DATABASE
The Journal of Biological Databases and Curation

Database, 2021, 1–20
doi:10.1093/database/baab012
Review

Review

Post-translational modifications in proteins: resources, tools and prediction methods

Shahin Ramazi[1,†] and Javad Zahiri[1,2,3,*,†]

[1]Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics, Faculty of Biological Sciences Tarbiat Modares University, Jalal Ale Ahmad Highway, P.O. Box: 14115-111, Tehran, Iran, [2]Department of Neuroscience, University of California San Diego, La Jolla, CA, USA and [3]Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

*Corresponding author: Email: Zahiri@modares.ac.ir

[†]These authors contributed equally to this work.

Citation details: Ramazi, S., Zahiri, J. Post-translational modifications in proteins: resources, tools and prediction methods. *Database* (2021) Vol. 2021: article ID baab012; doi:10.1093/database/baab012

Received 12 July 2020; Revised 20 February 2021

Abstract

Posttranslational modifications (PTMs) refer to amino acid side chain modification in some proteins after their biosynthesis. There are more than 400 different types of PTMs affecting many aspects of protein functions. Such modifications happen as crucial molecular regulatory mechanisms to regulate diverse cellular processes. These processes have a significant impact on the structure and function of proteins. Disruption in PTMs can lead to the dysfunction of vital biological processes and hence to various diseases. High-throughput experimental methods for discovery of PTMs are very laborious and time-consuming. Therefore, there is an urgent need for computational methods and powerful tools to predict PTMs. There are vast amounts of PTMs data, which are publicly accessible through many online databases. In this survey, we comprehensively reviewed the major online databases and related tools. The current challenges of computational methods were reviewed in detail as well.

Introduction

Posttranslational modifications (PTMs) are covalent processing events that change the properties of a protein by proteolytic cleavage and adding a modifying group, such as acetyl, phosphoryl, glycosyl and methyl, to one or more amino acids (1). PTMs play a key role innumerous biological processes by significantly affecting the structure and dynamics of proteins (2, 3). Generally, a PTM can be reversible or irreversible (4). The reversible reactions contain covalent modifications, and the irreversible ones, which proceed in one direction, include proteolytic modifications (5). PTMs occur in a single type of amino acid or multiple amino acids and lead to changes in the chemical properties of modified sites (6). PTMs usually are seen in the proteins with important structures/functions such as secretory proteins, membrane proteins and histones. These modifications affect a wide range of protein behaviors and characteristics, including enzyme function and assembly (7), protein lifespan, protein–protein interactions (8), cell–cell and cell–matrix interactions, molecular trafficking, receptor activation, protein solubility (9–14), protein folding (15) and protein localization (16).

Therefore, these modifications are involved in various biological processes such as signal transduction, gene expression regulation, gene activation, DNA repair and cell cycle control (17–19). PTMs occur in various cellular organelles including the nucleus, cytoplasm, endoplasmic reticulum and Golgi apparatus (5).

Proximity ligation assay (PLA) is a novel immunoassay technology that can be used to study PTMs (20). In addition to PLA, immunoprecipitation (IP) is utilized in several different PTM detection assays (21). However, the combination of mass spectrometry with IP strategy is a more effective method (22). Nevertheless, large-scale detection of PTMs is very costly and challenging. In recent years, computational methods for predicting PTMs have attracted a considerable attention (5, 16, 17, 23–26).

The rest of this paper is structured as follows. In the section 'The 10 most studied PTMs', the 10 most studied PTMs will be described. Major PTM databases will be reviewed in the section 'The 10 most studied PTMs' as well. In the section 'Involvement of PTMs in diseases and biological processes', involvement of PTMs in diseases and biological processes will be discussed. Then, computational methods for predicting PTMs will be described in the section 'Computational methods for predicting PTMs'. Finally, tools for PTM prediction will be reviewed in the section 'Tools for PTM prediction'.

The 10 most studied PTMs

There are more than 400 different types of PTMs (27) affecting many aspects of protein functions. According to the dbPTM (6), one of the most comprehensive PTM databases, there are 24 major PTMs, with more than 80 experimentally verified reported modified sites. Figure 1 provides a visualized summary of the current major PTM data according to the dbPTM. According to Figure 1, we can see that some of these major PTMs occur more frequently and have much more been studied. Three main PTMs, based on the dbPTM database, are phosphorylation, acetylation and ubiquitination, which comprise more than 90% (~827 000 sites out of ~908 000) of all the reported PTMs Accordingly, each amino acid undergoes at least three different PTMs, and Lys undergoes the largest number of PTMs (15 PTM types). Moreover, based on the whole dbPTM data, Cys and Ser are also modified with at least 10 PTM types. Finally, one can see that phosphorylation on Ser is the most reported PTM type.

Figure 1A shows a clustergram, indicating the division of the PTMs into four clusters as one can see each phosphorylation, and acetylation has been considered as a separate cluster due to their different patterns of modification on the amino acids. On the other hand, ubiquitination, methylation and amidation are the PTMs with many different target residues and have been clustered as a group. According to the clustergram, amino acids have been divided into five clusters. Amino acid Lys is the most different amino acid based on the PTM pattern.

Panels B and C in Figure 1 show the frequency of PTM types and amino acids in the dbPTM database in log scale, respectively. According to Figure 1, it is observed that phosphorylation, acetylation and ubiquitination are the most frequent PTMs.

Roughly speaking, according to the type of the modifications, these PTMs can be categorized into three main groups. First and second groups are those PTMs that include the addition of chemical and complex groups to the target residue, respectively. The first group and the second group include glycosylation, prenylation, myristoylation and palmitoylation. Those PTMs that contain addition of polypeptides to the target residue comprise the last group, and these PTMs are ubiquitylation and SUMOylation. Figure 2 shows a graphical timeline for the discovery of these major PTMs. In this timeline, the organisms in which each PTM was discovered for the first time also have been depicted. In the following subsections, the 10 most studied PTMs, out of these major ones, are described in more detail.

Phosphorylation

Protein phosphorylation was first reported in 1906 by Phoebus Levene with the discovery of phosphate in the protein vitellin (phosvitin) (28). However, it took another 20 years before Eugene Kennedy described the first enzymatic phosphorylation of proteins (43). This process is an important reversible regulatory mechanism that plays a key role in the activities of many enzymes, membrane channels and many other proteins in prokaryotic and eukaryotic organisms (44, 45). Phosphorylation target sites are Ser, Thr, Tyr, His, Pro, Arg, Asp and Cys residues (6), but this modification mainly happens on Ser, Thr, Tyr and His residues (46). This PTM includes transferring a phosphate group from adenosine triphosphate to the receptor residues by kinase enzymes (Figure 3A). Conversely, dephosphorylating or removal of a phosphate group is an enzymatic reaction catalyzed by different phosphatases (47). Phosphorylation is the most studied PTM and one of the essential types of PTM, which often happens in cytosol or nucleus on the target proteins (48). This modification can change the function of proteins in a short time via one of the two principal ways: by allostery or by binding to interaction domains (49).

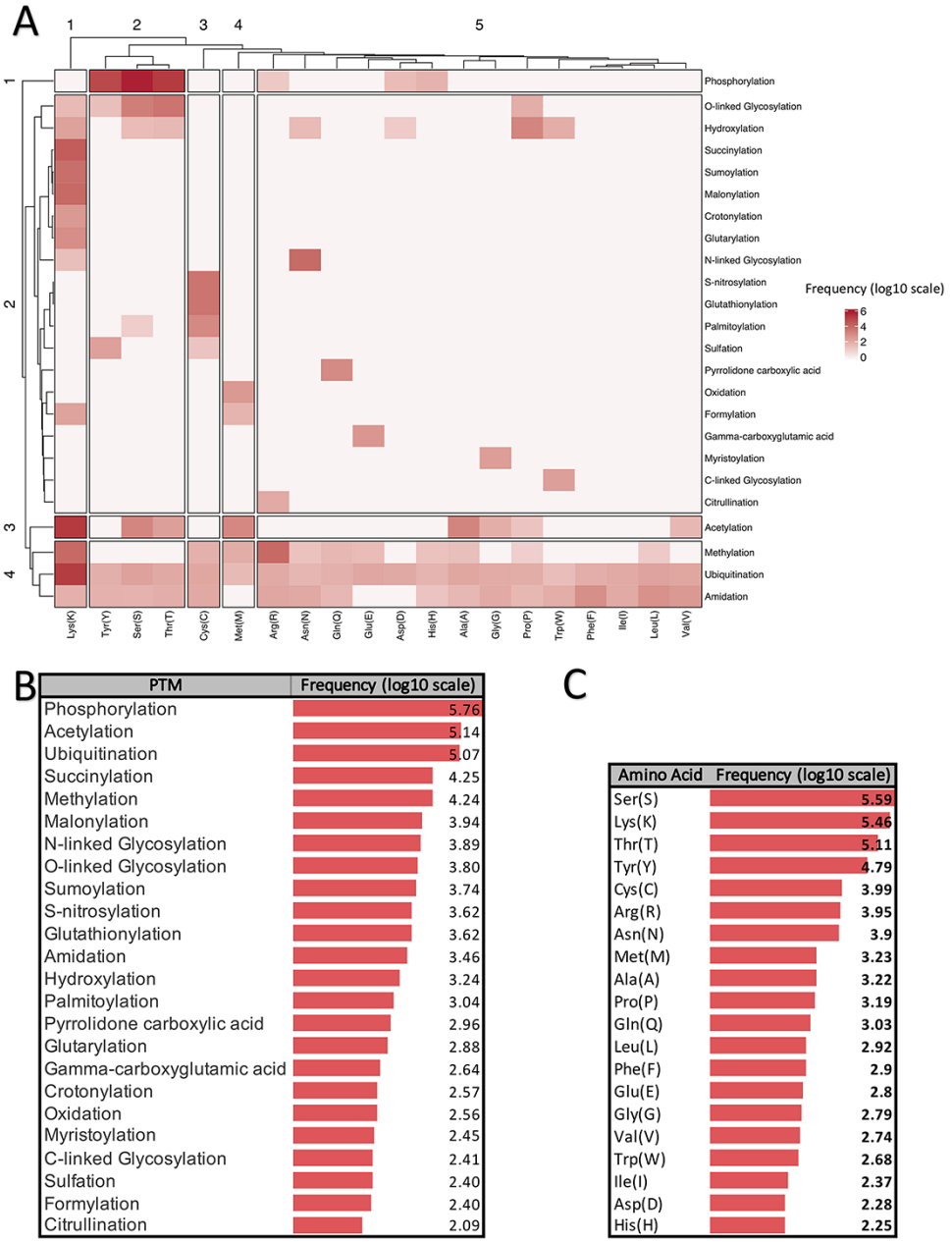

Figure 1. Summarized information of major PTMs (24 PTMs with more than 80 experimentally verified reported modified sites) according to the dbPTM databank (October 2020). All frequencies are shown in log scale. (A) Clustergram indicating the frequency of each PTM on different amino acids. (B) Frequency of major PTMs. (C) Frequency of each amino acid that was reported as a modified site.

Phosphorylation has a vital role in significant cellular processes such as replication, transcription, environmental stress response, cell movement, cell metabolism, apoptosis and immunological responsiveness (12, 50, 51). It has been shown that disruption in the pathway of phosphorylation can lead to various diseases such as cancer, Alzheimer's disease, Parkinson's disease and heart disease (24, 52, 53).

Acetylation

The first acetylation modification in proteins was discovered by V.G. Allfrey in 1964 in isolated calf thymus nuclei *in vitro* (31). Acetylation is catalyzed via lysine acetyltransferase (KAT) and histone acetyltransferase (HAT) enzymes. Acetyltransferases use acetyl CoA as a cofactor for adding an acetyl group (COCH3) to the ε-amino group of lysine side chains, whereas deacetylases (HDACs)

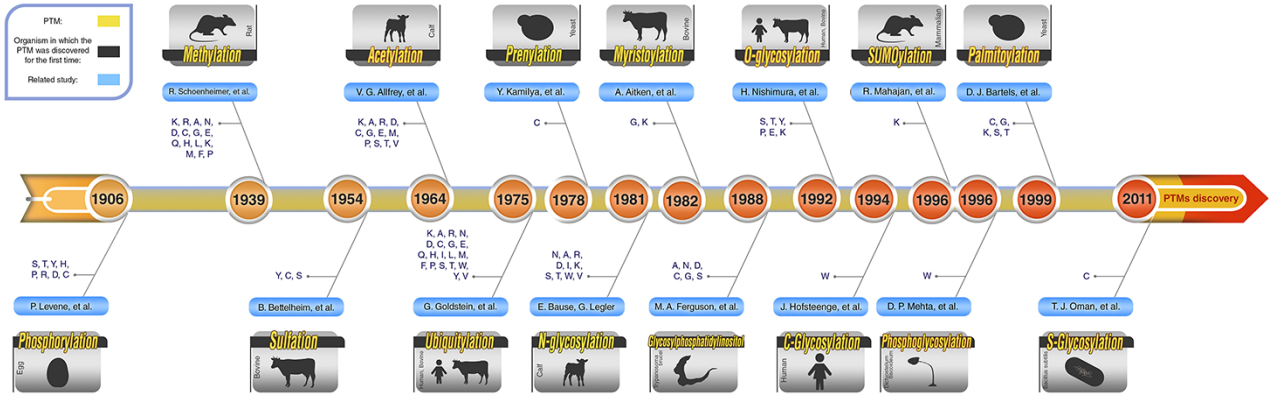

Figure 2. Schematic PTM discovery timeline for 10 major PTMs: phosphorylation (28), methylation (29), sulfation (30), acetylation (31), ubiquitylation (32), prenylation (33), myristoylation (34), SUMOylation (35), palmitoylation (36), different types of glycosylation (N-glycosylation (37), O-glycosylation (38), C-glycosylation (39) and S-glycosylation (40)), phosphoglycosylation (41) and glycosylphosphatidylinositol (GPI anchored) (42). For each PTM, target residue(s) and the organism in which the related PTM was discovered for the first time are shown.

remove an acetyl group on lysine side chains (Figure 3B) (54). There are three forms of acetylation: Nα-acetylation, Nε-acetylation and O-acetylation. Nα-acetylation is an irreversible modification, and the other two types of acetylation are reversible (55). These three forms of acetylation occur on Lys, Ala, Arg, Asp, Cys, Gly, Glu, Met, Pro, Ser, Thr and Val residues with different frequencies (6), although the acetylation is more reported on Lysine residue. Nε-acetylation is more biologically significant compared to the other types of acetylation (55).

Acetylation has an essential role in biological processes such as chromatin stability, protein–protein interaction, cell cycle control, cell metabolism, nuclear transport and actin nucleation (56–58). According to the available evidence, acetylated lysine is vital for cell development, and its dysregulation would lead to serious diseases such as cancer, aging, immune disorders, neurological diseases (Huntington's disease and Parkinson's disease) and cardiovascular diseases (56, 59, 60, 61).

Ubiquitylation

Ubiquitylation is one of the most important reversible PTMs. This modification was firstly studied in 1975 by Gideon Goldstein (32). This modification is a versatile PTM and can occur on all 20 amino acids (Figure 2). However, it occurs on lysine more frequently. This PTM has a major role in the degradation of intracellular proteins via the ubiquitin (Ub)–proteasome pathway in all tissues (62). In ubiquitylation, a covalent bond befalls between the C-terminal of an active ubiquitin protein (a polypeptide of 76 amino acids) and N^{ε} of a lysine residue of the protein (63). Ubiquitin can occur in mono- or poly-ubiquitination forms on substrate proteins through specific isopeptide bonds by receptors containing ubiquitin-binding domains. Ubiquitylation is catalyzed by an enzyme complex that contains ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin ligase (E3) enzymes (Figure 3C). Ubiquitinated proteins may be acetylated on Lys, or phosphorylated on Ser, Thr or Tyr residues, and lead to dramatically altering the signaling outcome (64). Ubiquitylation modification in substrate proteins can be removed by several specialized families of proteases called deubiquitinases (64).

Ubiquitination plays important roles in stem cell preservation and differentiation by regulation of the pluripotency (65). Ubiquitylation has also played a vital role in many various cell activities such as proliferation, regulation of transcription, DNA repair, replication, intracellular trafficking and virus budding, the control of signal transduction, degradation of the protein, innate immune signaling, autophagy and apoptosis (12, 66, 67). Dysfunction in the ubiquitin pathway can lead to diverse diseases such as different cancers, metabolic syndromes, inflammatory disorders, type 2 diabetes and neurodegenerative diseases (68–70).

Methylation

Research on methylation dates back to 1939 (29). Nonetheless, just recently, with the identification of new methyltransferases (such as protein arginine methyltransferases (PRMTs), and histone lysine methyltransferases (HKMTs)), has attracted more and more attention (71). Methylation is a reversible PTM, which often occurs in the cell nucleus and on the nuclear proteins such as histone proteins (1, 72). Methylation occurs on the Lys, Arg, Ala, Asn, Asp, Cys, Gly, Glu, Gln, His, Leu, Met, Phe and Pro residues in target proteins (6). However, lysine and arginine are the two main target residues in methylation, at least in eukaryotic cells (73, 74). One of the most biologically important roles of methylation is in histone modification. Histone proteins, after synthesis of their polypeptide chains, are methylated

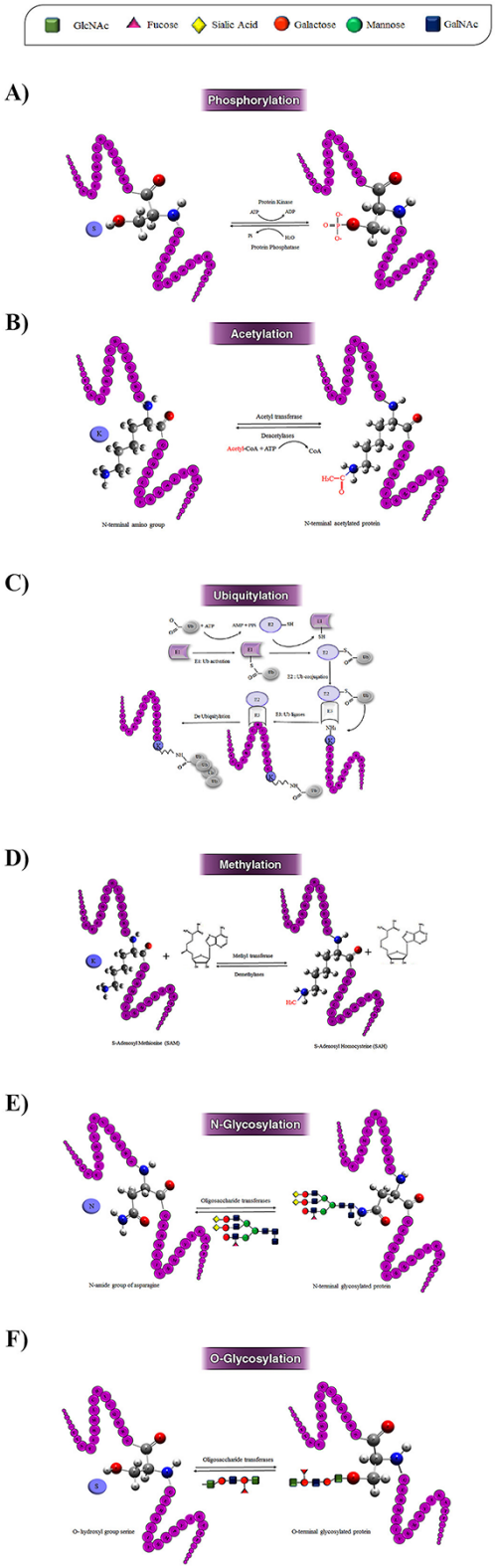

Figure 3. Schematic illustration of the 10 most studied PTMs including Phosphorylation (A), Acetylation (B), Ubiquitylation (C), Methylation (D), N-glycosylation (E), O-glycosylation (F), SUMOylation (G), S-palmitoylation (H), N-myristoylation (I), Prenylation (J), and Sulfation (k).

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.