikori, T. Ogura, M. Shibata and T. Ando, *Methods Enzymol*, 2010, **475**, 541-564.
278. S. Kalinin, T. Peulen, S. Sindbert, P. J. Rothwell, S. Berger, T. Restle, R. S. Goody, H. Gohlke and C. A. Seidel, *Nat Methods*, 2012, **9**, 1218-1225.
279. A. C. Ferreon, Y. Gambin, E. A. Lemke and A. A. Deniz, *Proc Natl Acad Sci U S A*, 2009, **106**, 5645-5650.
280. F. Huang, S. Rajagopalan, G. Settanni, R. J. Marsh, D. A. Armoogum, N. Nicolaou, A. J. Bain, E. Lerner, E. Haas, L. Ying and A. R. Fersht, *Proc Natl Acad Sci U S A*, 2009.
281. J. J. McCann, L. Zheng, D. Rohrbeck, S. Felekyan, R. Kuhnemuth, R. B. Sutton, C. A. Seidel and M. E. Bowen, *Proc Natl Acad Sci U S A*, 2012, **109**, 15775-15780.
282. P. Tompa, *Nat Chem Biol*, 2012, **8**, 597-600.
283. M. Wells, H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, D. I. Svergun, M. Blackledge and A. R. Fersht, *Proc Natl Acad Sci U S A*, 2008, **105**, 5762-5767.
284. S. Yang, L. Blachowicz, L. Makowski and B. Roux, *Proc Natl Acad Sci U S A*, 2010, **107**, 15757-15762.
285. D. Elmlund and H. Elmlund, *Annu Rev Biochem*, 2015, **e-pub ahead of publication**.
286. X. C. Bai, I. S. Fernandez, G. McMullan and S. H. Scheres, *Elife*, 2013, **2**, e00461.
287. G. Yu, D. Zhang, F. Guo, M. Tan, X. Jiang and W. Jiang, *PLoS One*, 2013, **8**, e59817.
288. D. Luque, J. Gomez-Blanco, D. Garriga, A. F. Brilot, J. M. Gonzalez, W. M. Havens, J. L. Carrascosa, B. L. Trus, N. Verdaguer, S. A. Ghabrial and J. R. Caston, *Proc Natl Acad Sci U S A*, 2014, **111**, 7641-7646.
289. I. Gutsche, A. Desfosses, G. Effantin, W. L. Ling, M. Haupt, R. W. Ruigrok, C. Sachse and G. Schoehn, *Science*, 2015, **348**, 704-707.
290. Z. Yan, X. C. Bai, C. Yan, J. Wu, Z. Li, T. Xie, W. Peng, C. C. Yin, X. Li, S. H. Scheres, Y. Shi and N. Yan, *Nature*, 2015, **517**, 50-55.
291. R. Zalk, O. B. Clarke, A. des Georges, R. A. Grassucci, S. Reiken, F. Mancia, W. A. Hendrickson, J. Frank and A. R. Marks, *Nature*, 2015, **517**, 44-49.
292. R. G. Efremov, A. Leitner, R. Aebersold and S. Raunser, *Nature*, 2015, **517**, 39-43.
293. H. Elmlund, V. Baraznenok, T. Linder, Z. Szilagyi, R. Rofougaran, A. Hofer, H. Hebert, M. Lindahl and C. M. Gustafsson, *Structure*, 2009, **17**, 1442-1452.
294. K. L. Tsai, C. Tomomori-Sato, S. Sato, R. C. Conaway, J. W. Conaway and F. J. Asturias, *Cell*, 2014, **157**, 1430-1444.
295. K. Murakami, H. Elmlund, N. Kalisman, D. A. Bushnell, C. M. Adams, M. Azubel, D. Elmlund, Y. Levi-Kalisman, X. Liu, B. J. Gibbons, M. Levitt and R. D. Kornberg, *Science*, 2013, **342**, 1238724.

296. A. Bartesaghi, D. Matthies, S. Banerjee, A. Merk and S. Subramaniam, *Proc Natl Acad Sci U S A*, 2014, **111**, 11709-11714.
297. H. Gao, M. Valle, M. Ehrenberg and J. Frank, *J Struct Biol*, 2004, **147**, 283-290.
298. D. Veesler, T. S. Ng, A. K. Sendamarai, B. J. Eilers, C. M. Lawrence, S. M. Lok, M. J. Young, J. E. Johnson and C. Y. Fu, *Proc Natl Acad Sci U S A*, 2013, **110**, 5504-5509.
299. H. Viadiu, O. Stemmann, M. W. Kirschner and T. Walz, *Nat Struct Mol Biol*, 2005, **12**, 552-553.
300. P. Yi, Z. Wang, Q. Feng, G. D. Pintilie, C. E. Foulds, R. B. Lanz, S. J. Ludtke, M. F. Schmid, W. Chiu and B. W. O'Malley, *Mol Cell*, 2015, **57**, 1047-1058.
301. E. van Duijn, A. Barendregt, S. Synowsky, C. Versluis and A. J. Heck, *J Am Chem Soc*, 2009, **131**, 1452-1459.
302. J. Marcoux, S. C. Wang, A. Politis, E. Reading, J. Ma, P. C. Biggin, M. Zhou, H. Tao, Q. Zhang, G. Chang, N. Morgner and C. V. Robinson, *Proc Natl Acad Sci U S A*, 2013, **110**, 9704-9709.
303. A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck and R. Aebersold, *Mol Cell Proteomics*, 2010, **9**, 1634-1649.
304. L. Fischer, Z. A. Chen and J. Rappsilber, *J Proteomics*, 2013, **88**, 120-128.
305. M. Suchanek, A. Radzikowska and C. Thiele, *Nature Methods*, 2005, **2**, 261–268.
306. A. Belsom, M. Schneider, L. Fischer, O. Brock and J. Rappsilber, *Mol Cell Proteomics*, 2016, **15**, 1105-1116.
307. K. Lars, P. Jingxi and L. Yu-Hong, *Chem. Soc. Rev.*, 2011, **40**, 1224–1234.
308. M. J. Chalmers, S. A. Busby, B. D. Pascal, G. M. West and P. R. Griffin, *Expert Rev. Proteomics*, 2011, **8**, 43-59.
309. A. Brock, *Protein Expr. Purif.*, 2012, **84**, 19–37.
310. D. Balasubramaniam and E. A. Komives, *Biochim. Biophys. Acta*, 2013, **1834**, 1202–1209.
311. L. Konermann, B. B. Stocks, Y. Pan and X. Tong, *Mass Spectrom. Rev.*, 2010, **29**, 651-667.
312. L. Wang and M. R. Chance, *Anal Chem.*, 2011, **83**, 7234-7241.
313. S. Gupta, V. Bavro, R. D'Mello, S. J. Tucker, C. Vénien-Bryan and C. M. R., *Structure*, 2010, **18**, 839-846.
314. L. N. Johnson, *FASEB J*, 1992, **6**, 2274-2282.
315. C. L. Brooks and W. Gu, *Mol Cell*, 2006, **21**, 307-315.
316. N. Gupta, S. Tanner, N. Jaitly, J. N. Adkins, M. Lipton, R. Edwards, M. Romine, A. Osterman, V. Bafna, R. D. Smith and P. A. Pevzner, *Genome Res*, 2007, **17**, 1362-1377.
317. J. V. Olsen and M. Mann, *Mol Cell Proteomics*, 2013, **12**, 3444-3452.
318. C. Choudhary and M. Mann, *Nat Rev Mol Cell Biol*, 2010, **11**, 427-439.
319. O. N. Jensen, *Curr Opin Chem Biol*, 2004, **8**, 33-41.
320. C. R. Landry, E. D. Levy, D. Abd Rabbo, K. Tarassov and S. W. Michnick, *Cell*, 2013, **155**, 983-989.
321. C. R. Landry, E. D. Levy and S. W. Michnick, *Trends Genet*, 2009, **25**, 193-197.
322. N. C. Tedford, A. B. Hall, J. R. Graham, C. E. Murphy, N. F. Gordon and J. A. Radding, *Proteomics*, 2009, **9**, 1469-1487.
323. S. Morandell, T. Stasyk, S. Skvortsov, S. Ascher and L. A. Huber, *Proteomics*, 2008, **8**, 4383-4401.
324. F. X. Theillet, A. Binolfi, T. Frembgen-Kesner, K. Hingorani, M. Sarkar, C. Kyne, C. Li, P. B. Crowley, L. Gierasch, G. J. Pielak, A. H. Elcock, A. Gershenson and P. Selenko, *Chem Rev*, 2014, **114**, 6661-6714.
325. R. J. Ellis, *Trends Biochem Sci*, 2001, **26**, 597-604.
326. J. M. Mouillon, S. K. Eriksson and P. Harryson, *Plant Physiol*, 2008, **148**, 1925-1937.
327. C. S. Szasz, A. Alexa, K. Toth, M. Rakacs, J. Langowski and P. Tompa, *Biochemistry*, 2011, **50**, 5834-5844.
328. F. X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb and P. Selenko, *Nature*, 2016, **530**, 45-50.
329. T. Rolland, M. Tasan, B. Charloteaux, S. J. Pevzner, Q. Zhong, N. Sahni, S. Yi, I. Lemmens, C. Fontanillo, R. Mosca, A. Kamburov, S. D. Ghiassian, X. Yang, L. Ghamsari, D. Balcha, B. E.

Begg, P. Braun, M. Brehme, M. P. Broly, A. R. Carvunis, D. Convery-Zupan, R. Corominas, J. Coulombe-Huntington, E. Dann, M. Dreze, A. Dricot, C. Fan, E. Franzosa, F. Gebreab, B. J. Gutierrez, M. F. Hardy, M. Jin, S. Kang, R. Kiros, G. N. Lin, K. Luck, A. MacWilliams, J. Menche, R. R. Murray, A. Palagi, M. M. Poulin, X. Rambout, J. Rasla, P. Reichert, V. Romero, E. Ruyssinck, J. M. Sahalie, A. Scholz, A. A. Shah, A. Sharma, Y. Shen, K. Spirohn, S. Tam, A. O. Tejeda, S. A. Trigg, J. C. Twizere, K. Vega, J. Walsh, M. E. Cusick, Y. Xia, A. L. Barabasi, L. M. Iakoucheva, P. Aloy, J. De Las Rivas, J. Tavernier, M. A. Calderwood, D. E. Hill, T. Hao, F. P. Roth and M. Vidal, *Cell*, 2014, **159**, 1212-1226.

330. M. P. Stumpf, T. Thorne, E. de Silva, R. Stewart, H. J. An, M. Lappe and C. Wiuf, *Proc Natl Acad Sci U S A*, 2008, **105**, 6959-6964.
331. T. Kind, M. Scholz and O. Fiehn, *PLoS One*, 2009, **4**, e5440.
332. D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucleic Acids Res*, 2013, **41**, D801-807.
333. P. Beltrao, V. Albanese, L. R. Kenner, D. L. Swaney, A. Burlingame, J. Villen, W. A. Lim, J. S. Fraser, J. Frydman and N. J. Krogan, *Cell*, 2012, **150**, 413-425.
334. J. F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth and M. Vidal, *Nature*, 2005, **437**, 1173-1178.
335. M. Babu, J. Vlasblom, S. Pu, X. Guo, C. Graham, B. D. Bean, H. E. Burston, F. J. Vizeacoumar, J. Snider, S. Phanse, V. Fong, Y. Y. Tam, M. Davey, O. Hnatshak, N. Bajaj, S. Chandran, T. Punna, C. Christopolous, V. Wong, A. Yu, G. Zhong, J. Li, I. Stagljar, E. Conibear, S. J. Wodak, A. Emili and J. F. Greenblatt, *Nature*, 2012, **489**, 585-589.
336. L. Giot, J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L. Hao, C. E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. A. Stanyon, R. L. Finley, Jr., K. P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R. A. Shimkets, M. P. McKenna, J. Chant and J. M. Rothberg, *Science*, 2003, **302**, 1727-1736.
337. S. Li, C. M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, P. O. Vidalain, J. D. Han, A. Chesneau, T. Hao, D. S. Goldberg, N. Li, M. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari, S. L. Wong, L. V. Zhang, G. F. Berriz, L. Jacotot, P. Vaglio, J. Reboul, T. Hirozane-Kishikawa, Q. Li, H. W. Gabel, A. Elewa, B. Baumgartner, D. J. Rose, H. Yu, S. Bosak, R. Sequerra, A. Fraser, S. E. Mango, W. M. Saxton, S. Strome, S. Van Den Heuvel, F. Piano, J. Vandenhaute, C. Sardet, M. Gerstein, L. Doucette-Stamm, K. C. Gunsalus, J. W. Harper, M. E. Cusick, F. P. Roth, D. E. Hill and M. Vidal, *Science*, 2004, **303**, 540-543.
338. G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann and B. Seraphin, *Nat Biotechnol*, 1999, **17**, 1030-1032.
339. S. Fields and O. Song, *Nature*, 1989, **340**, 245-246.
340. I. Lemmens, S. Lievens and J. Tavernier, *Methods Mol Biol*, 2015, **1278**, 447-455.
341. S. W. Michnick, P. H. Ear, C. Landry, M. K. Malleshaiah and V. Messier, *Methods Mol Biol*, 2011, **756**, 395-425.
342. O. Soderberg, M. Gullberg, M. Jarvius, K. Ridderstrale, K. J. Leuchowius, J. Jarvius, K. Wester, P. Hydbring, F. Bahram, L. G. Larsson and U. Landegren, *Nat Methods*, 2006, **3**, 995-1000.
343. L. Licata, L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A. P. Nardozza, E. Santonico, L. Castagnoli and G. Cesareni, *Nucleic Acids Res*, 2012, **40**, D857-861.

344. S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R. C. Lovering, B. Meldal, A. N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni and H. Hermjakob, *Nucleic Acids Res*, 2014, **42**, D358-363.
345. S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, C. Jandrasits, R. C. Jimenez, J. Khadake, U. Mahadevan, P. Masson, I. Pedruzzi, E. Pfeiffenberger, P. Porras, A. Raghunath, B. Roechert, S. Orchard and H. Hermjakob, *Nucleic Acids Res*, 2012, **40**, D841-846.
346. O. Rozenblatt-Rosen, R. C. Deo, M. Padi, G. Adelmant, M. A. Calderwood, T. Rolland, M. Grace, A. Dricot, M. Askenazi, M. Tavares, S. J. Pevzner, F. Abderazzaq, D. Byrdsong, A. R. Carvunis, A. A. Chen, J. Cheng, M. Correll, M. Duarte, C. Fan, M. C. Feltkamp, S. B. Ficarro, R. Franchi, B. K. Garg, N. Gulbahce, T. Hao, A. M. Holthaus, R. James, A. Korkhin, L. Litovchick, J. C. Mar, T. R. Pak, S. Rabello, R. Rubio, Y. Shen, S. Singh, J. M. Spangle, M. Tasan, S. Wanamaker, J. T. Webber, J. Roecklein-Canfield, E. Johannsen, A. L. Barabasi, R. Beroukhim, E. Kieff, M. E. Cusick, D. E. Hill, K. Munger, J. A. Marto, J. Quackenbush, F. P. Roth, J. A. DeCaprio and M. Vidal, *Nature*, 2012, **487**, 491-495.
347. W. Borcherds, F. X. Theillet, A. Katzer, A. Finzel, K. M. Mishall, A. T. Powell, H. Wu, W. Manieri, C. Dieterich, P. Selenko, A. Loewer and G. W. Daughdrill, *Nat Chem Biol*, 2014, **10**, 1000-1002.
348. J. J. Sakon and K. R. Weninger, *Nat Methods*, 2010, **7**, 203-205.
349. A. Rigort, F. J. Bauerlein, E. Villa, M. Eibauer, T. Laugks, W. Baumeister and J. M. Plitzko, *Proc Natl Acad Sci U S A*, 2012, **109**, 4449-4454.
350. Y. Feng, G. De Franceschi, A. Kahraman, M. Soste, A. Melnik, P. J. Boersema, P. P. de Laureto, Y. Nikolaev, A. P. Oliveira and P. Picotti, *Nat Biotechnol*, 2014, **32**, 1036-1044.
351. C. Dinant, M. E. van Royen, W. Vermeulen and A. B. Houtsmuller, *J Microsc*, 2008, **231**, 97-104.
352. I. C. Felli, L. Gonnelli and R. Pierattelli, *Nat Protoc*, 2014, **9**, 2005-2016.
353. D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Walchli, B. O. Smith, M. Shirakawa, P. Guntert and Y. Ito, *Nature*, 2009, **458**, 102-105.
354. P. Selenko, Z. Serber, B. Gadea, J. Ruderman and G. Wagner, *Proc Natl Acad Sci U S A*, 2006, **103**, 11904-11909.
355. J. F. Bodart, J. M. Wieruszeski, L. Amniai, A. Leroy, I. Landrieu, A. Rousseau-Lescuyer, J. P. Vilain and G. Lippens, *J Magn Reson*, 2008, **192**, 252-257.
356. B. C. McNulty, G. B. Young and G. J. Pielak, *J Mol Biol*, 2006, **355**, 893-897.
357. A. Binolfi, F. X. Theillet and P. Selenko, *Biochem Soc Trans*, 2012, **40**, 950-954.
358. K. H. Vousden and C. Prives, *Cell*, 2009, **137**, 413-431.
359. P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine and N. P. Pavletich, *Science*, 1996, **274**, 948-953.
360. H. Lee, K. H. Mok, R. Muhandiram, K. H. Park, J. E. Suk, D. H. Kim, J. Chang, Y. C. Sung, K. Y. Choi and K. H. Han, *J Biol Chem*, 2000, **275**, 29426-29432.
361. A. Fontana, P. P. de Laureto, B. Spolaore, E. Frare, P. Picotti and M. Zambonin, *Acta Biochim Pol*, 2004, **51**, 299-321.
362. M. Monti, A. Amoresano, S. Giorgetti, V. Bellotti and P. Pucci, *Biochim Biophys Acta*, 2005, **1753**, 44-50.
363. C. L. Trkulja, E. T. Jansson, K. Jardemark and O. Orwar, *J Am Chem Soc*, 2014, **136**, 14875-14882.
364. J. L. DeRisi, V. R. Iyer and P. O. Brown, *Science*, 1997, **278**, 680-686.
365. M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson and G. Dollinger, *Proc Natl Acad Sci U S A*, 2000, **97**, 5802-5806.

366. J. Luo, J. Fishburn, S. Hahn and J. Ranish, *Mol Cell Proteomics*, 2012, **11**, M111 008318.
367. C. Schmidt and C. V. Robinson, *Nat Protoc*, 2014, **9**, 2224-2236.
368. Y. Shi, J. Fernandez-Martinez, E. Tjioe, R. Pellarin, S. J. Kim, R. Williams, D. Schneidman-Duhovny, A. Sali, M. P. Rout and B. T. Chait, *Mol Cell Proteomics*, 2014, **13**, 2927-2943.
369. A. G. Hughson, G. F. Lee and G. L. Hazelbauer, *Protein Sci*, 1997, **6**, 315-322.
370. C. Zhang and R. B. Darnell, *Nat Biotechnol*, 2011, **29**, 607-614.
371. M. R. Sawaya, D. Cascio, M. Gingery, J. Rodriguez, L. Goldschmidt, J. P. Colletier, M. M. Messerschmidt, S. Boutet, J. E. Koglin, G. J. Williams, A. S. Brewster, K. Nass, J. Hattne, S. Botha, R. B. Doak, R. L. Shoeman, D. P. DePonte, H. W. Park, B. A. Federici, N. K. Sauter, I. Schlichting and D. S. Eisenberg, *Proc Natl Acad Sci U S A*, 2014, **111**, 12769-12774.
372. S. Minguet, M. Swamy, B. Alarcon, I. F. Luescher and W. W. Schamel, *Immunity*, 2007, **26**, 43-54.
373. G. Parkinson, C. Wilson, A. Gunasekera, Y. W. Ebright, R. H. Ebright and H. M. Berman, *J Mol Biol*, 1996, **260**, 395-408.
374. F. Posas and H. Saito, *Science*, 1997, **276**, 1702-1705.
375. R. L. Roberts and G. R. Fink, *Genes Dev*, 1994, **8**, 2974-2985.
376. T. Pawson and J. D. Scott, *Science*, 1997, **278**, 2075-2080.
377. R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe and D. Eisenberg, *Nature*, 2005, **435**, 773-778.

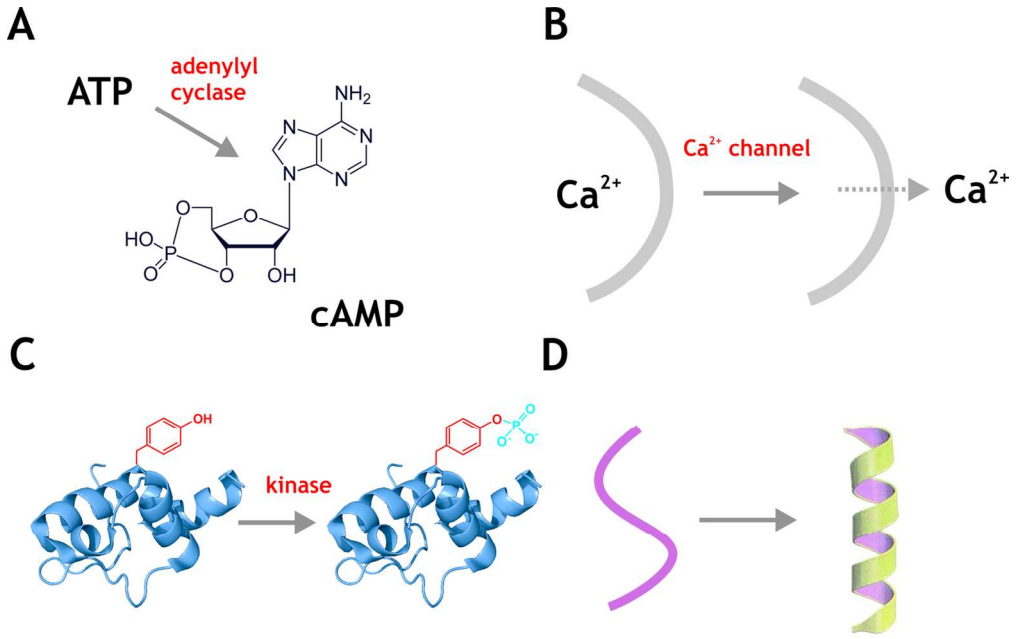


Figure 1 Covalent and conformational signals in the cell
Signal transduction in the cell relies on the appearance of a novel chemical entity, i.e. a novel spatial and chemical constellation of atoms perceived by a receptor protein with an appropriate binding surface. A) The classic example is the second messenger cAMP, which is generated form ATP by the action of the enzyme adenylyl cyclase. cAMP then binds to and activates a broad range of proteins. B) A messenger may also be "generated" by cellular translocation, as appears with Ca2+ ions, which, upon activation, are released from the endoplasmic reticulum into the cytoplasm where they can act through dedicated mediators (e.g. calmodulin). C) A covalent signal can also arise by the modification of a protein, as best exemplified by phosphorylation of Tyr residues by protein kinases. D) A novel constellation of atoms might also arise by the conformational change of a protein, as illustrated here by the appearance of a helical segment. This can bind to partners with the appropriate recognition site.

140x88mm (300 x 300 DPI)

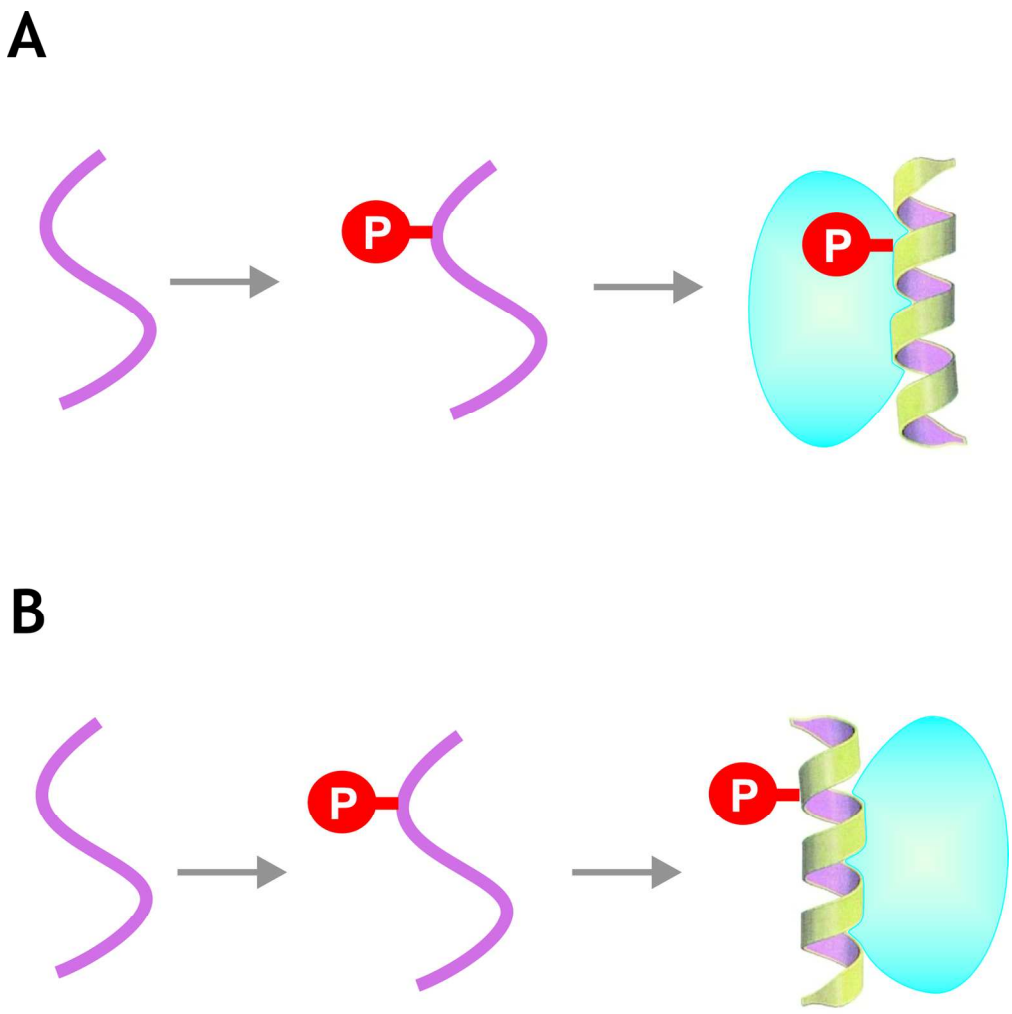


Figure 2 Covalent modification of a protein can generate a conformational signal
A) The modification of a protein can generate a covalent signal, when a PTM creates a novel binding surface recognized by a cognate binding domain (blue). B) The modification may also have its primary effect on the conformation of the protein, which either acts on the activity of the protein or serves itself as the signal, being recognized by a downstream effector.

158x158mm (300 x 300 DPI)

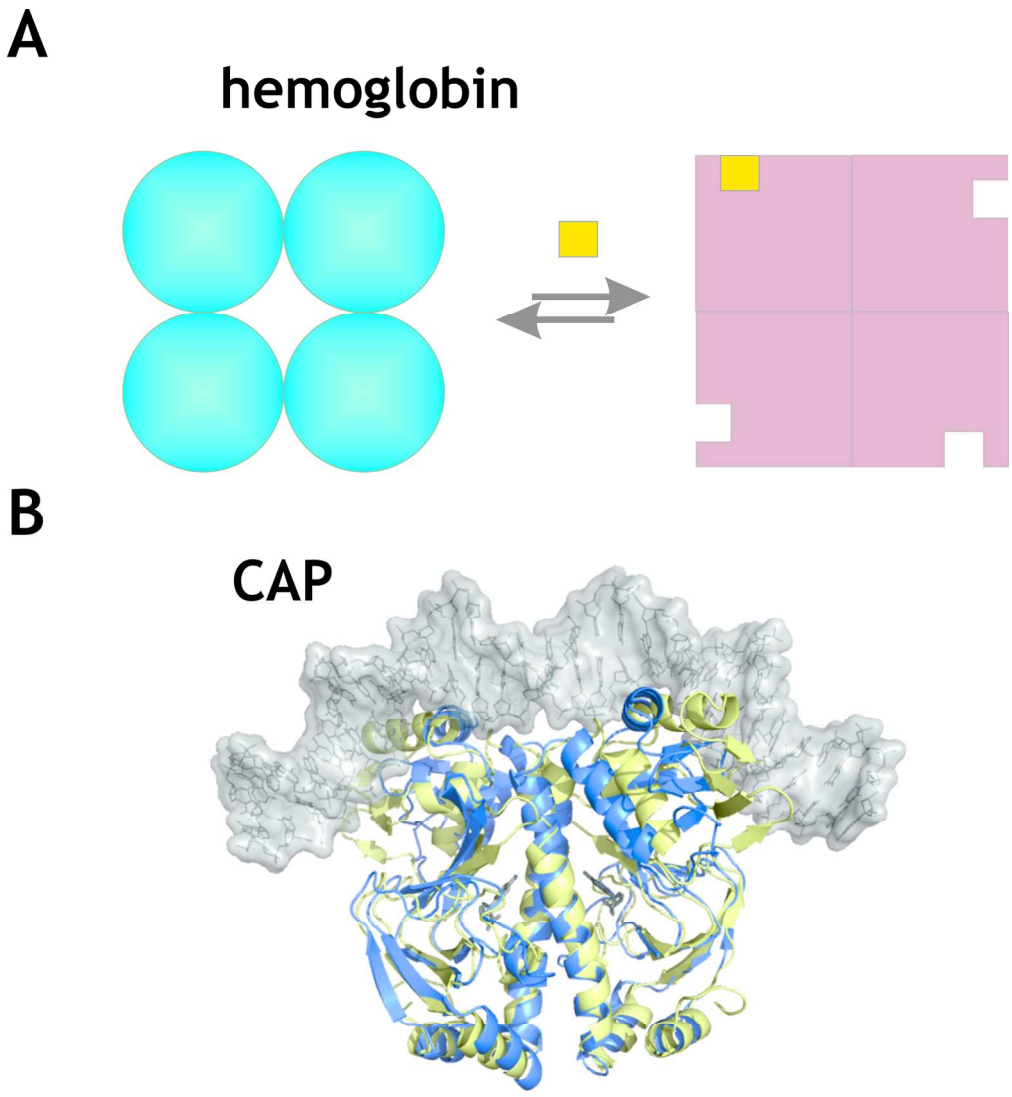


Figure 3 Allosteric signaling
A) The essence of classical allostery is a conformational change that spreads between subunits of a homo-oligomeric protein, such as hemoglobin. This scheme shows that binding of the ligand (oxygen, a yellow square) to one subunit shifts the structure of the entire tetramer to favor subsequent binding events at the other subunits. The initial conformational change thus spreads between subunits as a signal. The conformational change is subtle and details are not shown. B) Allosteric conformational change can also spread within one subunit, as demonstrated for catabolite activator protein (CAP). CAP is a cAMP-dependent transcription factor, which has a DNA binding helical region, rendered incompetent for DNA binding by a 90-degree rotation in the absence of cAMP (yellow, pdb 2WC288). When cAMP binds to both subunits (in a cooperative manner itself), it causes a significant conformational change, i.e. rotation of the helix, to assume a position relevant for DNA binding (blue, pdb 1J59373).

170x189mm (300 x 300 DPI)

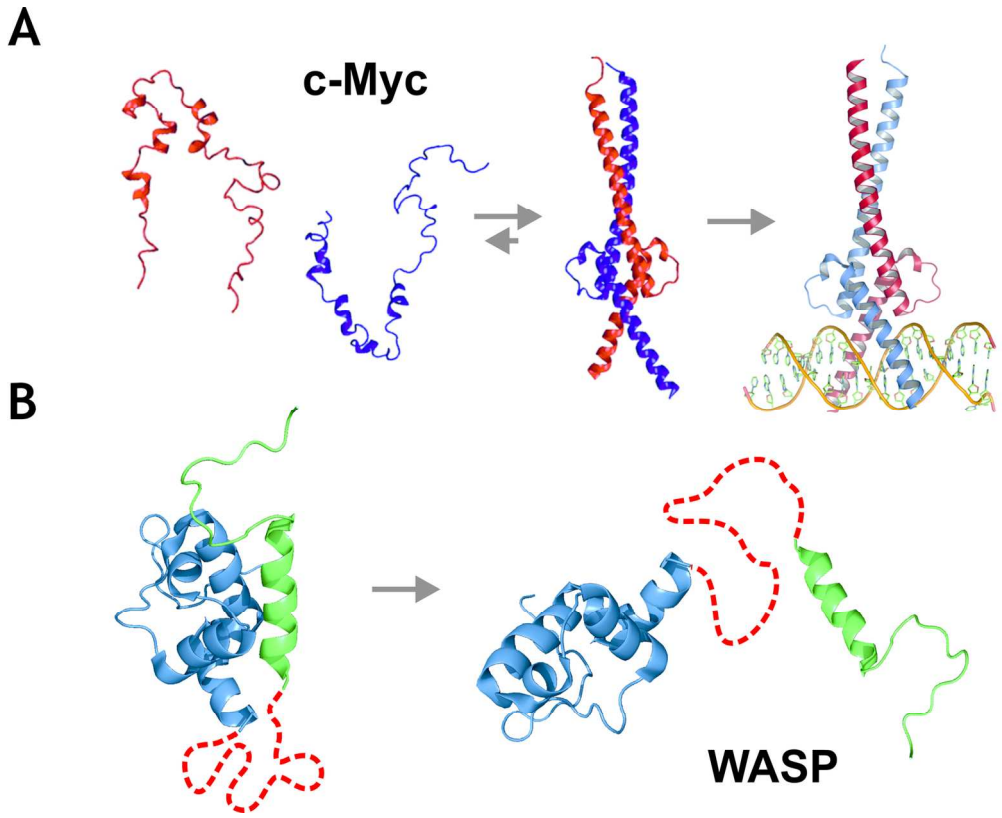


Figure 4 Conformational signaling can be generated by induced folding and unfolding
A) c-Myc belongs to the bZip transcription factor family, which are structurally disordered and inactive proteins in the monomeric state. Their activation is usually induced by a PTM (specific phosphorylation), which makes them assemble into a coiled coil dimeric structure. The mutual induced folding of the two subunits creates a signal (i.e. a DNA binding surface), due to which c-Myc binds the appropriate recognition element and initiates transcription. B) Induced unfolding can also generate a signal, as is the case with WASP protein. In the naïve state, the protein is kept inactive by the interaction between its N-terminal GBD (blue) and C-terminal VCA (green) domains. Upon activation by Cdc42 binding to GBD, the VCA region is released form GBD and is exposed for interaction with Arp2/3 to promote actin polymeriztation.

153x123mm (300 x 300 DPI)

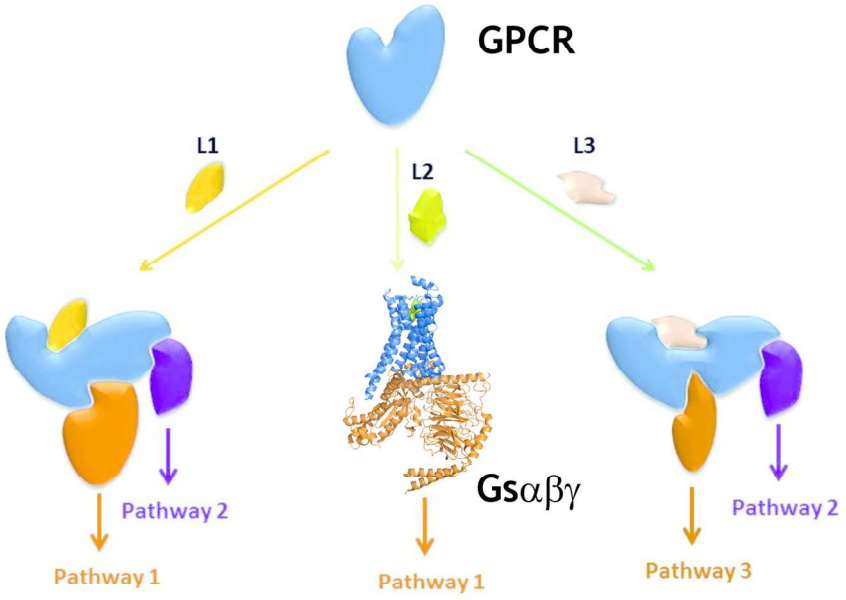


Figure 5 Conformational switches: alternative states of G-protein coupled receptors
Conformational signals can make switches of multiple states, which signal to distinct downstream signaling pathways. Canonical GPCR signaling proceeds through heterotrimeric G-protein partners, such as Gsαβγ. The highly dynamic GPCR structure (blue), however, can assume multiple conformational states that can signal through distinct messenger system in a ligand-specific manner (distinct ligands indicated as L1 through L3). For example, the existence of multiple conformations has been shown for β2-adrenergic receptor by NMR,77 but only in one state, in the presence of an agonist (green) and Gs heterotrimer (orange), could the structure of the receptor be solved by X-ray crystallography (pdb 3SN6).149

162x121mm (300 x 300 DPI)

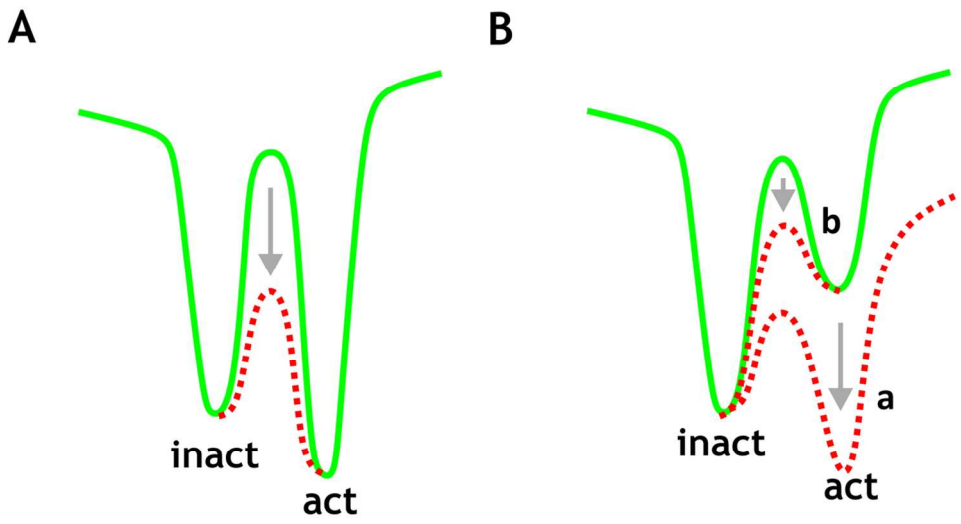


Figure 6 Thermodynamics of conformational signaling
Generation of a conformational signal can be conceived by two energetic pathways, both of which can be activated by any type of signaling input, i.e., a PTM, ligand, physical input, and interacting protein partner, or the action of specific chaperone-like proteins. A) If the active (act) conformation is more stable than the inactive (inact) one, but it is separated by a high energy barrier, activation can occur by bringing the barrier down. B) If the active conformation is less stable than the inactive one, signaling may remodel the conformational landscape (a) to make it more favorable, or (b) to bring it into a high-energy but kinetically stable state by transient interaction.

120x65mm (300 x 300 DPI)

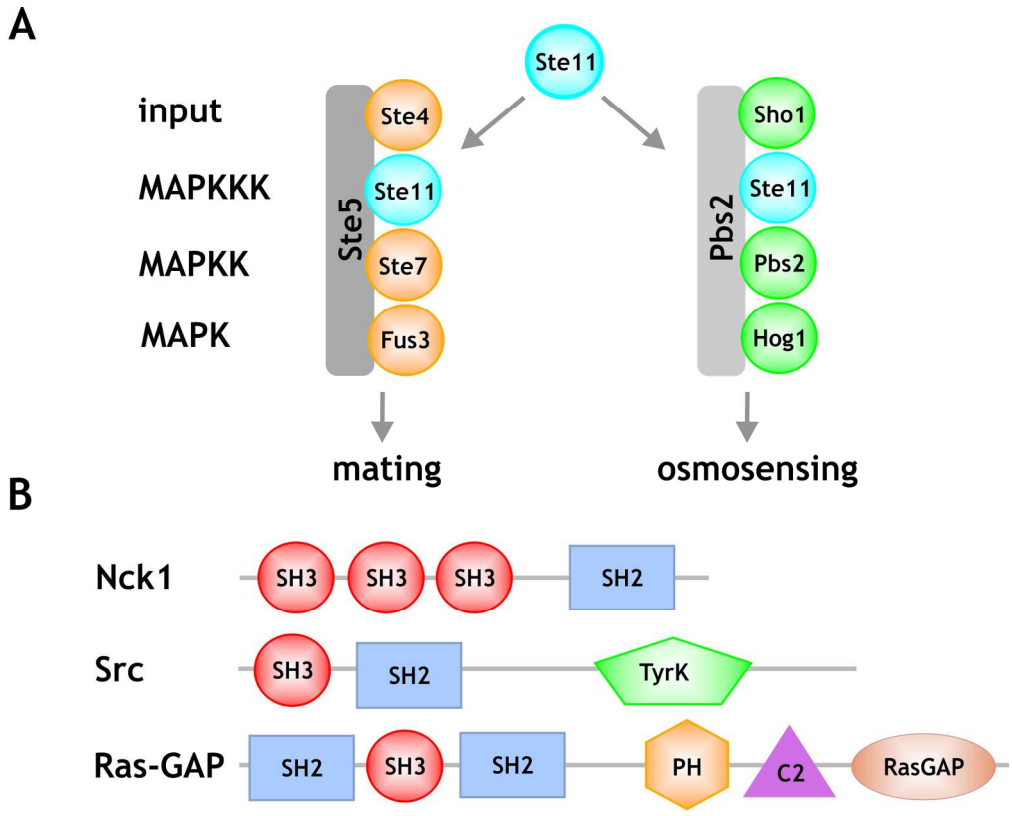


Figure 7 Modularity in signal transduction
A key principle of signal transduction is modularity, when the same functional unit is used in different evolutionary context(s). Distinct signaling pathways may use the same (type of) enzymatic activity, when the homologue of the same gene product is linked with different upstream and downstream signaling components. For example: (A) the same kinase can be recruited into distinct pathways (e.g. Ste11 MAPKKK in yeast used in the mating and osmosensing pathways165, 374, 375 by the assistance of different scaffold proteins (grey rectangles), or (B) the same domain can be used in combination with distinct other domains to constitute various signaling switches (e.g. the SH2 pTyr-binding is domain combined with other types of domains to create multidomain signaling proteins of complex signal integrating capacity176, 376).

181x145mm (300 x 300 DPI)

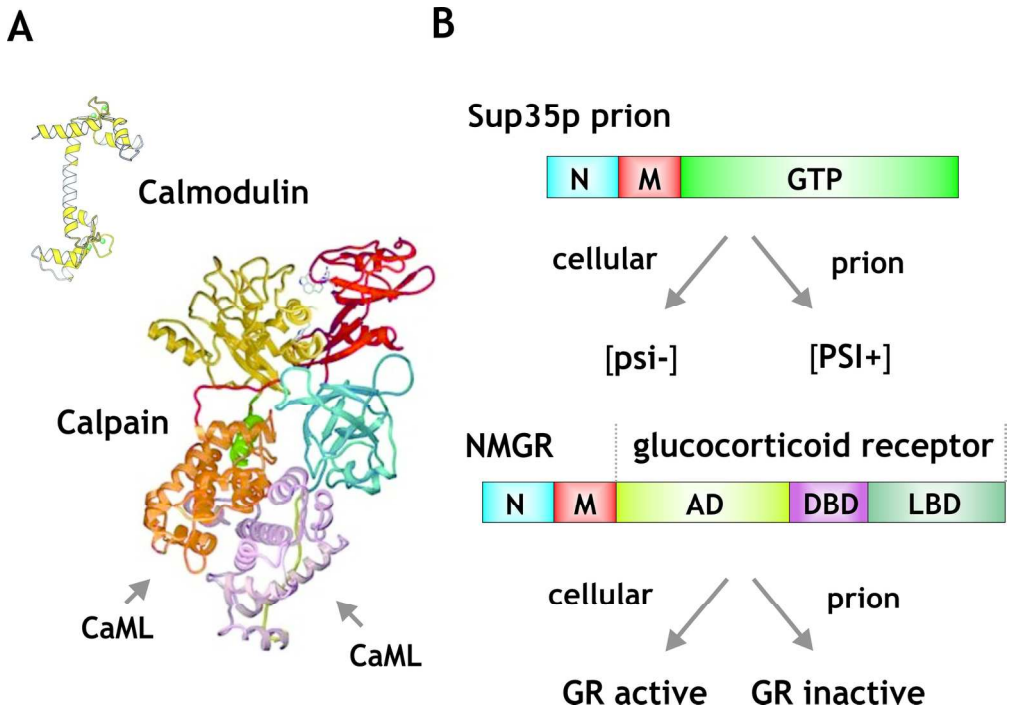


Figure 8 Modularity in conformational signaling
Modularity is also apparent in conformational signaling, when the same module capable of generating a conformational signal is used in different contexts. (A) For example, calmodulin, a calcium-sensing mini-protein that changes conformation upon calcium binding is incorporated into myriads of pathways as a stand-alone protein, but it is also part of the multi-domain protein calpain, a calcium-dependent cysteine protease of two subunits, both incorporating a calmodulin-like domain (CaML). (B) A modular conformational signal can also be created artificially. The NM region of the physiological yeast prion Sup35p can undergo a reversible change to a prion (amyloid) state. In [psi-] cells, Sup35p is soluble and functional (causing translation termination at stop codons). When the NM region converts to the prion state, Sup35p becomes insoluble and nonfunctional, causing an increased rate of nonsense suppression (translational read-through of stop codons), termed the [PSI+] phenotype. When the NM region is fused to glucocorticoid receptor (GR), it turns GR into a prion (NMGR), which loses normal GR function when it converts to the prion state.211 NM can thus be considered a module capable of generating a conformational signal.

164x114mm (300 x 300 DPI)

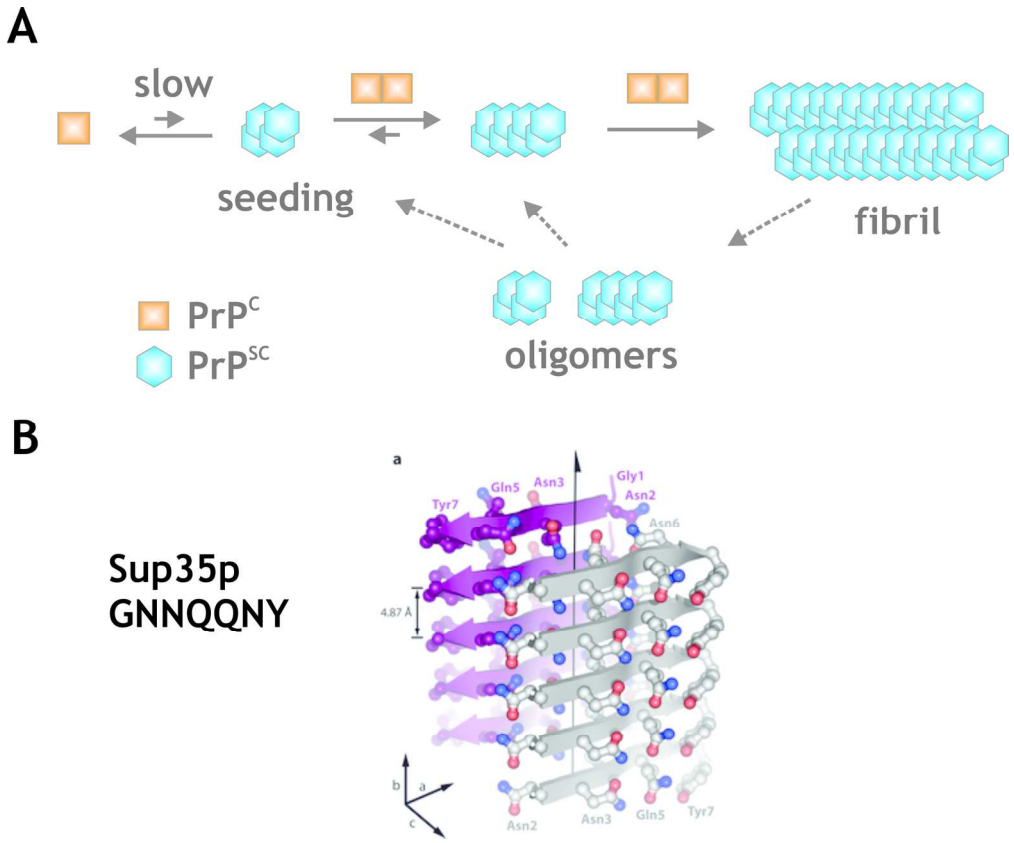
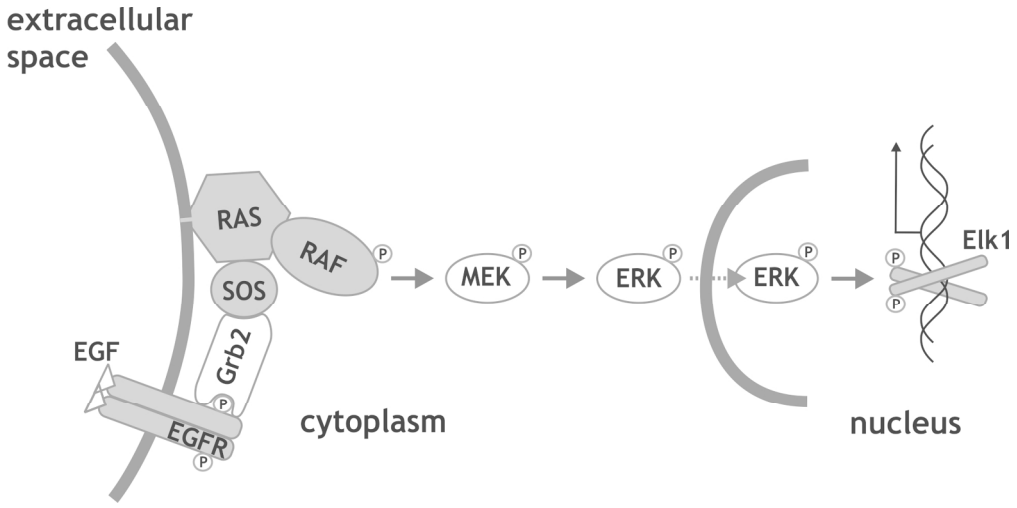


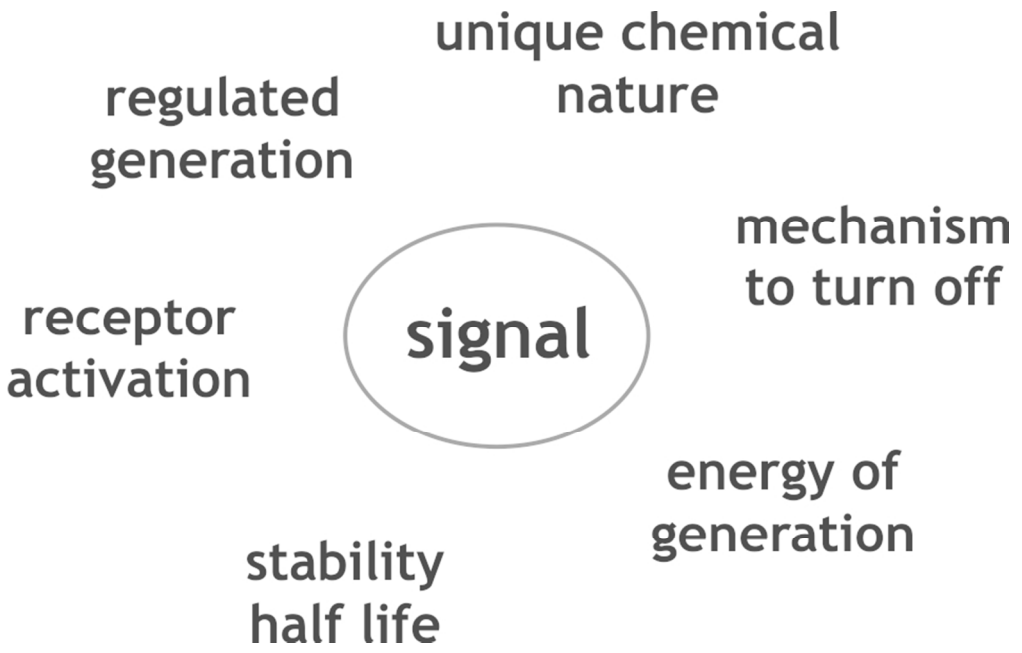
Figure 9 Conformational switches made of amyolids
Conformational signals are dynamic, they can make switches that can build pathways. (A) Prions are autocatalytic conformational switches, which can exist in a naïve (cellular, inactive, PrPC) and scrapie (prion, active, PrPSc) state. When the prion conformation arises, it can "signal" naïve molecules to undergo a conformational change and become prions themselves. Due to the autocatalytic nature of the conversion, the signal can affect many proteins, it can be amplified and even be transmitted between cells, very much like a signal in signaling cascades. (B) Formation of amyloids typically entails a transition to a cross-β structure, which forms the basis of autocatalytic transformation of the cellular form to the prion state. X-ray crystallography of a 7-residue segment of Sup35 prion domain reveals a tightly-packed β-conformational steric zipper structure.377 Reproduced with permission from Nature Publishing Group.

168x139mm (300 x 300 DPI)



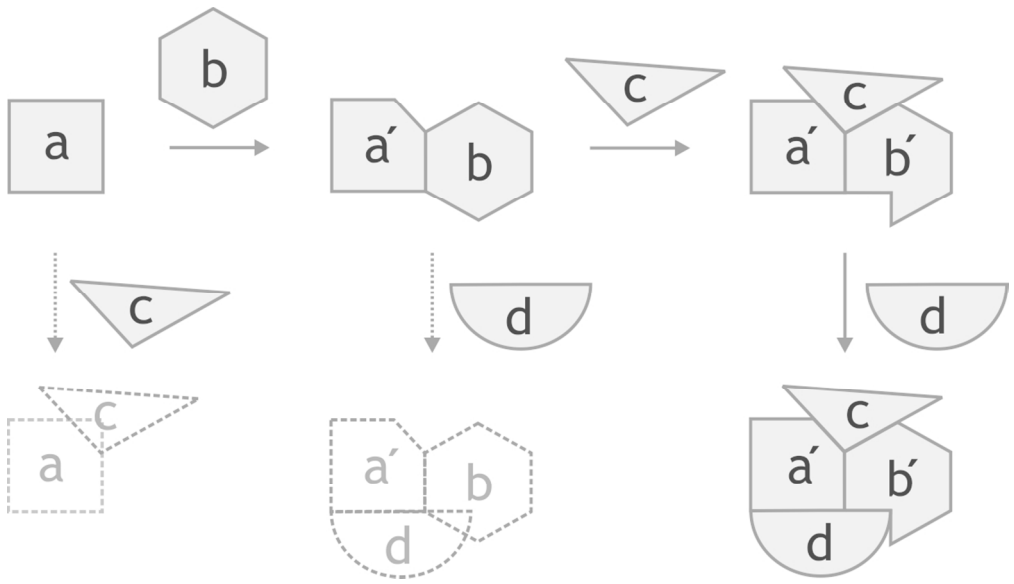
Scheme 1 MAPK cascade
The MAPK cascade transduces information from an extracellular signal (EGF) to activating functional genes (by the transcription factor Elk1) via a cascade of phosphorylation events of successive kinases (shown by P marks). Nevertheless, several proteins of the pathway (shown in dark) undergo large signaling conformational changes, which are not necessarily linked to phosphorylation.

249x123mm (200 x 200 DPI)



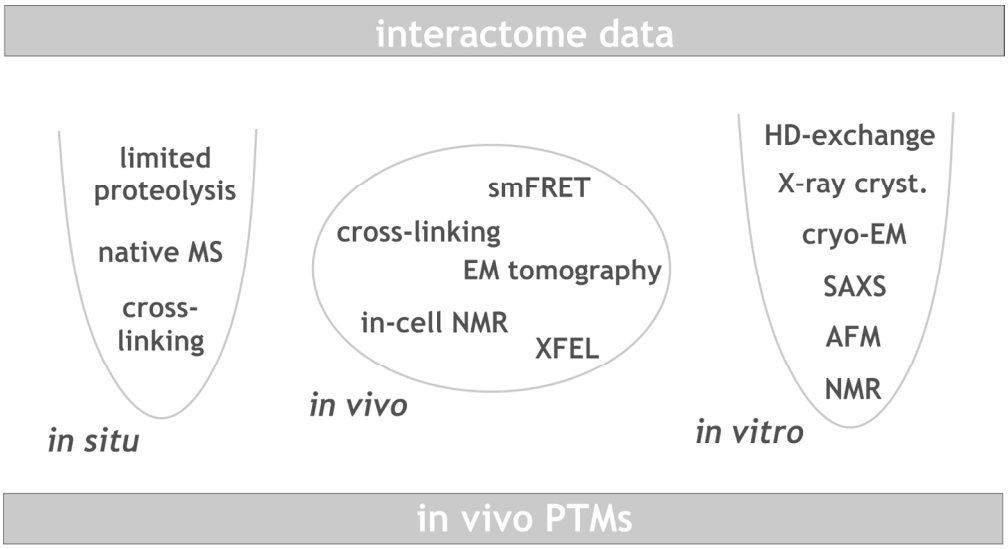
Scheme 2 Covalent and conformational signals have to meet the same requirements
There are several requirements for both a covalent (a small molecule, covalently modified protein) and a conformational (a protein of altered conformational state) signal, which they have to meet to be able to play a signaling role.

120x76mm (200 x 200 DPI)



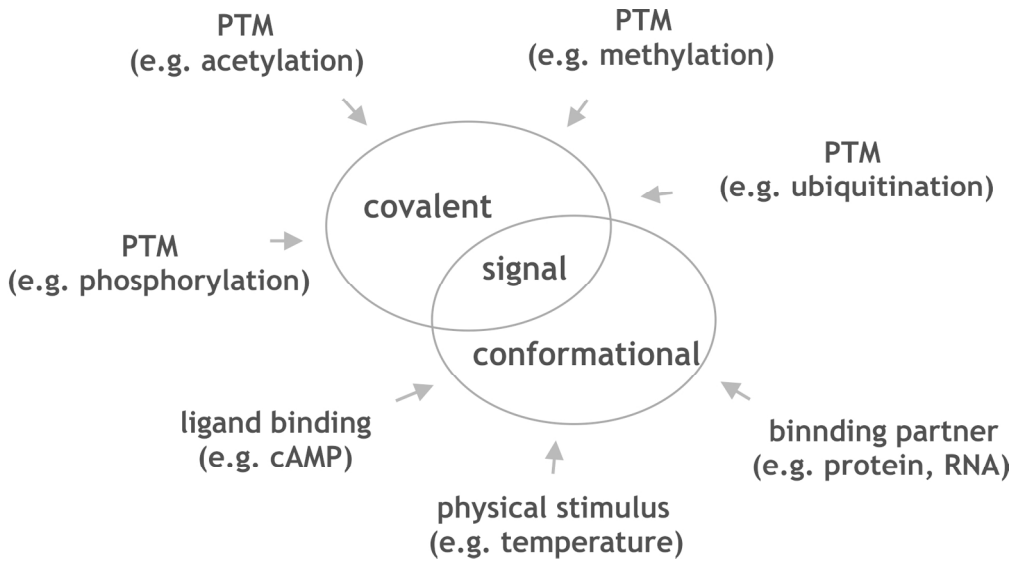
Scheme 3 Assembly pathway of a complex is determined by conformational signals
The assembly of macromolecular complexes does not occur by the random addition of components, but usually follows a concrete pathway. The scheme depicts such as pathway, in which proper assembly (full arrows and dark objects) is defined by conformational signals that develop upon the assembly of the components (object a and b turning to a' and b' upon interaction with b and c, respectively) in a predefined order. The addition of components out of their proper context (dashed arrows) results in dead-end, non-specific interactions (grey objects) that cannot proceed toward the final fully functional state.

156x89mm (200 x 200 DPI)



Scheme 4 Synergy of structural methods can address conformational signals
The scheme depicts that the HT identification of conformational signals takes much more than simply solving structures or describing structural changes of isolated proteins. Approaches applicable on isolated proteins in vitro have to synergize with studies in situ (on extracts) and in vivo. Furthermore, data generated by these approaches have then to be co-interpreted with HT interactome and PTM data, to set apart conformational signals and outline likely conformational signaling pathways in the cell.

227x122mm (200 x 200 DPI)



Scheme 5 Interplay of covalent and conformational signals
A PTM is thought to generate a covalent signal (a modified protein), whereas other stimuli are thought primarily to act by conformational changes. As shown through many examples throughout the manuscript, the mode of action of the two stimuli overlaps, often PTMs act by way of conformational changes that act as signals.

242x132mm (200 x 200 DPI)

Table 1 *First messengers in cellular signaling*[a]

Type	Message (source)	receptor[b]	Second message	action[c]	reference
physical	Light (external)	photoreceptor protein opsins (GPCR)	cyclic GMP-gated cation channel, hyperpolarization of photoreceptor cell	vision, phototaxis, photoperiodism	[23]
	Temperature[d] (external)	heat shock factor (HSF) transcription factor	Direct binding to Heat shock sequence element (HSE)	Upregulation of heat shock proteins (HSPs)	[24]
metabolite	Serotonin (neurotransmitter in gastrointestinal tract, blood platelets, and the central nervous system)	5-HT receptors (GPCRs and ligand-gated ion channels)	Heterotrimeric G protein activation	aggression, anxiety, appetite, learning, memory, thermoregulation	[25]
	Acetylcholine (neurotransmitter)	Acetylcholine receptor (GPCR)	Activated heterotrimeric G protein, PLC, IP_3	Elevated Ca^{2+} level, neuronal activation	[26]
steroid	Testosterone (from testis)	Androgen receptor (nuclear hormone receptor)	Direct transcription factor action	Genital development, generation of sperm	[27]
	Progesterone (from ovaries, placenta, adrenal glands)	nuclear progesterone receptor	Direct transcription factor action	menstrual cycle, pregnancy, embryogenesis	[28]
peptide	Insulin (from pancreas upon glucose stimulus)	Insulin receptor (Tyr kinase receptor)	Insulin receptor substrate phosphorylation	glucose uptake via glucose transporter	[29]
	Vasopressin (synthesized in hypothalamus)	receptor (GPCR)	Activated heterotrimeric G protein, PLC ($IP3/Ca^{2+}$), Adenylate cyclase (cAMP)	Retention of water, constriction of blood vessels	[30]
protein	EGF (secreted by various tissues such as submandibular	EGFR	Activating MAPK cascade	cell growth, proliferation, differentiation	[31]

	gland, parotid gland)				
	Interleukin-2 (cytokine, secreted by activated T cells)	IL-2 receptor	Activating MAPK, PI3K and JAK-STAT pathways	Activation of B-cells, macrophage, and natural killer cells	[32]

[a]list illustrating the types of possible primary signals
[b]often primary signals act on multiple receptors, only one illustrative is mentioned
[c]many hormones have multiple effects, only a few examples are given
[d]this is only one type of response to temperature

Table 2 *Second messengers in cellular signaling*[a]

Message	Turning on/source	First message	Turning off	Receptor[b]	Action[c]	Ref.
cAMP	From ATP by adenylate cyclase	Glucagon, adrenaline	Phosphodiesterases	PKA, calcium channels, GEFs	glycogen, sugar, and lipid metabolism	51
Ca^{++}	Released from intracellular stores (ER)	Glutamate, IP3, membrane depolarization	Uptake into intracellular stores	CaM, calnexin, protein kinase C	excitability, exocytosis, motility, apoptosis, transcription	52
cGMP	From GTP by guanylate cyclase	Atrial natriuretic factor, NO	Phosphodiesterases	protein kinase G, ion channels, cyclic nucleotide phosphodiesterases	glycogenolysis, apoptosis, relaxation of smooth muscle, vasodilation	53
NO (nitric oxide)	From L-arginine by NO synthase (NOS)	Acetylcholine, Ca2+/CaM, cytokines	Spontaneous decomposition within seconds	Guanylyl cyclase, K+ channels,	vascular tone, insulin secretion, inflammation, angiogenesis	54
inositol 1,4,5-trisphosphate (IP_3)	From phosphatidylinositol 4,5-bisphosphate (PIP_2), by phospholipase C (PLC)	Acetylcholine, Insulin	Inositol 5' phosphatase	IP_3 receptor	Release of Ca^{2+} from ER, cell proliferation, contraction	55
Diacylglycerol (DAG)	From PIP_2 by PLC	Acetylcholine, Insulin	DAG and monoacylglycerol (MAG) lipases	Protein kinase C	Cell proliferation, prostaglandin synthesis	56

[a]Only a few of the best known second messengers are given
[b]characteristic examples, as second messengers may be recognized by many receptors
[c]depending on the first messenger, the effect can be mutifarous

Table 3 *Conformational (mis)signaling by prions*[203]

Prion[a]	Prion region	Physiological/pathological	Function (cellular state)	Function (prion state)
PrP (human)	residues 90–230	pathological	PrP^{C}: cellular soluble form, function in long-term memory consolidation	PrP^{Sc}: prion (scrapie) form: disperse neuronal cell death causing spongiform encephalopathy(TSE)
Sup35p (yeast)	residues 1-123 (N domain) Q/N-rich	physiological	[psi-] phenotype: translation termination	[PSI+] phenotype: stop codon readthrough
Ure2p (yeast)	residues 1-89 Q/N-rich	physiological	[ure-o] phenotype: repression of genes of utilization of poor nitrogen sources when preferred nutrients such as ammonia or glutamine are present	[URE3] phenotype: loss of repression of nitrogen catabolic genes, utilization of poor nitrogen sources
HET-S (Podospora anserina)	residues 218–289: G/V-rich	physiological	[Het-s*] state: fusion of cells leads to the formation of a viable mixed cell (heterokaryon)	[Het-s] state: programmed cell death that results in heterokaryon incompatibility
CPEB (Aplysia californica)	residues 1-160: Q-rich	physiological	Low activity of mRNA polyadenylation	Functional prion: mRNA polyadenylation, initiation of translation required for synaptic stabilization

Table 4 *Novel developments in structural techniques*

Technique	specific	Ensemble/single molecule[a]	*In vitro/in vivo*[b]	Conformational change[c]	Reference[d]
X-ray crystallography	Stabilization of structure	Ensemble	*In vitro*	yes	[149]
	XFEL	Single molecule	*in vitro*	Yes	[249]
NMR	DNP	Ensemble	*In vitro*	Yes	[252]
	Specific labeling	Ensemble	*In vitro*	Yes	[253]
	Relaxation dispersion	Ensemble	*In vitro*	Yes	[254]
	Solid state	Ensemble	*In vitro*	Yes	[255]
AFM	Fast tapping AFM	Single molecule	*In vitro*	Yes	[256]
	Protein unfolding	Single molecule	*In vitro*	Yes	[257]
SAXS	EOM	Ensemble	*In vitro*	Yes	[258]
Electron microscopy	Cryo-EM	Single molecule	*In vitro*	yes	[259]
Mass-spectrometry	Native MS (ESI)	Single molecule	*In vitro*	Yes	[260]

[a] *if the technique has the capacity to provide structural information at the single molecule level*

[b] *if the technique can be applied under cellular conditions*

[c] *whether the technique is sensitive to local conformational changes*

[d] *a representative reference, many more can be found in the main body of the text*

Table 5 *Structural techniques applicable for uncovering conformational signals in situ/in vivo*

Technique	specific	Ensemble/single molecule[a]	*in situ/in vivo*	Conformational change[b]	Reference[c]
X-ray crystallography	XFEL	Single molecule	*in vivo* (also *in vitro*)	Yes	[249]
NMR	In-cell	Ensemble	*in vivo*	Yes	[347]
Single-molecule fluorescence	smFRET	Single molecule	*in vivo*	Yes	[348]
Electron microscopy	tomography	Single molecule	*in vivo*	Yes	[349]
Limited proteolysis	LiP-SRM	Ensemble	*in situ* (extract)	Yes	[350]
Cross-linking		Ensemble	*in vivo*	yes	[303]

[a] *if the technique has the capacity to provide structural information at the single molecule level*

[b] *whether the technique is sensitive to local conformational changes*

[c] *a representative reference, many more can be found in the main body of the text*

Table 6 *Select cases of conformational signaling*

Protein[a]	Type of action/cellular process	Input	Nature of signal/activation[b]	Signaling readout	Reference
CREB	cAMP-dependent transcription factor, various processes	Phosphorylation	Increase in helix conformation	Protein-protein interaction	[108]
CAP	Transcription factor, glucose metabolism	cAMP binding	Global conformational rearrangement, helix rotation	DNA binding	[88]
c-Myc	General transcription factor	Hetero-dimerization	Helix formation	DNA binding	[109]
β2-adrenergic receptor (GPCR)	Stimulation of sympathetic nervous system	Adrenaline (epinephrine) binding	Global conformational change: twisting of helices	Binding and activation of Gsαβγ	[149, 150]
CPEB (prion)	Protein translation in memory	Increase in expression	Transition to cross-β amyloid state	mRNA polyadenylation	[205]
CaM	Ca^{2+} signaling	Ca^{2+} binding	Unfolding of linker helix	Binding of hundreds of downstream partners	[58]
Hemoglobin	Oxygen transport	O_2 binding	Distributed conformational change	Allosteric (cooperative) activation of subunits	[86, 87]
BCL-xL	Apoptotic signaling	PUMA binding	Helix unfolding	Release and apoptotic activation of p53	[125]
WASP	Reorganization of actin cytoskeleton	Cdc42 binding	Global unfolding	Activation of Arp 2/3	[122]
Glucocorticoid receptor	Transcription factor on development, metabolism, and immune	Phosphorylation, interaction with DNA-binding domain (DBD) and co-	Transition to a more helical state	Interaction with co-activator proteins CBP, TFIIF, TBP	[116]

	response	activators

[a]as conformational signaling is very general, this is a very arbitrary and illustrative selection of cases, meant to demonstrate the diversity of this phenomenon

[b]strictly speaking, proteins undergo global conformational rearrangements, here only the most characteristic feature is mentioned