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.

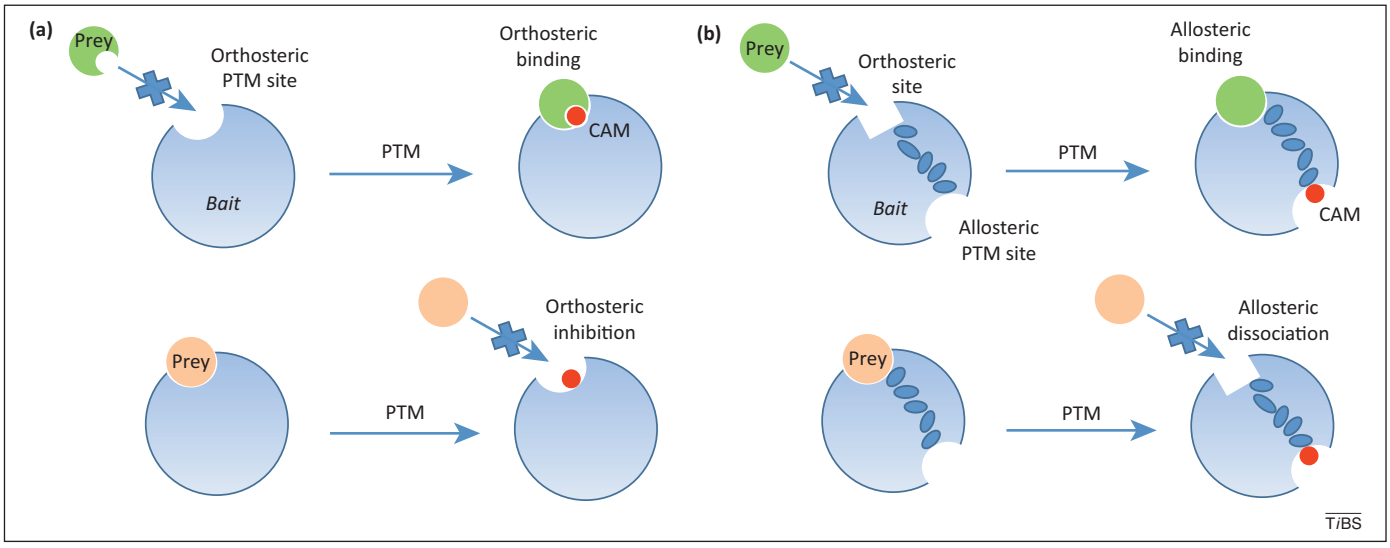

Figure 1. Post-translational modifications (PTMs) can perform their function through two distinct mechanisms. This classification is based on whether the covalently added module (CAM) is attached directly to (orthosteric) or away from (allosteric) the functional site. **(a)** Depicts orthosteric regulation 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