Review

Cell PRESS

Allosteric post-translational modification codes

Ruth Nussinov[1,2*], Chung-Jung Tsai[1], Fuxiao Xin[3] and Predrag Radivojac[3]

[1] Basic Science Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, Frederick National Laboratory, NCI, Frederick, MD 21702, USA
[2] Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Sackler Institute of Molecular Medicine, Tel Aviv University, Tel Aviv 69978, Israel
[3] School of Informatics and Computing, Indiana University, 150 South Woodlawn Avenue, Bloomington, IN 47405, USA

Post-translational modifications (PTMs) have been recognized to impact protein function in two ways: (i) orthosterically, via direct recognition by protein domains or through interference with binding; and (ii) allosterically, via conformational changes induced at the functional sites. Because different chemical types of PTMs elicit different structural alterations, the effects of combinatorial codes of PTMs are vastly larger than previously believed. Combined with orthosteric PTMs, the impact of PTMs on cellular regulation is immense. From an evolutionary standpoint, harnessing this immense, yet highly specific, PTM code is an extremely efficient vehicle that can save a cell several-fold in gene number and speed up its response to environmental change.

PTMs expand proteome complexity with little evolutionary cost

Signaling pathways control how cells perceive and respond to the environment. One major way that pathway complexity and cellular life is regulated is through PTMs. PTMs can involve covalently linking chemical groups, lipids, carbohydrates or (poly)peptide chains to amino acids of the target molecule during or after its translation. Similar to noncovalent binding, PTM events can take place at the functional site (orthosteric PTMs) or away (allosteric PTMs). Orthosteric PTMs work via direct recognition. Allosteric PTMs can lead to conformational and dynamic changes; their introduction perturbs the protein structure because it needs to accommodate them. As databases show, PTMs are common and extensive: current data suggest more than five confidently identified PTM sites per (modified) protein in the human genome, and every fifth protein is modified by multiple PTM types [1]. PTMs frequently take place in disordered regions, which can help modification enzymes recognize and catalyze the reactions [2]. Similar to noncovalent binding, PTM events can lead to dissociation of a binding partner if the perturbations that they elicit are large enough to weaken the interaction; this can result from the cumulative effect of multiple (homo- or heterotypic) PTMs. From an evolutionary standpoint, multiple PTM sites, types, and combinations could be an advantageous route to adapt a signaling protein to an increasing number of binding partners while retaining the same number of genes in the genome (Box 1).

The large number of PTMs per molecule argues that many cannot be accommodated by recognition domains and thus must act allosterically (Figure 1). Although the literature richly describes the mechanisms of direct recognition, this is not the case for allosteric PTMs. This review first explains the allosteric mechanisms through which PTMs work. Because the allosteric effects of PTMs depend on their type, protein environment, and other PTMs on the protein, the number of possible PTM codes is vastly larger than has been recognized (Figure 2). For example, in transcription factor p53 there are at least 50 PTM sites [3]; the FoxO family of forkhead transcription factors is regulated by specific combinations of PTMs, including phosphorylation, acetylation, and ubiquitylation, where distinct FoxO PTM combinations act as a 'FoxO code' [4]. Seventeen possible PTM acceptor residues were described in FOXO3a (Fokhead box O3) alone, and it was estimated that single and binary multiple modifications could give rise to thousands of different PTM isoforms [5].

A combinatorial code imparts high specificity in a way that is similar to a jigsaw puzzle. The tight packing among all molecules means geometrical fitting: it is difficult to replace one molecule with another, particularly if it has a different shape. The shapes must fit together and there is only one way to achieve it. Because there are many types of PTMs, and a protein is typically modified at many sites, the advantages of using protein domains and whole proteins in a combinatorial manner can be further enhanced by transient PTMs [6]. PTMs are recognized by specific domains; therefore, different combinations would lead to different assemblies. Collectively, this further emphasizes the fundamental importance of PTMs in signaling [7,8] and the extraordinary extent to which evolution has exploited their occurrence. At the same time, it underscores the crucial role of allostery [9–11] in signal propagation and, consequently, in cell activity. Combined, the functional site and the allosteric PTMs provide powerful discriminatory readout codes that have been harnessed by evolution at relatively small cost to regulate biological processes.

Corresponding author: Nussinov, R. (NussinoR@helix.nih.gov)
Keywords: allostery; propagation; signaling proteins; conformational ensembles; signaling pathways; protein structure; population shift; conformational selection; induced fit.
* The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Box 1. The advantages of combinatorial PTMs from an evolutionary standpoint

Cellular signaling is complex and dynamic. Complexity is essential because the cell needs to respond to extremely large and variable combinations of conditions; dynamism is crucial, because cellular responses need to be fast. Complexity and dynamics can be mediated by protein–protein interactions and by PTMs. The strategies adopted by evolution to address complexity include genomic rearrangements, duplication events, and alternative splicing. Duplication of protein–protein interaction domains [64], among which are PTM recognition domains, is a particularly common mechanism that facilitates emergence of protein interactions and expansion of the functional repertoire [65]. The combinatorial code of proteins is a powerful theme in cell regulation [66]. Just as different PTMs and protein domains can be combined to create a protein with a unique function, different proteins can be combined to create protein complexes with a unique function [67]. One example is the specific organization of transcription factors (TFs) and cofactors in enhanceosomes, leading to gene-specific transcription initiation [68]. Enhanceosomes consist of multiple TFs bound to DNA recognition elements (REs) and their cofactors. The REs are separated by spacers, which disfavor those TFs that are either too large to fit together or too small. The interferon (IFN)-β enhanceosome crystal structures [69] show that there are few protein–protein interactions even though consecutive REs overlap. Data suggest that the organization of the REs cooperatively enhances the binding of TFs to neighboring REs and restricting others [70]. A key advantage of combinatorial patterns is that even though there can be small differences between species in the number of genes, they can present large differences in complexity [71]. Dynamicity is accomplished by propagation of the signals across the cell. Rapid propagation is helped by the functional modular organization of the cellular network [72]; by pre-encoded sequences in key regions of the proteins, such as loops and linkers which facilitate conformational transitions [73]; by tight packing at protein–protein interfaces which can be achieved by conformational disorder [74]; by large multimolecular complexes; and by PTM codes, which provide a way to regulate protein function on a very short time scale.

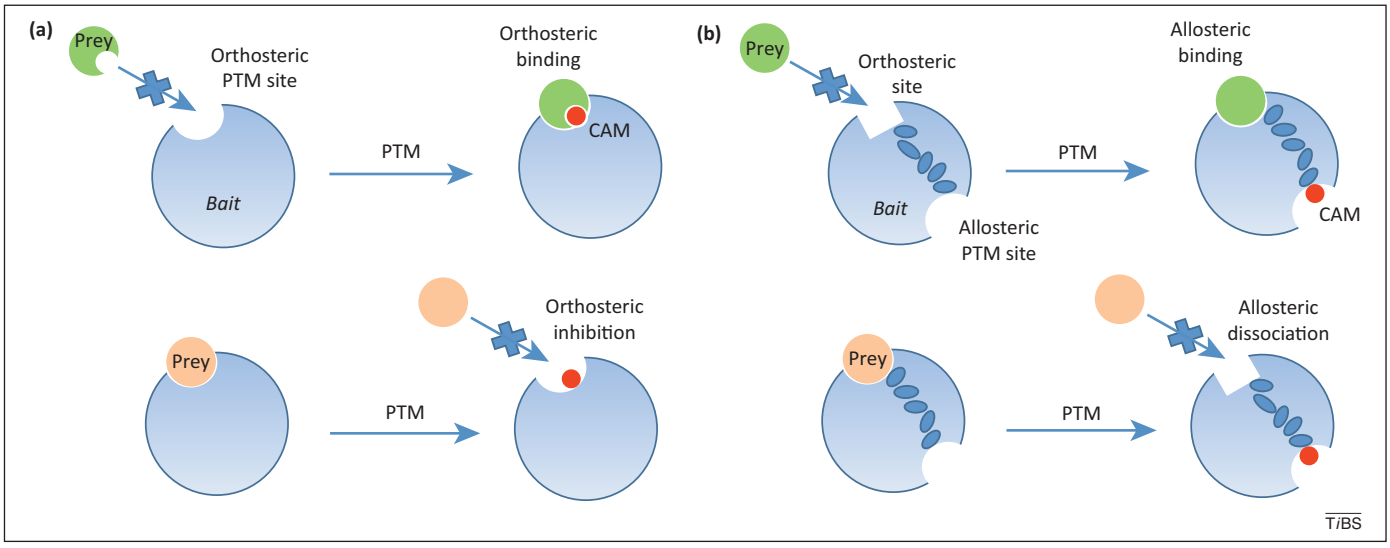

Figure 1. Post-translational modifications (PTMs) can perform their function through two distinct mechanisms. This classification is based on whether the covalently added module (CAM) is attached directly to (orthosteric) or away from (allosteric) the functional site. **(a)** Depicts orthosteric regulation by a PTM, in which the CAM (red) functions through direct interaction with a PTM recognition domain of a substrate (green or orange) at the functional site. The PTM can either promote or stabilize binding of the substrate (green, prey) to the enzyme (blue, bait) (upper row); or inhibit binding or promote disassociation (lower row) of a substrate (orange). This is the more commonly described mechanism in the literature. **(b)** Depicts allosteric regulation by a PTM, in which the CAM is located in the vicinity or far away from the functional site. The conformational change at the CAM site is illustrated by the strain energy (blue ellipsoids) that is created by the CAM and propagated to the active site (at the top). Similar to orthosteric PTMs, allosteric PTMs can either promote or stabilize substrate binding to the enzyme (upper row), or inhibit or dissociate enzyme–substrate interactions (lower row).Through the allosteric mechanism of regulation by PTMs, nature can take advantage of the enormous diversity of PTM types, combinations, and sites to achieve specific interactions among homologs in a protein family.

Modes of PTM functions

We classify PTM functions into two major categories: (i) those that are at the functional site; adopting drug terminology, we call these orthosteric; and (ii) those that are elsewhere in the molecule, away from the functional site; we refer to these as allosteric. Orthosteric PTMs function either via direct recognition by recognition domains or by blocking active sites through direct interference with binding. By contrast, allosteric PTMs function through conformational changes [12]. Since allosteric PTMs are away from the active sites, it can be expected that they are less evolutionarily conserved than orthosteric PTMs. Although not discussed in this review, chemical modifications on lipids and particularly on DNA can also follow such orthosteric/allosteric classification. Figure 1 illustrates orthosteric (Figure 1a) and allosteric (Figure 1b) PTM types.

PTMs at the functional site that act via direct recognition

We first relate to some of the major PTM types. They are of fundamental importance, provide key codes for protein function, and act in combination with allosteric PTMs. Ubiquitin recognition domains are the largest group of PTM recognition domains due to the large number of ubiquitin-type modifications [7]. Different ubiquitin chain types function in distinct cellular processes and pathways; however, current data suggest that all can target proteins for degradation [13], particularly Lys48- and Lys11-linked chains [13,14]; Lys63-polyubiquitin has a role in endocytosis, DNA-damage response and signaling [15]. Lys48-linked chains can be recognized by the ubiquitin-associated (UBA) domain of hR23; Lys63-linked chains can be recognized by the compact Npl4 zinc finger (NZF) recognition domain of TAK1 (TGF-β-activated kinase 1)-binding

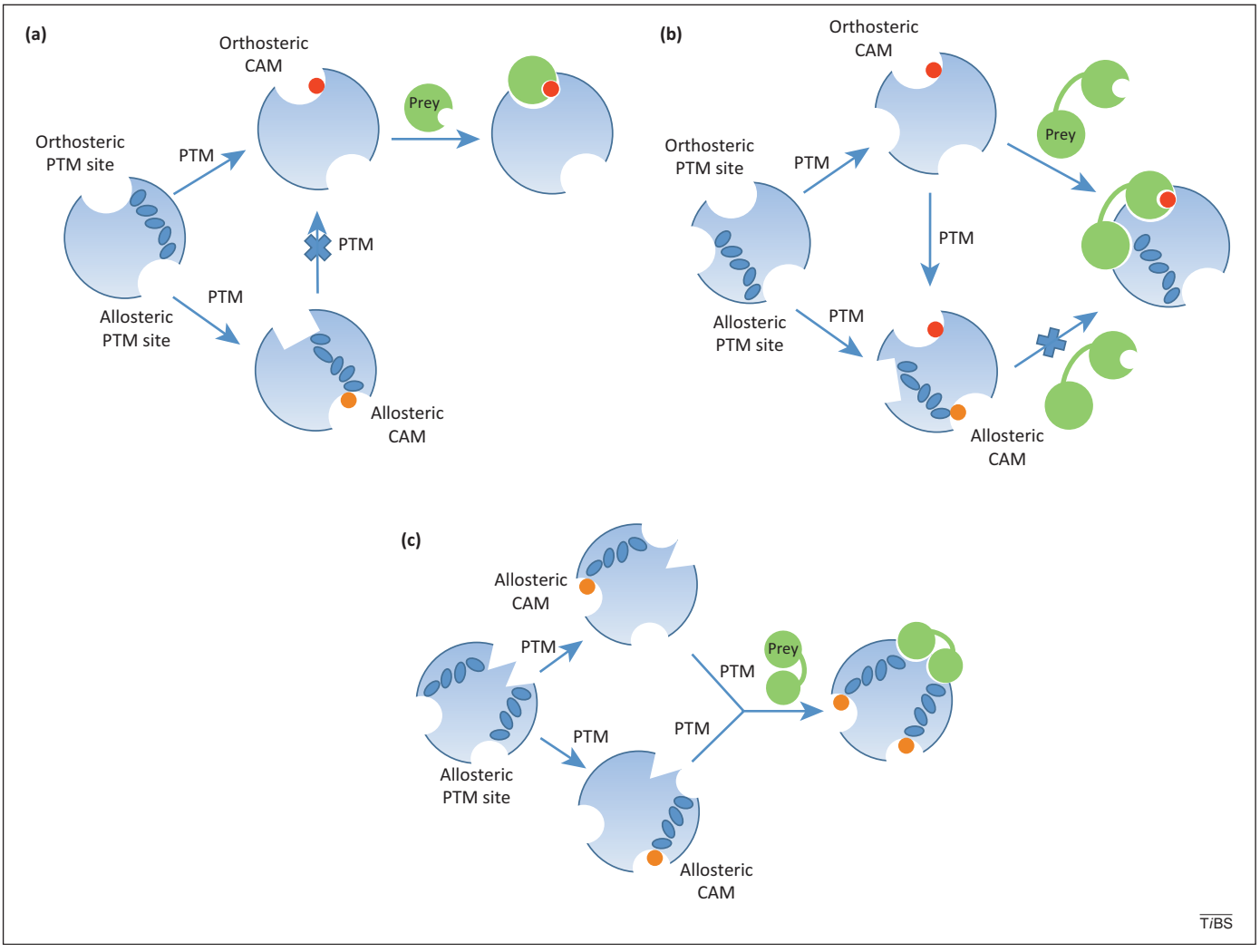

Figure 2. Allosteric post-translational modifications (PTMs) can vastly increase the complexity of combinatorial PTM codes. An allosteric PTM can play two distinct roles in a combinatorial PTM code: it can exclude binding of a substrate (green, prey) by preventing the addition of a required orthosteric PTM **(a)**, or it can control substrate binding by causing a conformational change in a nonorthosteric binding face **(b)**. In (a), the strain energy created by the allosteric covalently added module (CAM) propagates (blue ellipsoids) to the orthosteric PTM site, resulting in a conformational change at the orthosteric PTM site that prevents the addition of CAMs (red) and thus obstructs subsequent binding. In (b), the allosteric CAM (brown) causes a conformational change at another binding surface. Because the orthosteric PTMs and the distal site are required for binding a substrate (green, prey), the allosteric CAM disrupts the binding. Through a similar mechanism, allosteric PTMs can also create a positive combinatorial code (promoting binding) as depicted in **(c)** with two consecutive allosteric PTMs, rather than a negative code (disruption).

protein 2 (TAB2); and monoubiquitin can be recognized by the ubiquitin-interacting motif (UIM) domain of vps27. In addition, ubiquitin-like PTMs such as small ubiquitin-like modifier (SUMO) also play key roles [16,17]; for instance, linear SUMO chains are recognized by the UBAN [ubiquitin binding in ABIN, A20-binding inhibitor of NF-κB (nuclear factor-κB) activation] domain of NEMO (NF-κB essential modulator).

The diversity of phosphate recognition domains is limited. Phosphorylation of serine, threonine, tyrosine, aspartic acid, and histidine is a fundamental regulatory signal in the cell (Figure 3a), and a series of phosphorylation events is a common strategy in signal amplification [6]. Examples of recognition domains include Src homology 2 (SH2), phosphotyrosine-binding domain (PTB), and 14-3-3. Acetylation of the lysine ε-amino group is also frequent (Figure 3b). Acetylation of histones is a key epigenetic event. In addition to the regulation of transcriptional activity, histone acetylation can enhance protein–protein interactions (PPIs) [18] and compete with ubiquitylation on the same lysine, therefore preventing degradation [19]. Acetylated lysines are recognized by the bromodomain [6]. Methylated lysines can be recognized by the chromodomain, and methylated arginines by the tudor domain [6].

PTMs at the functional site that block active sites (Figure 3c)

Orthosteric PTMs can also act by blocking an otherwise available functional site. The bulky SUMO modification can block the binding of partner proteins if the SUMOylation site is close to a protein-binding site [20,21]. For example, SUMO attachment to the N-terminal helix of E2 ubiquitin-conjugating enzyme E2-25K impairs its interaction with the E1 ubiquitin-activating enzyme, which uses the same helix [22]; Cdc25B, which regulates the entry into mitosis, is regulated by PTMs that block binding to 14-3-3. Phosphorylation of Ser321 by a cyclin-dependent kinase blocks the 14-3-3 binding to Ser323. Loss of 14-3-3 binding

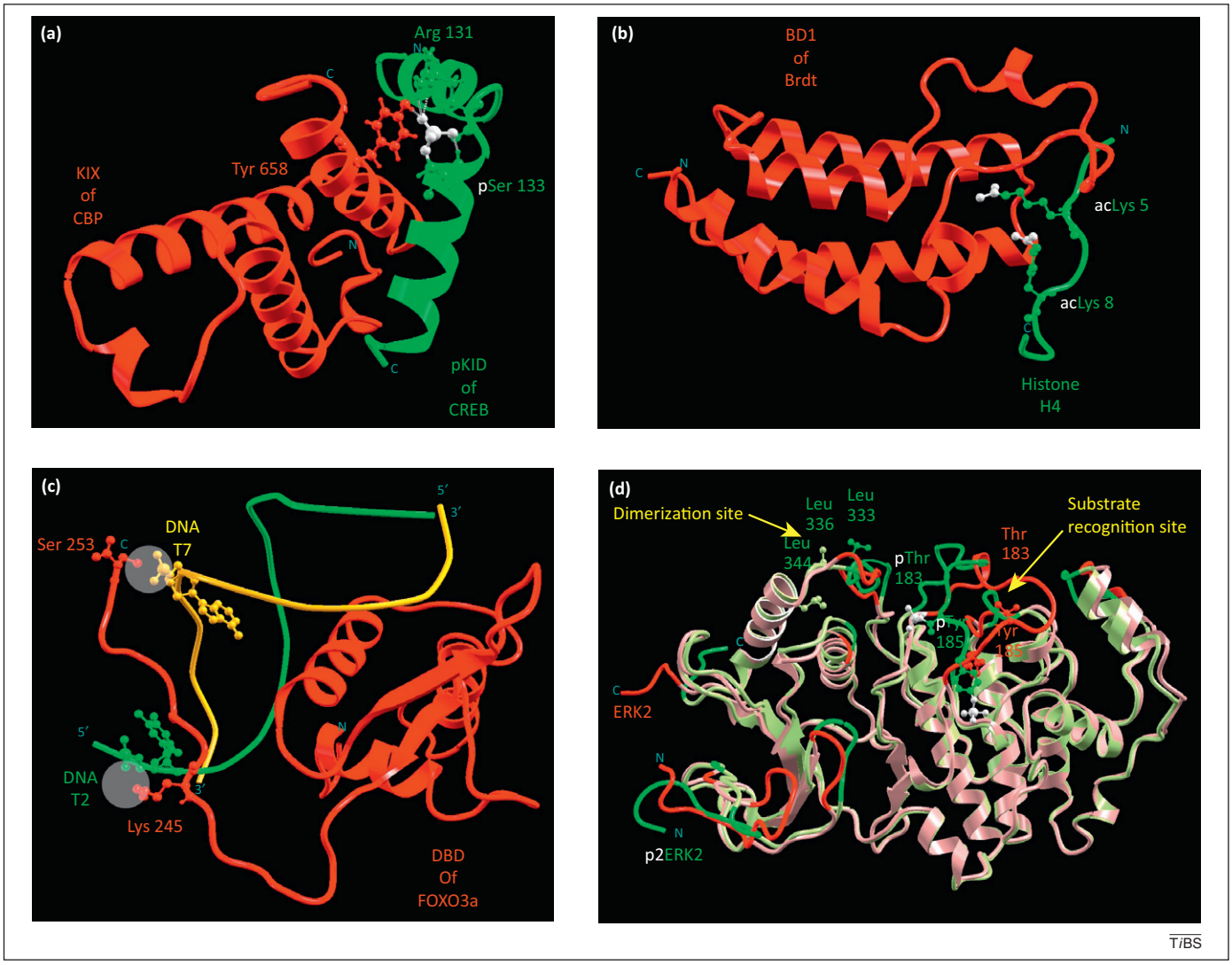

Figure 3. The structural basis of cellular function through combinatorial post-translational modifications (PTMs). **(a, b)** Illustrate two recruitment events that are facilitated through binding to single orthosteric PTM (a) and double orthosteric PTMs (b) on the partners, respectively. In (a), the phosphate of phosphoserine (pSer 133) in the pKID domain of cAMP-response element binding protein (CREB) mediates (green) CREB:CBP complex binding by forming four hydrogen bonds (broken lines), including one with Tyr658 of the KIX domain of CBP (red) (PDB 1kdx [60]). This example illustrates the direct involvement of a covalently added module (CAM) in an orthosteric PTM. In (b), two acetylation marks (double orthosteric PTMs) on a histone H4 tail (green) are recognized by the bromodomain (BD1) of Brdt (red), a testis-specific member of the BET protein family (PDB 2wp2 [61]). That BD1 fails to bind monoacetylated H4 tail illustrates the essential role of combinatorial PTMs in interaction specificity. **(c)** Illustrates how orthosteric PTMs can disrupt interactions to attenuate a functional binding event. In (c), the crystal structure of human FOXO3a DNA-binding domain (DBD) (red) is shown complexed with a 13-bp DNA duplex (green and yellow) (PDB 2uzk [62]) that contains a FOXO consensus binding sequence (GTAAACA). Two C-terminal DBD residues that can be post-translationally modified (Lys245, Ser253), highlighted by transparent circles, interact with the DNA phosphate group. Acetylation of Lys245 by CREB or phosphorylation of Ser253 by protein kinase B (PKB) has been suggested to disrupt the protein–DNA contacts and thereby reduce FOXO transcriptional activity. In **(d)**, an example of double allosteric PTMs that activate the target enzyme through conformational changes is shown. Two MAP kinase ERK2 structures are superimposed [63]; one is the inactivated form (labeled ERK2 in red, PDB 1erk) and the other is the activated, dually phosphorylated form (p2ERK2 in green, PDB 2erk). The ribbons in dark color highlight those matched residues which are separated by more than 2.0 Å. The conformational changes due to the dual allosteric PTMs (pThr183 and pTyr185) are reflected by the darkly colored ribbon region, which corresponds to the substrate recognition site and the dimerization site. In this figure, only side chains of PTMs plus residues involved in dimerization are drawn in ball-and-stick model, and the CAM(s) are highlighted in white. Abbreviations: pKID, kinase-inducible domain of CREB; CBP, cyclic-AMP response element binding protein (CREB) binding protein (CBP); KIX a domain of CREB binding protein; BET, Bromo and extra terminal family; FOXO3a, Forkhead box O3.

increases substrate access to the catalytic site of Cdc25B. By contrast, unphosphorylated Ser321 appears to help stabilize 14-3-3 binding to Ser323 and thus decrease Cdc25B activity [23]. The Ser321 phosphatase docking site overlaps that of cyclin-dependent kinase; thus, the phosphatase and kinase affect Cdc25B activity not only through catalysis but also through restricting each other's access to their target substrate [24].

PTMs away from the functional site that cause a conformational change

The roles of PTMs as allosteric modulators are well established. The covalent linkage of a large and often charged group perturbs its immediate molecular environment. The atoms around the linked group have to adapt themselves to optimize their interactions. Existing contacts can be broken and others formed. In turn, the effect of these changed contacts propagates to the next layer of atoms, which also readjust their interactions with their neighborhood. This propagation is called 'population shift' [10,11,25]. In this way, the allosteric effect resulting from a PTM perturbation propagates across the structure, similar to waves initiating from the spot where a thrown stone hits the water. However, unlike waves in a pond, allosteric propagation takes place through numerous distinct pathways; a few major, others minor. Eventually, the propagation

reaches the active site to upregulate or downregulate the activity of the protein. The allosteric effect can be expressed in small or large changes in conformation or protein dynamics.

The conformational and dynamic changes elicited by the allosteric PTMs are diverse. They can stem from electrostatic repulsion (or attraction), or van der Waals forces, attractive or repulsive; and they can lead to broadly different functional consequences. Here, we provide examples focusing on ubiquitylation and phosphorylation. Ubiquitylation can cause conformational changes in dimeric enzymes, transiently inactivating them. This type of regulation has been shown for the thyroid hormone-activating type 2 deiodinase (D2, an endoplasmic reticulum-resident type 1 integral membrane enzyme) [26] (Figure 4a). D2 is a homodimer with interacting surfaces at its transmembrane and globular cytosolic domains. Upon binding its natural substrate T4, D2 is ubiquitylated. Ubiquitylation interferes with the globular surfaces that are crucial for dimerization. The effect propagates to the active site, abolishing catalytic activity. A different scenario is observed in the epithelial Na^+ channel (ENaC), where intracellular ubiquitylation causes a conformational change that controls extracellular proteolytic channel activation, illustrating that an intracellular, post-translational modification of a transmembrane protein can affect its extracellular conformation [27]. Combined, these two examples indicate how ubiquitylation-induced conformational changes can work via different mechanisms even in transmembrane proteins.

Phosphorylation-elicited conformational changes are also common (e.g., [28–33]). Such changes can regulate the circadian clock machinery [34]. Circadian clocks are regulated by negative feedback loops that relate to

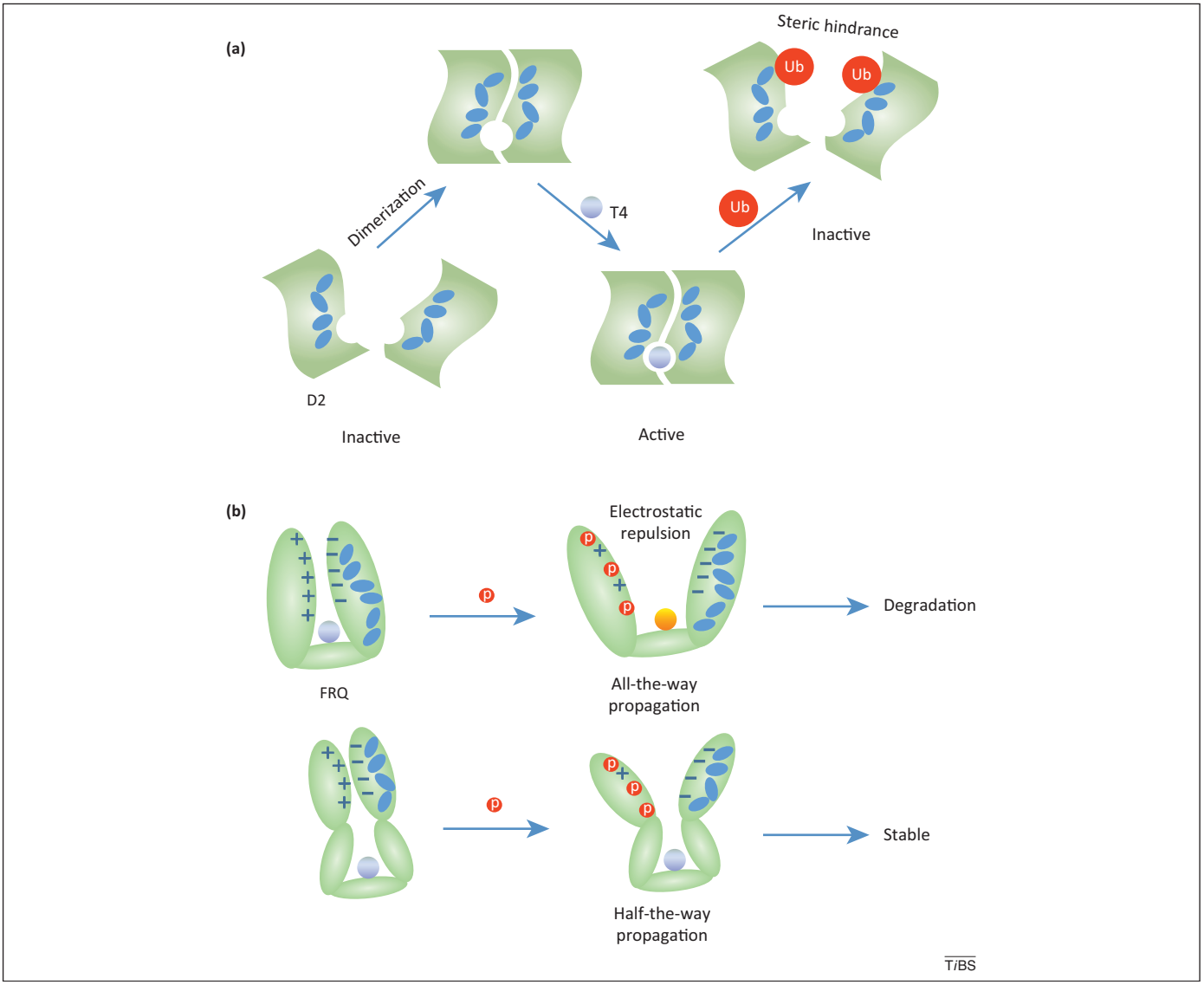

Figure 4. Schematic diagrams of two allosteric post-translational modification (PTM) examples involving ubiquitylation and phosphorylation. **(a)** A ubiquitylation event transiently inactivates thyroid hormone-activating type 2 deiodinase (D2) by allosterically disrupting its dimerization, which is essential for substrate (T4) binding. The allosteric action of the ubiquitylations at the dimer interface is indicated by the interacting blue ellipsoids, which propagate the steric hindrance down to the T4 binding site. **(b)** A progressive phosphorylation event up to a certain threshold number at the N-terminal region of circadian clock protein, FREQUENCY (FRQ), allosterically exposes its protected middle region (silver circle) and leads to its own degradation (above). The allosteric action is illustrated by the interacting blue ellipsoids, which propagate the created electrostatic repulsion between the N-terminal and C-terminal of FRQ down to the exposed middle region (gold circle). To contrast the importance of the action of allosteric propagation, a half-the-way propagation, which fails to expose the protected middle region (silver circle), is also shown (bottom).

rhythmic synthesis of transcriptional repressors that rhythmically repress their own transcription. A striking example relates to the *Neurospora crassa* clock protein FREQUENCY (FRQ) (Figure 4b). FRQ is progressively phosphorylated at up to 113 sites during the day and is eventually degraded. Two amphipathic motifs in FRQ interact to bring the positively charged N-terminal region into proximity with the negatively charged middle and C-terminal regions. This interaction leads to the recruitment of casein kinase 1a (CK1a), which then progressively phosphorylates the N-terminal domain of FRQ at up to 46 sites. The increasing number of phosphorylated sites decreases its isoelectric point and creates charge repulsion between the N terminus and C terminus. The progressive increase in charge repulsion triggers an allosteric conformational change that leads to an open conformation. This allows CK1a to access a previously hidden PEST sequence in the negatively charged central region of FRQ, which leads to FRQ degradation [35]. Another example of phosphorylation allosterically regulating the circadian machinery comes from the cyanobacterial circadian oscillator. In this system, the proteins KaiA and KaiB alternately stimulate autophosphorylation and autodephosphorylation of KaiC, respectively, with a periodicity of approximately 24 h. In this example, dynamics-driven protein allostery of autophosphorylation and autodephosphorylation of KaiC orchestrates a cycle of biochemical events. Thus, although allosteric effects can be enthalpic by causing a conformational change, they can also be entropic [36], and entropy-driven allostery can also be elicited by PTMs [37].

Other PTMs such as nitration and acetylation can also lead to diverse conformational changes (e.g., in the C terminus of the host-encoded cellular prion protein PrP^c [38]) as do the elongated and complex sugar patterns in glycosylation [39]. However, they too can regulate function through dynamic changes. N-linked glycosylation of human interleukin-7 receptor α (IL-7Rα) allosterically enhances its binding affinity to human IL-7 300-fold, but no significant conformational changes were induced by glycosylation of IL-7Rα; biophysical observations point to entropic allosteric effects [40] similar to those observed in N-glycans in the cystic fibrosis transmembrane conductance regulator (CFTR) [41].

Combinatorial 'PTM code'

A specific post-translational modification site does not necessarily correspond to a single, specific functional readout; and a complete list of PTM sites and types does not spell the 'PTM code' of a protein. Instead, the functional PTM code is most likely to be encoded in a combinatorial manner and be cell context-dependent [4,5,42–46]. Intuitively, if there were only one assigned function per gene product, the limited number of genes in the genome could not account for the huge biological complexity observed. Thus, it is not surprising that a single protein is capable of performing diverse biological functions through a reservoir of 'specific' combinatorial PTMs [4,44–46], in which each combination of PTMs spells a different, although related, function, such as enzyme activation or deactivation, or recruitment or release of a domain, protein, or complex. PTM codes can also act additively to fine tune regulation. The distinct PTM combinations in the epigenetic chromatin code that specify such functional variability of a set of proteins provide an example [47]. The effects cascade down the signaling pathway and determine cellular response. Mechanistically, a combinatorial PTM landscape is encoded to either create a specific binding surface (Figure 3b), or to disrupt an existing association (Figures 2 and 3c); the outcome can affect intramolecular or intermolecular interactions.

Why a combinatorial PTM code?

Signal transduction via PTMs is initiated by an event that stimulates an enzyme to attach (or to remove) a particular PTM to a target substrate protein. At the same PTM site, there might be competition from some other signaling event to attach a different PTM type. The activated, post-translationally modified enzyme can also activate (and sometimes post-translationally modify) another PTM enzyme in a cascading way, which further diversifies the PTM pattern. Thus, the complexity of PTM readouts, which are recognized by effector proteins or complexes, is expected: combinatorial PTMs can arise through several distinct signal transduction pathways, and a particular combination can specify a function.

Regulation based solely on an orthosteric PTM pattern has been usefully classified into several categories: cooperative (multisite); sequential (order-dependent); antagonistic, and mutually exclusive PTMs [48]. It can be further extended if we include an intramolecular PTM recognition code (not associated with an effector), or if the modifying enzyme is attaching to a complex and modifying several proteins in the complex; in such a case the target protein is a protein complex, and the PTMs are in several chains (intermolecular crosstalk) [49]. Figure 2 illustrates how the complexity of combinatorial PTM codes can be vastly broadened by including allosteric PTMs in addition to the orthosteric PTMs, and Box 1 describes the advantages of combinatorial PTMs from an evolutionary standpoint.

A PTM code might work only in a specific cellular context; however, appending contexts to combinatorial codes would immensely complicate its characterization. To account for PTM complexity, we suggest the following code/function format: each PTM code gives the total number of PTM sites (including the 'must be vacant' sites) and a description of the individual site and type. The functional description would provide the PTM code with its effector molecule and the associated action. The detailed description of each PTM site, in addition to its type, would also need to specify whether the site modification is pre-acquired (sequential, orthosteric, or allosteric), orthosteric, or allosteric. Because a specific code spells a function, each PTM code in a complete combinatorial list for a given protein or complex is assumed to be independent. However, the functions spelt by the distinct codes are related; for example, controlling association/dissociation. The dominant (overriding) [4] status of PTM codes with opposite functions should also be registered. Such a complete list of PTM codes could provide the foundation for deciphering the complex biological logic of signaling via PTMs.

Technically, a protein containing n PTM sites could result in an extremely large number of configurations;

for example, 2^n distinct molecules if each site can be modified by only one PTM type. However, because of the evidence that some groups of sites are antagonistic, that others are spatially compartmentalized or work in concert to provide a single functionality, and a possibility that yet others exert weak or no functional influence [50], it is unlikely that each distinct configuration is reachable or that it codes for a different function. For example, that there are ≥800 human proteins already verified to contain ≥10 confidently identified PTM sites in UniProtKB might argue for an extreme complexity of the human proteome. Instead, it is more reasonable to assume that there exist (generally overlapping) subsets of PTM configurations, each facilitating or carrying out a distinct context-specific function, and fine tuning is possible within each particular configuration set. The exponential nature of this combinatorial PTM code could then facilitate a wide range and diversity of (potentially) specifically regulated cellular responses. Combined with our still rudimentary understanding of protein function as a whole, this suggests that the full extent of PTM functional repertoire is far from understood. Next, we provide some combinatorial PTM examples and analyses based on specific cases in the literature and high-throughput data.

Analysis of combinatorial PTMs

Combinatorial PTM codes can comprise orthosteric PTMs, allosteric PTMs, or both. It is inherently more difficult to describe an allosteric or mixed code. This is because unlike orthosteric codes, allosteric PTM codes depend on the context of the protein. Description of specific allosteric codes is challenging: the collective effect of a PTM combination will vary across different proteins. It will depend on the protein surface environment and its chemical properties; on protein size, compactness, composition, architecture, spatial organization of specific residues, loops, and linkers through which the allosteric pathways can propagate; and the protein binding partners and environment (e.g., membrane, compartment, ions, cofactors, DNA/RNA). Yet, it is this vast combinatorial complexity that can help to fine tune function.

Epigenetic regulation can provide an example of a combinatorial PTM code [51,52]. HMGA1a, a small heterochromatin-associated high mobility group protein, is highly modified. Proteomic approaches to identify potential combinatorial modification patterns on HMGA1a [53] observed that the main combinatorial PTMs are N-terminal acetylation, Arg25 methylation and phosphorylation of the three most C-terminal serine residues. In human prostate cancer cells, the three C-terminal serine residues are phosphorylated, Arg25, Arg57, and Arg59 can be monomethylated and dimethylated; Ser35, Thr52, Thr77, Ser98, Ser101, and Ser102 are phosphorylated when Arg25 is methylated; and more. The most abundant forms of modified HMGA1a possess N-terminal acetylation and phosphorylation of two of the three residues Ser98, Ser101, and/or Ser102. Collectively, these implicate an HMGA1a PTM code [53]. Comparative analysis of histones from wild type embryonic stem cells (ESCs) and ESCs deficient in Suz12, a core component of the polycomb repressive complex 2, revealed a dramatic reduction of histone H3K27 methylation and an increase in H3K27 acetylation. This uncovered an antagonistic methyl/acetyl switch at H3K27. This effect was accompanied by H3K36 acetylation and methylation [54]. Additional examples of combinatorial histone modifications were also documented [55].

The idea of a PTM-driven combinatorial pattern is supported by our large-scale analysis of the relationship between the number of partners in a PPI network and the number of post-translationally modified sites available for each protein. For example, in the human PPI network, we find that hub proteins (defined here as proteins with ≥10 partners) have a significantly larger number of confidently identified PTM sites (3.9 vs 1.7; $P = 3.2 \times 10^{-48}$; t-test) compared with non-hub proteins (proteins with <10 partners). Similarly, multifunctional proteins (defined here as those with more than one leaf term in the Gene Ontology) are significantly enriched in confidently identified PTMs (3.4 vs 2.1; $P = 1.1 \times 10^{-16}$, Molecular Function Ontology; 2.8 vs 2.0, $P = 8.5 \times 10^{-7}$, Biological Process Ontology; t-test). Therefore, a larger number of PTM sites could lead to a greater number of PPIs and might be a convenient evolutionary mechanism to create multiple binding interfaces in a single protein.

Allosteric PTMs can facilitate disruption of complexes

Interactions not only form; they also need to break. To date, conformational changes have generally been viewed as important for recognition and binding. Although binding and disruption are mechanistically similar events, the role of conformational changes in disrupting interactions have been largely overlooked. Yet, the allosteric effects elicited by PTMs can lead to dissociation of PPIs nearby and far away (Figures 1b and 2). The strain energy following covalent linkage (or removal) is expected to be larger than that generated by noncovalent events; this is because covalent bonds are stronger and geometrically restricting, and therefore can generate the extra strain required to accommodate them, which can facilitate PTM-elicited allosteric dissociation. The dissociation of ubiquitylated proteins from the adaptor proteins that mediate their association with the scaffold (cullin) in the E3 ligase machinery might provide an example of such an outcome of the allosteric propagation across the substrate protein. The strain generated by the more complex, or multiple [6,7], ubiquitin chains might facilitate dissociation.

Allosteric PTMs can also facilitate disruption of complexes through conformational changes induced by attractive/repulsive van der Waals forces or electrostatic repulsion, similar to FRQ [35] (Figure 4b). The location of the PTM can be near the interaction site and involve more than one PTM. The high mobility group nucleosomal (HMGN) protein family, which regulates chromatin function, illustrates such a scenario. The HMGN2 nucleosome-binding domain binds to the acidic patch in the H2A–H2B dimer and to nucleosomal DNA near the entry/exit point, thus stapling the histone core and the DNA. During mitosis, phosphorylation of Ser24 and Ser28 of the histone core, which are close to the acid patch, induces dissociation of the HMGNs from the nucleosome [56]. The negative charges created by phosphorylation result in unfavorable electrostatic interactions with the acidic patch and

destabilization of the HMGN2–nucleosome complex. Another striking example involves Unc-51-like kinase 1 (Ulk1), a key initiator for mammalian autophagy, which is dramatically dephosphorylated upon starvation and subsequently dissociates from AMPK. Ser638 is dephosphorylated first, followed by Ser758, which is closer to the interaction site. Phosphorylation at Ser638 and Ser758 is crucial for the Ulk1–AMPK association [57]. The proposed role of PTMs in mediating allosterically elicited dissociation leads to questions related to turnover rates [58]. For example, how does the E3 machinery 'know' whether the ubiquitin chain should (or should not) be elongated or branched? Such questions are crucial to the fate of the protein substrate, because they can distinguish between signal transduction and degradation signals, and can control subcellular localization shuttling. If no other factors are involved, whether the substrate is further (poly)-ubiquitylated or dissociates can depend on the relative turnover rates between these (ubiquitylation and dissociation) processes in the substrate–E3 interactions.

Concluding remarks

PTMs are common, frequent, and varied; thus, the number of possible PTM combinations is huge. Further, variants also include different linkages of the same and of different PTM types. This combinatorial complexity provides PTM codes, where each code spells a specific function. PTMs are either at the functional site, or elsewhere, in which case they work via allosteric effects that change the protein surface nearby or far away. To date, attention has largely focused on the mechanisms of PTMs at the functional site and their recognition by specific recognition domains. Here we focus on allosteric PTMs. The variable PTM chemistry (charge, polarity, hydrophobicity), volume, shape, size, and protein environment can lead to hugely different, yet 'specific', allosteric effects. Integration of allosteric PTM combinations with the codes spelt by orthosteric PTMs suggests that the number of possible PTM codes is vastly larger than is currently believed. From an evolutionary standpoint the advantage to the cell is clear: the cell can function with fewer genes, because a single gene product can fulfill many (finely tuned) functions through variations in its PTM code, which can be dynamically regulated. Although a development of a PTM system required the evolution of PTM conjugation and removal enzymes and their regulators, thereby adding genetic complexity, the evolutionary cost is relatively low because any given enzyme can act on many substrates (as shown by the kinases, phosphatases, and the E2/E3 ubiquitylation enzymes). From a cellular standpoint, the key feature is acquiring functional specialization, which is exquisitely achieved by the PTM codes.

The landscape of functional allosteric PTM codes is hugely complex and dynamic, making a precise description of a combinatorial PTM code a daunting challenge. Yet, in a sense there lies its power: it highlights the unlimited number of PTM codes that are available to the organism as it evolves. Additionally, patterns can have similar functional consequences among protein interaction domains, DNA-binding motifs, or homologous architectures. Techniques such as NMR and molecular dynamic simulations, which describe dynamic ensembles of proteins, might also provide some clues. We expect that as more structural and functional data come out, the major PTM combinatorial codes and their functional consequences will be uncovered within their cellular context; and that these play key roles in deciding cell fate. Single and multiple PTMs are known to lead to such effects; allosteric combinatorial codes can refine and better specify distinct functions under varied physiological conditions.

Note added in proof

Additional potential allosteric PTM sites are listed in [59].

Acknowledgments

This project was funded, in whole or in part, with Federal funds from the National Cancer Institute, National Institutes of Health (NIH), under contract number HHSN261200800001E. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, and National Science Foundation (NSF) award DBI-0644017.

References

1 The UniProt Consortium (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.* 37, D169–D174
2 Radivojac, P. *et al.* (2007) Intrinsic disorder and functional proteomics. *Biophys. J.* 92, 1439–1456
3 Meek, D.W. and Anderson, C.W. (2009) Post-translational modification of p53: cooperative integrators of function. *Cold Spring Harb. Perspect. Biol.* 1, a000950
4 Calnan, D.R. and Brunet, A. (2008) The FoxO code. *Oncogene* 27, 2276–2288
5 Benayoun, B.A. and Veitia, R.A. (2009) A post-translational modification code for transcription factors: sorting through a sea of signals. *Trends Cell Biol.* 19, 189–197
6 Deribe, Y.L. *et al.* (2010) Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* 17, 666–672
7 Komander, D. (2009) The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.* 37, 937–953
8 Grabbe, C. *et al.* (2011) The spatial and temporal organization of ubiquitin networks. *Nat. Rev. Mol. Cell Biol.* 12, 295–307
9 Kar, G. *et al.* (2010) Allostery and population shift in drug discovery. *Curr. Opin. Pharmacol.* 10, 715–722
10 Tsai, C.J. *et al.* (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. Biosyst.* 5, 207–216
11 del Sol, A. *et al.* (2009) The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* 17, 1042–1050
12 Pollard, T.D. *et al.*, eds (2007) *Cell Biology*, W.B. Saunders
13 Xu, P. *et al.* (2009) Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* 137, 133–145
14 Jin, L.Y. *et al.* (2008) Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell* 133, 653–665
15 Chen, Z.J. and Sun, L.J. (2009) Nonproteolytic functions of ubiquitin in cell signaling. *Mol. Cell* 33, 275–286
16 Lomelí, H. and Vázquez, M. (2011) Emerging roles of the SUMO pathway in development. *Cell. Mol. Life Sci.* 68, 4045–4064
17 Wilkinson, K.A. and Henley, J.M. (2010) Mechanisms, regulation and consequences of protein SUMOylation. *Biochem. J.* 428, 133–145
18 Kovacs, J.J. *et al.* (2005) HDAC6 regulates Hsp90 acetylation and chaperone-dependent activation of glucocorticoid receptor. *Mol. Cell* 18, 601–607
19 Yang, X.J. (2004) Lysine acetylation and the bromodomain: a new partnership for signaling. *Bioessays* 26, 1076–1087
20 Bergink, S. and Jentsch, S. (2009) Principles of ubiquitin and SUMO modifications in DNA repair. *Nature* 458, 461–467
21 Creton, S. and Jentsch, S. (2010) SnapShot: the SUMO system. *Cell* 143, 848
22 Pichler, A. *et al.* (2005) SUMO modification of the ubiquitin-conjugating enzyme E2-25K. *Nat. Struct. Mol. Biol.* 12, 264–269

23 Astuti, P. *et al.* (2010) Mitotic phosphorylation of Cdc25B Ser(321) disrupts 14-3-3 binding to the high affinity Ser(323) site. *J. Biol. Chem.* 285, 34364–34370
24 Hirschi, A. *et al.* (2010) An overlapping kinase and phosphatase docking site regulates activity of the retinoblastoma protein. *Nat. Struct. Mol. Biol.* 17, 1051–1057
25 Gunasekaran, K. *et al.* (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins Struct. Funct. Genet.* 57, 433–443
26 Sagar, G.D.V. *et al.* (2007) Ubiquitination-induced conformational change within the deiodinase dimer is a switch regulating enzyme activity. *Mol. Cell. Biol.* 27, 4774–4783
27 Ruffieux-Daidie, D. and Staub, O. (2011) Intracellular ubiquitylation of the epithelial Na(+) channel controls extracellular proteolytic channel activation via conformational change. *J. Biol. Chem.* 286, 2416–2424
28 Edreira, M.M. *et al.* (2009) Phosphorylation-induced conformational changes in Rap1b: allosteric effects on switch domains and effector loop. *J. Biol. Chem.* 284, 27480–27486
29 Johnson, L.N. and Lewis, R.J. (2001) Structural basis for control by phosphorylation. *Chem. Rev.* 101, 2209–2242
30 Lee, N.Y. and Koland, J.G. (2005) Conformational changes accompany phosphorylation of the epidermal growth factor receptor C-terminal domain. *Protein Sci.* 14, 2793–2803
31 Zhang, H. *et al.* (2011) Carboxyl-group footprinting maps the dimerization interface and phosphorylation-induced conformational changes of a membrane-associated tyrosine kinase. *Mol. Cell. Proteomics* 10, M110.005678
32 Giannopoulos, P.N. *et al.* (2009) Phosphorylation of prion protein at serine 43 induces prion protein conformational change. *J. Neurosci.* 29, 8743–8751
33 Limardo, R.G.R. *et al.* (2011) p38γ activation triggers dynamical changes in allosteric docking sites. *Biochemistry* 50, 1384–1395
34 Menet, J.S. and Rosbash, M. (2011) A new twist on clock protein phosphorylation: a conformational change leads to protein degradation. *Mol. Cell* 43, 695–697
35 Querfurth, C. *et al.* (2011) Circadian conformational change of the Neurospora clock protein FREQUENCY triggered by clustered hyperphosphorylation of a basic domain. *Mol. Cell* 43, 713–722
36 Tsai, C.J. *et al.* (2008) Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* 378, 1–11
37 Chang, Y.G. *et al.* (2011) Flexibility of the C-terminal, or CII, ring of KaiC governs the rhythm of the circadian clock of cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14431–14436
38 Gong, B.B. *et al.* (2011) Probing structural differences between PrP(C) and PrP(Sc) by surface nitration and acetylation: evidence of conformational change in the C-terminus. *Biochemistry* 50, 4963–4972
39 Hiruma-Shimizu, K. *et al.* (2010) Chemical synthesis, folding, and structural insights into O-fucosylated epidermal growth factor-like repeat 12 of mouse Notch-1 receptor. *J. Am. Chem. Soc.* 132, 14857–14865
40 Walsh, S.T.R. (2010) A biosensor study indicating that entropy, electrostatics, and receptor glycosylation drive the binding interaction between interleukin-7 and its receptor. *Biochemistry* 49, 8766–8778
41 Glozman, R. *et al.* (2009) N-glycans are direct determinants of CFTR folding and stability in secretory and endocytic membrane traffic. *J. Cell Biol.* 184, 847–862
42 Li, S.S. and Shang, Y.F. (2007) Regulation of SRC family coactivators by post-translational modifications. *Cell. Signal.* 19, 1101–1112
43 Sims, R.J. and Reinberg, D. (2008) Is there a code embedded in proteins that is based on post-translational modifications? *Nat. Rev. Mol. Cell Biol.* 9, 815–820
44 Verhey, K.J. and Gaertig, J. (2007) The tubulin code. *Cell Cycle* 6, 2152–2160
45 Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature* 403, 41–45
46 Buratowski, S. (2003) The CTD code. *Nat. Struct. Biol.* 10, 679–680
47 Taverna, S.D. *et al.* (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* 14, 1025–1040
48 Seet, B.T. *et al.* (2006) Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* 7, 473–483
49 Young, N.L. *et al.* (2010) Systems-wide proteomic characterization of combinatorial post-translational modification patterns. *Expert Rev. Proteomics* 7, 79–92
50 Landry, C.R. *et al.* (2009) Weak functional constraints on phosphoproteomes. *Trends Genet.* 25, 193–197
51 Eliuk, S.M. *et al.* (2010) High resolution electron transfer dissociation studies of unfractionated intact histones from murine embryonic stem cells using on-line capillary LC separation: determination of abundant histone isoforms and post-translational modifications. *Mol. Cell. Proteomics* 9, 824–837
52 Flick, F. and Lüscher, B. (2012) Regulation of sirtuin function by posttranslational modifications. *Front. Pharmacol.* 3, 29
53 Young, N.L. *et al.* (2010) Collective mass spectrometry approaches reveal broad and combinatorial modification of high mobility group protein A1a. *J. Am. Soc. Mass Spectrom.* 21, 960–970
54 Jung, H.R. *et al.* (2010) Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Mol. Cell. Proteomics* 9, 838–850
55 Suganuma, T. and Workman, J.L. (2011) Signals and combinatorial functions of histone modifications. *Annu. Rev. Biochem.* 80, 473–499
56 Kato, H. *et al.* (2011) Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12283–12288
57 Shang, L.B. *et al.* (2011) Nutrient starvation elicits an acute autophagic response mediated by Ulk1 dephosphorylation and its subsequent dissociation from AMPK. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4788–4793
58 Turner, B.M. (2008) Simplifying a complex code. *Nat. Struct. Mol. Biol.* 15, 542–544
59 Fuxreiter, M. *et al.* (2011) Dynamic protein-DNA recognition: beyond what can be seen. *Trends Biochem. Sci.* 36, 415–423
60 Radhakrishnan, I. *et al.* (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* 91, 741–752
61 Moriniere, J. *et al.* (2009) Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* 461, 664–668
62 Tsai, K.L. *et al.* (2007) Crystal structure of the human FOXO3a–DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res.* 35, 6984–6994
63 Canagarajah, B.J. *et al.* (1997) Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell* 90, 859–869
64 Vogel, C. *et al.* (2004) Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* 14, 208–216
65 Tsai, C.J. *et al.* (2009) Protein–protein interaction networks: how can a hub protein bind so many different partners? *Trends Biochem. Sci.* 34, 594–600
66 Wu, J.I. *et al.* (2009) Understanding the words of chromatin regulation. *Cell* 136, 200–206
67 Freedman, J.A. *et al.* (2009) A combinatorial mechanism for determining the specificity of E2F activation and repression. *Oncogene* 28, 2873–2881
68 Pan, Y.P. *et al.* (2010) Mechanisms of transcription factor selectivity. *Trends Genet.* 26, 75–83
69 Panne, D. (2008) The enhanceosome. *Curr. Opin. Struct. Biol.* 18, 236–242
70 Pan, Y.P. and Nussinov, R. (2011) The role of response elements organization in transcription factor selectivity: the IFN-β enhanceosome example. *PLoS Comput. Biol.* 7, e1002077
71 Ho, L. and Crabtree, G.R. (2010) Chromatin remodelling during development. *Nature* 463, 474–484
72 Nussinov, R. (2011) How do dynamic cellular signals travel long distances? *Mol. Biosyst.* 8, 22–26
73 Ma, B. *et al.* (2011) Dynamic allostery: linkers are not merely flexible. *Structure* 19, 907–917
74 Csermely, P. *et al.* (2011) Disordered proteins and network disorder in network representations of protein structure, dynamics and function. *Curr. Protein Pept. Sci.* 13, 19–33