Protein post-translational modifications: In silico prediction tools and molecular modeling

Martina Audagnotto *, Matteo Dal Peraro *

Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

A R T I C L E   I N F O

Article history:
Received 4 December 2016
Received in revised form 17 March 2017
Accepted 21 March 2017
Available online 31 March 2017

A B S T R A C T

Post-translational modifications (PTMs) occur in almost all proteins and play an important role in numerous biological processes by significantly affecting proteins’ structure and dynamics. Several computational approaches have been developed to study PTMs (e.g., phosphorylation, sumoylation or palmitoylation) showing the importance of these techniques in predicting modified sites that can be further investigated with experimental approaches. In this review, we summarize some of the available online platforms and their contribution in the study of PTMs. Moreover, we discuss the emerging capabilities of molecular modeling and simulation that are able to complement these bioinformatics methods, providing deeper molecular insights into the biological function of post-translational modified proteins.

© 2017 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Post-translational modifications (PTMs) occur on a large number of proteins de facto increasing the actual complexity of the proteome. PTMs consist in a covalent modification of amino acids of the primary protein sequence [1] and have the effect to create a much larger array of possible protein species. In response to specific physiological requirements, PTMs play a crucial role in regulating many biological functions [2], such as protein localization in the cell [3,4], protein stability [5], and regulation of enzymatic activity [6]. To date more than 90,000 individual PTMs were detected using biochemical and biophysical analyses [7]. In particular, it was observed that almost 5% of the human genome encodes enzymes in charge of catalyzing reactions leading to PTMs [8], highlighting once more the importance of these chemical modifications of the proteome.

Enzymes often are responsible for regulating these chemical modifications in proteins, as in the case of phosphorylation, acetylation, methylation, carbonylation or hydroxylation [9]. For instance, protein kinases can phosphorylate a given protein target to induce a signaling cascade, while this PTM can be further removed by specific protein phosphatases. These enzymes are found indeed in important signaling pathways, like G-protein [9,10] and Wnt signaling [11,12]. On the other side, PTMs not induced by specific enzymes (e.g., carbonylation or oxidations) were observed to be responsible of non-specific protein damage involved in neurodegenerative diseases, cancer and diabetes [13–15].

During the past 30 years, experimental techniques used for mapping and quantifying PTMs have seen an impressive progress. In particular, liquid chromatography (LC) with mass-spectrometry (MS) protein-based analysis allowed the detection of thousands of PTMs across entire proteomes [16]. The study of PTMs in their biological context was achieved thanks to advancements in fluorophore chemistry, fluorescence spectrometry, and peptide and antibody synthesis [17]. However, the identification and characterization of PTMs are still limited by the poor knowledge of the underlying enzymatic reactions and their final effects on protein stability and dynamics. In this context, in silico methods, often based on the current knowledge of PTMs, are a promising strategy to perform preliminary analysis and prediction that can guide further in vivo and in vitro experiments, leading to expand our understanding of the role of PTMs in cellular processes.

In this review, we provide an overview of some of the existing computational approaches used to study the most common PTMs, which we classified based on the covalent attachment of (i) small chemical groups, (ii) lipids or (iii) small proteins to the main peptide chain. Most of these tools are presented as online webserver, providing a user-friendly interface for PTM site identification. Although in this review we mainly focus our attention on this kind of resources, standalone software, like for instance PEAKS PTM [18], GlycoMaster [19] or MODa [20], are also available but won’t be covered here.

2. Covalent attachment of small chemical groups

2.1. Phosphorylation

Phosphorylation is the most studied PTM that involves the covalent addition of a small chemical group [21]. It is a reversible enzymatic
reaction, which consists in the attachment of a phosphate group to the side chain of an arginine, lysine, histidine, tyrosine, serine or threonine residue [22] (Fig. 1A). It plays a key role in almost every cellular process, including metabolism, division, organelle trafficking, membrane transport, immunity, learning and memory [23,24], and function of target proteins [25]. It can activate [26,27] and inhibit [28,29] enzyme activity through allosteric conformational changes, facilitate the recognition of other proteins [30–33], promote protein-protein association [34–36] or dissociation [37] and also induce order-to-disorder transition [38,39] (Table 2).

It was estimated that 30% of the total proteome is phosphorylated at least at one residue [40,41]. However, this simple switch mechanism is in reality more complex since multiple enzymes can act on multiple sites of the same protein creating a highly connected network of interactions and modifications. For example, it was shown by high-resolution mass spectrometry that 37,248 phosphorylation sites are present on 5705 proteins in adipocyte cells [42]. Other phospho-proteomic analyses demonstrated that proteins, on average, could be phosphorylated on at least five different sites, although these results could suffer from biases coming from high stoichiometry of the complexes [43–45].

Advances in mass spectrometry, both in terms of speed and sensitivity, allowed identifying and quantifying thousands of phosphorylation sites in different species [42,45]. The conservation of the functional phosphorylation sites in species like mice, rats and flies is a feature used by biologists for selecting specific sites of interest for functional characterization. Therefore, mapping the phosphorylation sites on proteins is an important step in order to understand the catalytic process and the effects of signal transduction events. However, it is still in general difficult to identify specific phosphorylation sites. In silico predictions play an important role in this field. Several methods (Table 1) were implemented in order to predict the target phosphorylation sites from the sequence- and structure-based analysis of the specific protein kinases’ catalytic domain, such as KinasePhos2.0 [46] or GPS [47]. In particular, GPS is a group-based phosphorylation algorithm, which predicts kinase-specific

---

**Fig. 1.** Schematic representation of PTMs discussed in this review.
| PTM type       | Covalent attachment of small chemical groups | Year | Description | Method | Information                                                                 |
|---------------|---------------------------------------------|------|-------------|--------|----------------------------------------------------------------------------|
| Phosphorylation| NetPhos 3.1                                 | 1999 | K-specific and K-independent | ANN    | Prediction based on 17 different kinases                                    |
|               | Scansite                                    | 2003 | K-specific | WM     | Identification of short protein sequence motifs that are recognized by modular signaling domains or mediated specific interaction with proteins |
|               | PhosphoSitePlus                             | 2004 | K-specific | –      | Repository of human and mouse phosphorylation sites                        |
|               | GPS                                         | 2005 | K-specific | –      | Prediction based on 71 PK groups (e.g. Aurora-A, Aurora-B and NIMA)         |
| Glycosylation | KinasePhos 2.0                              | 2007 | K-specific | SVM    | SVM coupled with protein coupling pattern and repository of in vivo and in vitro phosphorylation sites |
|               | PhosphoELM                                  | 2010 | K-specific | –      | Prediction based on evolutionary information                                |
|               | PhosphoPhosphotolog bigPI                   | 2015 | K-independent | –    | Database for cross-species comparison                                       |
|               | O-GlycBase                                  | 1999 | GPF-anchor | KB     | Prediction for proteoma and metazoa                                        |
|               | GlycoMod                                    | 2001 | N,-O-glycosylated | Experimental determined | Match between the experimentally determined masses and the predicted protease (SWISSPROT and TrEMBL databases) |
| S-nitrosylation| YinoYang                                    | 2001 | N,-C,-O-glycosylated | NN    | Prediction based on eukaryote protein sequences                           |
|               | NetNGlyC                                    | 2002 | N-glycosylated | NN    | Prediction for procaryotes and 3D model of glycoproteins based on a PDB structure without attached glycans |
|               | GlyProt                                    | 2005 | N-glycosylated | SWEET-II | Prediction of glycosylation sites and the propensity of association with modified residues |
| Methylations  | iSN-3-PseAAC                                 | 2012 | N-glycosylated | RF     | Combination of different structure and residues pattern information           |
|               | MeMo                                        | 2006 | R, L-methylated | SVM    | Prediction based on orthogonal binary coding scheme for representing protein sequence fragments |
|               | BPP-PPMS                                    | 2009 | R, L-methylated | BPP and SVM | Prediction on experimental data and on the basis of sequence information (SASA and secondary structures) |
|               | BPB-MA                                      | 2009 | K, R, E-N-methylated | SVM | Prediction of physiological properties (binding and disorder information) |
|               | BMes                                        | 2012 | R, K-methylated | SVM    | Differentiation between K-methylated Histone and non-Histone                 |
|                 | MethX                                       | 2014 | K-methylated histone | SVM    | Prediction of physiological proteins, sequence evolution, biochemical and structural disorder information |
| N-acetylation  | PSSMe                                       | 2016 | R-methylated | IGF    | Prediction based on species-specific models and prediction of yeast and mammalian physiological properties |
|               | NetAcet                                     | 2004 | Ne-acetylated | NN     | Prediction based on dataset of 246 acetylated substrates                     |
|               | PAIL                                        | 2006 | Ne, K-acetylated | BD     | Prediction based on protein-protein interaction information                   |
|               | N-Ace                                       | 2010 | K, A, G, M, S- and T-acetylated | SVM | Prediction based on protein-protein interaction information                   |
|               | ASEB                                        | 2012 | K-acetylated | EBM    | Prediction based on protein-protein interaction information                   |

(continued on next page)
phosphorylation sites in 71 protein kinase groups, such as Aurora-A, Aurora-B and Nm23-like protein kinases. Other methods were instead implemented in a way to predict the phosphorylation sites simply from the substrate primary sequences. For example, Scansite [48] is built on combined experimental binding and/or substrate information to derive a weighted matrix-based scoring that predicts protein-protein and protein-phospholipid interactions, as well as phosphorylation sites. NetPhos [49] instead is based on artificial neuronal network that allows the users to choose between a generic predictions based only on the substrate protein sequence or kinase-specific predictions.

Recently, in order to overcome the limitations due to a training set implemented in a way to predict the phosphorylation sites simply from the substrate primary sequences. For example, Scansite [48] is built on combined experimental binding and/or substrate information to derive a weighted matrix-based scoring that predicts protein-protein and protein-phospholipid interactions, as well as phosphorylation sites. NetPhos [49] instead is based on artificial neuronal network that allows the users to choose between a generic predictions based only on the substrate protein sequence or kinase-specific predictions.

2.2. Glycosylation

Protein glycosylation is one of the most relevant and complex post-translational modifications in the cell [54,55], which is thought to influence almost half of all proteins in nature [56]. It consists of a covalent interaction between a glycosyl donor of a glycan and a glycosyl acceptor amino acid side chain of a protein [57] (Fig. 1B). Protein glycosylation can be divided in four main categories based on the linkage between the amino acid and sugar: N-linked glycans, O-linked glycan, GPI anchors and C-mannosylation. In N-glycosylation, a sugar is covalently attached to an amino group of an asparagine [58], while O-mannosylation occurs when an α-D-mannopyranosyl moiety is attached in human and mouse proteomes (PhosphositePlus [52]), as well as rat, fly, yeast and worm (PhosphoELM [53]).

Table 1 (continued)

| PTM type | Covalent attachment of small chemical groups | Year | Description | Method | Information |
|----------|-------------------------------------------|------|-------------|--------|-------------|
| BRABSB-PHKA | http://www.bioinfo.bio.cuhk.edu.hk/bpbbpha/ | 2012 | K-acetylated | Bi-BSB | Prediction for human-specific lysine acetylated sites |
| PTKacePred | http://bioinfo.ncu.edu.cn/inquiries_PTKacePred.aspx | 2012 | K-acetylated | SVM | Prediction based on aminoacid composition, evolutionary similarity and physiochemical properties |
| LAcP | http://www.scbi.org/lPTM/ | 2014 | K-acetylated | LRM | Prediction based on physiochemical properties |

| PTM type | Covalent attachment of acyl chains | Year | Description | Method | Information |
|----------|----------------------------------|------|-------------|--------|-------------|
| Palmitoylation | CSS-Palm | 2008 | Palmitoylated sites | Clustering and scoring algorithm | Prediction for budding yeast |
| CKAAP-Palm | http://ocor.org/wiki/CKSAAP-Palm | 2009 | Palmitoylated sites | SVM | Prediction based on protein sequences |
| N-myristoylation | SeqPalm | 2015 | Palmitoylated sites | SMOT | Correlation between the disruption of palmitoylation sites and diseases |
| NMT | http://mendel.imp.ac.at/myristate/ | 2002 | N-myristoylated sites | PSIC algorithm | Identification of the N-myristoylated sites |
| Prenylation | Myristoylator | 2004 | N-myristoylated sites | NN | Prediction based on first sequences and physical properties |
| PrePS | http://mendel.imp.ac.at/sat/PrePS/ | 2005 | | MCC | Prediction based on first sequences and physical properties |

| PTM type | Small proteins | Year | Description | Method | Information |
|----------|----------------|------|-------------|--------|-------------|
| SUMOHydro | http://protein.caou.edu.cn/others/SUMOHydro/ | 2012 | SUMOHydro | SVM | Prediction based on amino acid hydrophobicity |
| SUMOAMVR | Prediction.aspx | 2014 | SUMOHydro | GPS and PSO | Investigation on the relationship between sumoylation and sumo interaction process |
| JASA | http://www.jasa.fr | 2014 | SUMOHydro | SV | Investigation on the impact of sumo sites in human diseases |
| pSumo-CD | http://www.jci-bioinfo.cn/pSumo-CD | 2016 | SUMOHydro | CD | Identification of database hits, analysis of physicochemical properties and systematic pattern search |
| Ubiquitylation | UbPred | 2009 | Ubiquitylated sites | RF | Prediction based on first sequences and structural information |
| Ubiprober | http://bioinfo.ncu.edu.cn/ubiprober.aspx | 2013 | Ubiquitylated sites | ML | Prediction of species-specific ubiquitinated sites from experimental data |
| Ubiq-Lys | http://www.jci-bioinfo.cn/ubiq-lys | 2014 | Ubiquitylated sites | RF | Prediction based on evolutionary information |
| UbiNet | http://140.138.144.145/~ubinet/index.php | 2016 | Ubiquitylated sites | - | Repository of experimental data, ubiquitylated substrates and protein-protein interactions |

phosphorylation sites in 71 protein kinase groups, such as Aurora-A, Aurora-B and Nm23-like protein kinases. Other methods were instead implemented in a way to predict the phosphorylation sites simply from the substrate primary sequences. For example, Scansite [48] is built on combined experimental binding and/or substrate information to derive a weighted matrix-based scoring that predicts protein-protein and protein-phospholipid interactions, as well as phosphorylation sites. NetPhos [49] instead is based on artificial neuronal network that allows the users to choose between a generic predictions based only on the substrate protein sequence or kinase-specific predictions.

Recently, in order to overcome the limitations due to a training set implemented in a way to predict the phosphorylation sites simply from the substrate primary sequences. For example, Scansite [48] is built on combined experimental binding and/or substrate information to derive a weighted matrix-based scoring that predicts protein-protein and protein-phospholipid interactions, as well as phosphorylation sites. NetPhos [49] instead is based on artificial neuronal network that allows the users to choose between a generic predictions based only on the substrate protein sequence or kinase-specific predictions.
extensively used in glycoproteomics [68–70]. None of these tools can produce, however, a detailed molecular characterization of glycosylated proteins. The high heterogeneity of glycans and the difficulty of obtaining them in large amounts still preclude investigating the role of glycosylation at the molecular level. In the past years, the number of glycoconjugates’ crystallographic protein structures have increased [71–76], nevertheless, a complete chemical and structural description of a glycan structure is still challenging. Mass spectrometry as well as different web-servers (Table 1) currently provide information about existing glycosylation sites. Indeed, in the last decade, several algorithms, trained with sequences or sequence-based information, have been developed to improve prediction of glycosylation sites. Some of these resources are based on neuronal network algorithms, such as NetNGlyc, NetO Glyc [77] for prokaryotes, or YingOYang [78] for eukaryotes. Other useful tools are the GlycoMod [79] server for prediction of glycosyl’s structure based on experimental determined masses, and the NGlyPred [80] server, which incorporates both structure and residue pattern information. More recent developments include prediction of glycosylation sites based on machine learning algorithms (i.e., GlycoMine [81]), an approach that has produced a significant improvement with respect to prediction performances of NetNGlyc [82] and NectO Glyc [77].

### 2.3. S-nitrosylation

S-nitrosylation (SNO) consists in the covalent attachment of a nitric oxide (NO) to cysteine thiol moieties (Fig. 1C). Compared to phosphorylation, SNO is not catalyzed by an enzyme, but it depends on the chemical reactivity between the nitrosylation agent and the target, thus the specific residues’ environment influences the reactivity of the target protein. Concentration of the nitrosylation agent and the protein, as well as the stability of the S—NO bond under physiological conditions, influences in turn the specificity of this reaction.

Over the past 2 decades, hundreds of soluble [83–91] or membrane [92,93] proteins have been identified to be S-nitrosylated. The SNO modification not only modulates protein stability and activities [94,95], but also plays an important role in a variety of biological processes, such as cell signaling, transcriptional regulation, apoptosis and chromatin remodeling [96] (Table 2). Increasing evidences indicate that aberrant S-nitrosylation is implicated in various diseases like cancer [97], Parkinson’s [98,99], Alzheimer’s [100] and amyotrophic lateral sclerosis [101]. Thus the identification of SNO sites in proteins can be also very important for the development of drugs.

Although S—NO bonds are highly labile and redox-sensitive, several techniques managed to detect SNO in cells. There are methods for the direct detection of S-nitrosylated sites, such as the measurement of S—NO characteristic absorbance at 340 nm, electrospray ionization mass spectrometry (ESI-MS) [102] and NMR with $^{15}$N [103]. Ozone chemiluminescence [104] and specific reduction with Cu+/cysteine [105] at pH 6 are indirect chemical methods that are instead based on the analysis of the cleavage products of SNO. Biotin switch assays and chemical reduction/chemiluminescence assays are specific and sensitive methods for measuring low levels of intracellular S-nitrosylated proteins. These experiments are laborious and low-throughput due to the labile nature and low abundance of SNO. Therefore, computational methods represent again a valid alternative to timely and reliably identifying SNO protein sites for further experimental verification. Several benchmark datasets were developed during the past years. SNO SID [106] tests the prediction performance on 65 positive and 65 negative samples, while GPS-NO [107] was developed based on 549 experimentally verified SNO sites. A support vector algorithm machine (SVM) [108] and a nearest neighbor algorithm (NNA) [109] were also proposed to predict SNO sites. However, no web server was later developed for any of these methods, so that their current usage is quite limited. Alternative web-servers are iSNO-PseAAc [110], which identifies nitrosylated proteins on an independent dataset, predicting sites with 90% accuracy [110], and GPS-SNO [107], which also represents a valid tool for an experimentalist providing information for hundreds of potentially S-nitrosylated substrates that have not been yet experimentally determined [110] (Table 1).

### 2.4. Methylation

Protein methylation is a reversible PTM that modifies the nitrogen atoms of either the backbone or side-chain of several types of amino acids, such as lysine, arginine, histidine, alanine and asparagine [111–117]; methylation has been also reported at cysteine residues (S-methylation) [118] (Fig. 1D). Despite this variability, most studies have been predominantly focused on lysine and arginine modifications. Methylation research dates back to 1939, but just recently has attracted more and more attention [111] with the identification of new methyltransferases, like protein arginine methyltransferases (PRMTs) [119–121] or histone lysine methyltransferases (HKMTs) [122–124], which catalyze mono [125] or double [111,126] methylation. In particular, the methylation of the N-terminal tails of the histone plays an important role in gene expression regulation [127], genome stability [128] and nuclear architecture [129] influencing several biological processes such as transcription [130,131] and chromosome maintenance [132] (Table 2). Methylation can also occur on the C-5 position of the cytosine ring of the DNA (DNA methylation) resulting in its association with several human diseases such as cancer, mental retardation (Angelman syndrome) or diabetes mellitus [133]. Although different biological processes are linked to DNA and histone methylation, there seems to be a mutual relationship between these processes, which could play an important role in gene expression [134].

Methylated proteins, as well as methylation regulatory enzymes, are involved in several human diseases such as cancer [126,135,136], cardiovascular diseases [137], multiple sclerosis [138] and neurodegenerative disorders [139]. Thus, the inhibition of these enzymes with small molecules could be an effective therapeutic means of intervention [140]. Moreover, as it is key to identifying methylation sites, understanding methylation mechanistic and dynamic features is as important. In the past years several experimental methods were developed to study the molecular mechanism of methylation. Mutagenesis of potential methylated residues, methylation of a specific antibody [141], as well as Chip-Chip [142] were extensively used for this purpose. Recently, mass spectrometry experiments have been also applied.
allowing the identification of 249 arginine methylated protein sites in 131 proteins from T cells [143]. However, these techniques are usually very expensive and laborious limiting the research of potential methyl-ation sites.

Computational predictions of methylation sites have helped handle these limitations providing an important resource for reducing the number of experiments needed to determine protein methylation sites. Eight web-servers for prediction of methylation sites are currently available (Table 1). MeMo [144] is one of the first online tools to become available. It uses a support vector machine (SVM) as a prediction algo-rithm. Its dataset is based on a curated selection of all methylated resi-dues annotated in SWISS-PROT [145], 264 experimentally manually verified methyl-lysine and 107 methyl-arginine extracted from roughly 1700 scientific articles. MeMo [144] appears to be a powerful tool for predicting methylated-arginine sites when compared to methylated-lysine. However, its accuracy is affected by the lack of training data available at the time of development. Lately, the reliability of the predic-tion was improved by BPP-PFMS [146], where a Bi-Profile Bayesian approach was used to define methylated and non-methylated sites based on known experimental data [147,148]. The data set was increased to 363 candidates containing methylated arginines and 977 methylated lysine proteins. The combination of Bi-Profile Bayesian features with a larger data set improved the methylation prediction accuracy up to 92% for methylated lysine proteins and 88% for methylated arginine proteins [149]. It was observed that protein methylation mainly occurs in regions that are easily accessible and intrinsically disordered, thus MASA [149] used Solvent Accessible Surface Area (SASA) and secondary structure information for predicting methylated sites. This web-server allows the prediction not only of methylated lysines and methylated ar-ginines, but also methyl-glutamates. However, most of these methods use only primary sequence information without taking into account any physicochemical property of residues. With the aim of improving the quality of the prediction, a novel approach called PMes [150] was intro-duced, which considers physicochemical properties of amino acids surrounding methylation sites. A specific lysine-methylation prediction tool for histones was also proposed: METHK [151] uses amino acids’ composition, SASA, amino acid pair composition (i.e., the frequency of amino acid pairs in the primary sequence), amino acid index and protein disorder regions for discriminating between methylated lysine sites in histones and in non-histone proteins. More recently, another web-server has been introduced for in vivo or in vitro species-specific methylation sites’ identification: PSSMe [152] was tested on a large-scale experimental methylated site dataset originated from different species, revealing that methylation patterns are indeed species dependent.

2.5. N-acetylation

Protein acetylation is a covalent post-translational modification where the acetyl group from acetyl coenzyme A (acetyl CoA) is transferred either to the ε-amino group of terminal residues (Nε-acetylation) or to the e-amino group of internal lysine at specific sites (Nε-acetylation) [153–157] (Fig. 1E). Although Nε-acetylation is more common (roughly 85% in eukaryotic proteins), Nγ-lsine acetylation is more biologically important [156–163]. Indeed Nε-acetylation on internal lysines is a reversible post-translational modification involved in several biological processes, such as transcription regulation [159,161], protein expression and stability [153,164–167], DNA repair [162], apoptosis [160,163] and nuclear import [158] (Table 2). Aberrant lysine acetylation is linked with cancer [157,168–170], neurodegenerative disorders [171–173] and cardiovascular diseases [174–178]. Thus, the identification of acetylation sites is important for shedding light on the acetylation mechanism at the basis of numerous diseases [179].

Experimentally several techniques were applied to explore N-acetylation, such as radioactivity detection [180], immunity affinity detection and chromatin immunoprecipitation [181]. The development of high-throughput technologies like immune-precipitation combined to mass spectrometry increased also the number of detected acetylated proteins [182]. However, the experimental detection of acetylated sites is inefficient, expensive and have implicitly low throughput [183]. Therefore, computational tools represent alternative methods for studying the acetylation modifications and provide information for further experiments. Some web-servers (Table 1) dealt only with one specific type of acetylation such as NetAcet [184] for instance. NetAcet [184] attempted to predict only Nε-acetylation sites using a neuronal network trained on yeast data and extendable only to mammalian acetylated substrates. However, NetAcet [184] suffered from the limited size of the training dataset available at that time of development. Several web-servers aimed to predict acetylated lysine. PAIL [185] was the first in silico tool for Nε-lsine acetylation sites’ prediction. The Bayesian discriminant algorithm [186] was employed on a training set of 246 experimentally verified acetylated sites. Despite a small data set, PAIL [185] is able to achieved an accuracy of 85%. BRABSS-PHKA [187] is a human-specific lysine acetylation predictor, which combines a bi-relative adaptive binomial score Bayesian algorithm with a support vector machine. Another method in lysine acetylation prediction is PSKace-Pred [188], where a position-specific view was considered for the characterization of acetylated proteins.

Protein sequences’ information, evolution similarity and physio-chemical properties can help in discriminating between acetyl-lysines and non-acetyl-lysines, improving lysine sites’ evaluation. LACEp [189] is based on a logistic regression model, where the physicochemical property of the amino acids and the transition probability of adjacent amino acids were considered during the prediction process. It also allows predicting acetylated sites not only for lysines, but also for glycine, methionine serine and threonine residues. This is actually done by N-Acet [190], where physicochemical properties (e.g., non-bonded energy, absolute entropy) and solvent accessibility were included in the original prediction code.

The status of lysine acetylation can also be influenced by the en-zymes that catalyze the reaction. Although lysine acetyltransferases (KATs) act usually on a multiple-subunit complex, it is still difficult to determine which KATs are responsible for the acetylation of a given pro-tein. ASEB [191] was the first server for KAT-specific human acetylated lysine prediction that not only evaluates possible lysine acetylation sites, but also provides information about the responsible KAT enzyme.

3. Covalent attachment of acyl chains

Protein lipidation is a unique post-translational modification, which has the result of directly controlling the interaction of soluble protein with biological membranes affecting in turn cellular organization and trafficking. In this section we give an overview of several types of lipidation, their mechanism, involvement in diseases and the computa-tional resources used for predicting lipidation sites.

3.1. Palmitoylation

Palmitoylation consists in the attachment of a 16-carbon acyl chain to cysteine residues via a thioesteric bond [192,193] (Fig. 1F). Among all PTM lipidations, palmitoylation is the only reversible one and can dynamically regulate protein function, as in the case of H-Ras and N-Ras [3,194]. Two families of enzymes regulate the palmitoylation/depalmitoylation process: palmitoyltransferases (PATs), which catalyze the attachment of a palmitate from CoA to specific cysteines, and Acyl Protein Thioesterases (APTs), which remove the palmitate acyl chain. Palmitoylation occurs both in soluble and membrane proteins playing a critical role in the regulation of key biological processes, such as protein membrane trafficking, signaling, cell growth and development [195] (Table 2). Aberrant palmitoylation is associated to a variety of human diseases including neurological disorders (e.g., Huntington disease’s [196] or Alzheimer’s disease [197]) and cancer [198–203].
However, the S-palmitoylated proteome is not yet well defined and little is known about the mechanism that regulates S-palmitoylation and its consequences. In fact, the identification of palmitoylation sites is not simple due to the lack of a distinct sequence motif on the substrates [204]. Mass spectrometry allows the identification of several palmitoylated proteins in cells and tissues, which can be further experimentally characterized using Acyl Biotin Exchange (ABE) or Acyl Resin Assisted Capture (Acyl-RAC) techniques [205–207]. Metabolic labeling and click chemistry probes [208,209] were developed to recognize palmitoylation sites in order to shed light on the molecular mechanism and dynamics of palmitoylation. All these experimental techniques are time and money consuming, thus computer-aided methods are a necessary alternative for predicting palmitoylation sites (Table 1). CSS-palm [210] was one of the first methods to be developed for searching novel palmitoylated proteins in budding yeast. It is based on a clustering and scoring algorithm, where 263 experimentally verified palmitoylation sites are used as a training set, manually collected from the scientific literature. CKSAAP-PALM [211] is another computational method to predict palmitoylation sites based on protein sequences. An encoding scheme composed by k-spaced amino acid pairs is at the basis of this approach [211], which improved accuracy compared to former strategies. SwissPalm [212] has been recently introduced, which provides information from the comparison of different palmitoyl-proteomic studies and allows the users to easily search for the protein of interest, determine the predicted S-palmitoylation sites, identify orthologues and compare them across palmitoyl-proteomes. SeqPalm [213] has been recently developed in order to get insights into the correlation between the disruption of palmitoylation sites and diseases. This new computational method allows for the identification of palmitoylation sites based on amino acid compositions, autocorrelation of amino acid physicochemical properties and amino acid position-weighted matrices.

3.2. N-myristoylation

Myristoylation is a covalent and irreversible attachment of a 14-carbon fatty acid to N-terminal Gly residues [214] of eukaryotic or viral proteins (Fig. 1G). This PTM facilitates in turn the interaction with membranes or a hydrophobic protein domain [215–219]. The substrates involved in myristoylation are generally characterized by the consensus motif Met-Gly-X-X-Ser/Thr at the N-terminus. This PTM acts predominantly by removal of the main methionine residues in order to expose the subsequent glycine [220]. Less frequently, it can also expose an internal glycine by proteases’ cleavage [221]. These mechanisms are both catalyzed by the N-myristoyl transferase (NMT), a 50 kDa enzyme expressed in most organisms [222,223].

Myristoylation is involved in several critical cellular processes, such as signaling pathways, apoptosis [221] and extracellular protein export [224] (Table 2). Usually myristoylation acts with other post-translational modifications like palmitoylation [225–227], or in combination with positively charged residues [228] in order to enhance membrane-protein interactions. Several diseases are linked to N-myristoylation like cancer, epilepsy, Alzheimer’s disease and viral and bacterial infections [229]. The experimental detection of N-myristoylation includes radioactive techniques like the use of 3H or 14C radioactive myristate that requires a long exposure period (weeks to months). To the best of our knowledge only two online web-servers are available to predict myristoylation sites (Table 1). NMT [230] uses a trial set that combines experimentally proved myristoylated proteins with potential myristoylated candidates. Based on structural and biochemical characterization of the N-myristoyl-transferase, a set of descriptors was suggested for better predicting myristoylated sites. This protocol improved the previous pattern suggested in PROSITE [231] (pattern code: PDCC00008), which gave numerous false negative predictions.

Another N-myristoylated site predictor is called Myristoylator [222], which is based on a machine learning model that uses several combined neuronal networks and a test set of positive and negative sequences. Although this predictor seems to increase the specificity, it was trained to predict myristoylation only on terminal glycines, thus a priori knowledge of the proteolytic scission site is necessary when using this web-server.

3.3. Prenylation

Prenylation is a PTM leading to the attachment of a 15-carbon (farnesylation) or a 20-carbon (geranylgeranylation) lipid to cytoines catalyzed by farnesyltransferases or by protein geranylgeranyl transferases I, respectively (Fig. 1H). These isoprenyl anchors promote not only protein-membrane [233–237], but also protein–protein interactions [238–240] (Table 2). Several diseases are correlated to this PTM, like cancer [241,242], premature aging disorders [243,244], neurite [244] and hepatites C and D [245]. Protein prenylation occurs also in a wide range of parasites, leading to the use of protein farnesyltransferase inhibitors in protozoan parasitic diseases [246].

The most common approach for detecting prenylation is to use expensive radiolabeling techniques [203,204]. Initially, the prenylation motif was suggested to be CaaX, i.e. consisting of a cysteine (C) followed by two alphatic residues (aa) and a terminal residue X. However, further kinetic studies and mutation experiments showed a more flexible and complex recognition motif for prenylation [247]. PrePS [248] is the only online tool available, which is based on modeling of the substrate-enzyme interactions for each prenyltransferase.

4. Small proteins’ modifications

An important field in cell signaling is the characterization of the covalent and reversible attachment of ubiquitin (ubiquitylation) and small ubiquitin-related modifiers (sumoylation). This peculiar class of PTMs provides new protein–protein interfaces remodeling the target proteins [249]. In this section we review the latest findings on sumoylation and ubiquitylation with a particular attention on the in silico tools recently developed.

4.1. Ubiquitylation

Ubiquitylation is a three step process where, first, the ubiquitin is activated by a ubiquitin-activating enzyme (E1), then conjugated to a ubiquitin-conjugating enzyme (E2), and finally transferred by a ubiquitin-ligase enzyme (E3) to a substrate molecule via an isopeptide bond with an internal lysine (Fig. 1I). This reversible modification is implicated in the regulation of several cellular processes, like protein degradation [250–252], cell cycle division, the immune response [253], lysosomal trafficking [254] and control of insulin [255] (Table 2). The aberration of ubiquitylation is linked to human pathologies varying from inflammatory neurodegenerative diseases to different forms of cancers [253,256].

Despite the availability of several ubiquitin-protein ligase complex structures [257–261], the ubiquitylation reaction mechanism is still poorly understood. It has been recently hypothesized that structural disorder of the substrate could actually facilitate this process. Analysis of sequences by mutant yeast strain experiments [262] showed that most of the ubiquitylation sites are in the disordered and flexible regions of a protein. On the basis of this observation UbiPred was developed [262] (Table 1): a ubiquitylation site predictor based on a support vector machine algorithm (SVM), which allows studying the correlation between ubiquitylation and protein half-life. In order to overcome the lack of accuracy and training data deficiency, UbiProber [263] and iUbiq [264] were designed (Table 1). UbiProber predicts both general and species-specific ubiquitylation sites using large-scale experimental data as training set, while iUbiq is based on evolutionary information
incorporated into the general form of pseudo-amino acid composition. However, all these in silico tools do not account for E3 binding/recognition sites, although it was shown to be an important feature for ubiquitylation. UbiNet [265] (Table 1) is the first server that allows studying the regulatory network among E3 and ubiquitylated proteins.

4.2. Sumoylation

Sumoylation is a PTM characterized by a covalent attachment of the Small Ubiquitin-like Modifier (SUMO) to specific lysine residues via an enzymatic reaction (Fig. 1L). Sumoylation sites are identified by a canonical consensus sequence Ψ-K-X-E (where Ψ is a hydrophobic amino acid, such as A, L, I, M, P, F, V or W; X any amino acid residue) [266,267], and by SUMO-interacting motifs (3–4 aliphatic residues linked to acid and/or phosphorylatable amino acids) called SIM [268]. Both these features are essential for characterizing the biological significance of sumoylation. This modification is involved in several cellular processes, like protein binding, subcellular transport, gene expression, DNA repair, chromosome assembly and cellular signaling [269–272] (Table 2). Aberrant sumoylation is correlated not only to Alzheimer’s and Parkinson’s diseases [273], but also to cancer [274] and diabetes [275], highlighting the importance of detecting sumoylation sites. Mass spectrometry-based proteomic studies allow mapping hundreds of proteins identifying different sumoylation and SIM sites [276–279]. However, the limitations due to the reversibility of this modification and the difficult identification of peptides from trypsin digestion impose some limitations to the use of this technique. Computer-aided prediction represents a good alternative to reduce the number of potential targets to explore for further experimental verifications (Table 1).

While web-servers available for SIM prediction are only GPS-SUMO [280] and JASSA [281], several online methods for sumoylation sites’ prediction are currently available. SUMOHydro [282] is based on a support vector machine (SVM) combined with amino acid hydrophobicity, while SUMOAMVR [283] considers also other structural features, like average accessible surface area (AASA), secondary structure and evolutionary information of amino acids. Recently, a new in silico tool based on the covariance discriminant (CD) algorithm was developed in order to avoid errors caused by disparity in training data sets [284].

5. PTMs cross-talk

The hypothesis of a correlation between PTMs within the same protein (PTMs cross-talk) [285] has emerged in the proteomic field in recent years. For instance, the regulatory interplay among PTMs was observed for histones [286] and other proteins like p53 [287,288], RNA polymerase II [288] or β-tubulin [289]. In particular, the importance of PTM cross-talk was recognized in several biological pathways (e.g. DNA damage response [290] and protein stability regulation [291–293]) pointing to a strong relationship between PTM cross-talk and protein functions.

While the structural and functional understanding of combinatorial PTMs has been initially limited by technological limitations, recent advances in proteomics have allowed integrating information on different types of modifications [294,295]. However, with the latest experimental methods it is also difficult to identify the whole set of PTM sites in the proteins. In this emerging context, computational methods are poised to support the study of PTM cross-talk. The first unified tool for a simultaneous prediction of PTM sites was ModPred [296], which predicts and analyses simultaneously multiple types of PTM sites in order to gain structural and functional information on the protein regulatory mechanism of multiple PTMs. Recently, a new webserver, PTM-X [297], allows the prediction of PTM cross-talk based on experimentally published data. The difference compared to ModPred is represented by the necessity to know a priori the PTM candidate sites.

6. Structural and dynamical characterization of PTMs

Despite the important role played by PTMs, their structural and dynamical effects on protein function remain poorly understood from a molecular point of view, due to the labile transient nature of most of these modifications and the lack of adequate experimental techniques able to detect and characterize them and the underlying chemical mechanism of formation. The online tools previously discussed are valid methods to overcome some of these experimental limitations and predict putative PTM sites, but they do not usually provide any information about the impact of post-translational modifications from a mechanistic point of view.

Molecular modeling and molecular simulation (such as molecular dynamics, MD), which are based on empirical atomistic force fields [298–301], are a powerful strategy for studying biological systems at single molecule resolution and nanosecond-to-millisecond time scales. Although this computational approach allows nowadays the study of protein processes and properties that are not easily accessible experimentally, there are still some apparent limitations regarding the availability of accurate parameters that would allow the investigation of PTMed proteins. In the past years several improvements have been made in order to expand this approach also to non-standard biomolecules. Within the AMBER force field atomic charges and parameters were developed for phosphorylated residues as phosphoserine, phosphothreonine, phosphotyrosine, phosphohistidine [302], and S-nitrosylated residues (S-nitrocysteine [303]) and methylation (trimethyllysine [304,305]). Similarly, within the CHARMM force field there are parameters for methylated lysines and arginines, as well as acetylated lysines and palmitoylated cysteines [306]. There are also ad hoc comprehensive atomistic force field parameters for treating the description of the link between carbohydrates and proteins such as in GLYCAM for AMBER [307], CHARMM [308] and a modified version of GROMOS [309,310]. In theory, within these schemes, there are existing strategies to parameterize virtually any kind of non-standard amino acids, as for the case of PTMs; in practice, the development of new force field models always involves time-consuming parameterization protocols and rigorous a posteriori validations of the quality and robustness of the new models.

Moving to lower resolution, coarse-grained force fields can also be very useful for studying the impact of PTMs on protein function. In this domain there are no specific parameters for the description of PTMed residues. The Martini force field [311], for instance, provides parameters for treating non-covalently bound sugar molecules or phosphate groups but a complete general representation of modified residues is not yet available. However, a recent work described new parameters for modeling palmitylated cysteines [312] that were used to study H-Ras, and contributed to show the influence of this PTM in regulating the partition of the protein with the membrane.

While for a long time PTMs were not usually considered in modeling and molecular simulation works, the recent availability of more comprehensive experimental data along with accurate force field parameters have thus allowed investigating protein properties taking also into account the effect of PTMs on their structure and stability. Recent examples of this approach have revealed the impact of PTMs for the HIV-1 fusion peptide structure [313], rhodopsin [314], calnexin [315] and phosphatidylinositol 4-kinase [316].

Answering to the need of new and better molecular models to more realistically describe proteins, some automatic tools for generating force field parameters for new molecular species have become available, such as ParaChem or SwissParam [317] compatible with the CHARMM force field, s4md-forcefield tools for AMBER [318] and ATB for GROMOS [319]. However, none of them directly focus on the parameterization of PTMs, likely because of the complexity of the development of parameters required for most PTMs. Therefore, the necessity of having computational tools allowing an automatic parameterization of PTMed protein structures to be used in MD simulations resulted in the
development of some new web-servers, such as FF-PTM (http://selene.princeton.edu/FFPTM/) and Vienna-PTM (http://vienna-ptm.univie.ac.at). FF-PTM focuses on expanding the existing AMBER force field (i.e., ff03) including parameters for 32 PTMs. In particular, it is characterized by parameters that describe the attachment of small molecules (e.g., phosphorylation, methylation or acetylation) and the covalent interaction with acyl chains such as palmitic acid (palmitoylation) and geranylgeranyl pyrophosphate (geranylation). On the other hand, Vienna-PTM is a web platform designed for introducing PTMs on PDB structures to run simulations using the GROMOS 54A7 and 45A3 force fields.

7. Summary and outlook

The chemical modification of amino acids plays an important role in a myriad of cellular processes that range from protein localization to disease development and aging. Enormous efforts, which combine the development of both experimental and computational methods, have been put in the past 2 decades in order to understand PTM mechanisms and effects for protein structure, dynamics and function.

In this review we summarized the main in silico tools mainly available as an online server for studying PTMs (Table 1). Recently, an integrative platform (dbPTM: http://dbptm.mbc.nctu.edu.tw/) has also become available. Originally developed as a comprehensive database of experimentally verified PTMs, dbPTM collects all available databases and webserver resources considering also other PTMs, like succinylation and S-glutathionylation. Although this platform does not provide an exhaustive description for the case of lipidation, it includes also standalone software (not discussed here), offering thus a complementary source of information to this review.

With an increasing amount of experimental data available every day, we think that, as the existing ones will keep improving their performance, many other methods will emerge in the future. Although most of the web-servers available are based on a sequence-based analysis of training data sets, some of them have also started to take into account other interesting properties, such as evolutionary information (e.g., PhosphoOrtholog, ilBiq), SASA (e.g., MASA, METHK, SUMOAMVR) and physicochemical properties (e.g., Pmes, SUMOhydro).

Altogether, these approaches are only rarely considering the molecular features associated with PTMs and the molecular impact they have for protein function in general. Within this context, molecular modeling and simulations have the capability to complement experimental and bioinformatics techniques providing a molecular description of the effect of PTMs on protein structures and stability. However, the lack of suitable tools and parameters for treating PTMs in proteins has limited so far the characterization of these covalent modifications. As some automatic tools (e.g., FF-PTM and Vienna-PTM) have recently appeared providing an online platform to parameterize post-translational modified proteins suitable for running atomistic MD simulations with AMBER or GROMOS force fields, for most PTMs ad hoc parameterizations still need to be developed. Similarly, coarse-grained force fields still lack reliable and robust models for dealing with PTMs, as well as systematic protocols to produce accurate parameters.

Nowadays, with the constant increment of computing power and the availability of always more accurate force fields for biomolecules, which accompany the tireless advances on the experimental side, it is possible to achieve a more precise and faithful description of biological systems in their physiological conditions using molecular modeling and simulation. For instance, the advances in lipidomic analyses have provided a much more detailed view of membrane composition, allowing the construction of more realistic membrane models [320,321] to better investigate membrane biophysical properties and their interplay with integral and peripheral membrane proteins [322,323]. Several computational tools have been developed with this aim in mind, such as CHARMM-GUI [308] and LipidBuilder [324].

Along the same lines, it is also clear that protein PTMs are another important layer of complexity that integrally defines and modulates protein function and, for this reason, needs to be considered at all levels of both experimental and computational investigation. Therefore, the computational advances of bioinformatics and physics-based molecular modeling and simulation techniques appear as a fundamental requirement to complement the experimental investigation of PTMs’ impact on cellular processes.

Acknowledgment

M.D.P. laboratory is supported by the Swiss National Science Foundation (grant number 200020_157153).

References

[1] Walsh C. Posttranslational modification of proteins: Expanding nature’s inventory. Roberts and Company Publishers; 2006.
[2] Walsh CT, et al. Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed 2005;44:7342–72.
[3] Rocks O, et al. An acylation cycle regulates localization and activity of palmitoylated Ras isoforms. Science 2005;307:1746–52.
[4] Fairbank M, et al. RING finger palmitoylation of the endoplasmic reticulum Cyp78 E3 ubiquitin ligase. FEBS Lett 2012;586:2488–93.
[5] Maeda A, et al. Palmitoylation stabilizes unliganded rod opsin. Proc Natl Acad Sci 2010;107:8428–33.
[6] Teer T. Tyrosine phosphorylation: thirty years and counting. Curr Opin Cell Biol 2009;21:140–6.
[7] Khouy GA, et al. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 2011;1:90.
[8] Venne AS, et al. An improved workflow for quantitative N-terminal charge-based fractional diagonal chromatogram (ChaFRACD) to study proteolytic events in Arabidopsis Thaliana. Proteomics 2015;15:2458–69.
[9] Frenmont R, et al. Protein kinases that phosphorylate activated G protein-coupled receptors. FASEB J 1995;9:175–82.
[10] Guevich EV, et al. G protein-coupled receptor kinases: more than just kinases and not only for GPCRs. Pharmacol Ther 2012;133:40–6.
[11] Gao C, et al. Regulation of Wnt/β-catenin signalling by posttranslational modifications. Cell Biosci 2014;4:1.
[12] Jiang W, et al. Histone H3K27me3 demethylases KDM6A and KDM8 mediate distinctive endoderm differentiation from human iSCs by regulating WNT signaling pathway. Cell Res 2013;23:322–30.
[13] Ren R-P, et al. Proteomics of protein post-translational modifications implicated in neurodegeneration. Transl Neurodegener 2014;3:1.
[14] Krueger KE, et al. Posttranslational protein modifications current implications for cancer detection, prevention, and therapeutics. Mol Cell Proteomics 2006;5:1–8.
[15] McLaughlin RJ, et al. Where, how, and when: positioning posttranslational modification within type I diabetes pathogenesis. Curr Diab Rep 2016;16:1–9.
[16] Plowes RS, et al. Mass spectrometry-based proteomics. Nature 2003;422:188–207.
[17] Lothrop AP, et al. Deciphering post-translational modification codes. FEBS Lett 2013;587:1247–57.
[18] Zhang J, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics 2012;11 M111. 010587.
[19] He L, et al. GlycoMaster DB: software to assist the automated identification of N-linked glycopeptides by tandem mass spectrometry. J Proteome Res 2014;13:3881–95.
[20] Na S, et al. Fast multi-blind modification search through tandem mass spectrometry. Mol Cell Proteomics 2012;11 M111. 010199.
[21] Cohen P. The origins of protein phosphorylation. Nat Cell Biol 2002;4:E127–30.
[22] Mandell DJ, et al. Strengths of hydrogen bonds involving phosphorylated amino acid side chains. J Am Chem Soc 2007;129:820–7.
[23] Manning G, et al. The protein kinase complement of the human genome. Science 2002;296:1912–34.
[24] Manning G, et al. Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci 2002;27:514–20.
[25] Johnson LN, et al. Structural basis for control by phosphorylation. Annu Rev Biochem 1993;62:199–223.
[26] Russo AA, et al. Structural basis of cyclin-dependent kinase activation by phosphorylation. Nat Struct Mol Biol 1996;3:696–700.
[27] Hurley J, et al. Regulation of isocitrate dehydrogenase by phosphorylation involves long-range conformational change in the free enzyme. J Biol Chem 1990;265:3599–602.
[28] Weidmann JP, et al. How tyrosine 15 phosphorylation inhibits the activity of cyclin-dependent kinase 2-cyclin A. J Biol Chem 2007;282:3173–81.
[29] Filipiakopoulos P, et al. Structural coupling of SH2–kinase domains links Fes and Abi substrate recognition and kinase activation. Cell 2008;134:793–803.
Lederkremer GZ. Glycoprotein folding, quality control and ER-associated degrada-
tion by the concerted action of protein kinases CK1 and GSK3. Biochemistry 2007;
16:1992–10.

Canagarajah BJ, et al. Activation mechanism of the MAP kinase ERK2 by dual
phosphorylation. Cell 1997;98:859–69.

Piña LA, et al. How do protein kinases recognize their substrates? Biochemi-
ety BioActa Mol Cell Res 1996;134:191–225.

Cohen P. The regulation of protein function by multisite phosphorylation—a 25 year update. Trends Biochem Sci 2000;25:596–601.

Humphrey SJ, et al. Dynamic adipocytokine phosphotyrosine regulates Akt directly
or through mTORC2. Cell Metab 2013;17:1009–17.

Imperiali B, et al. Asparagine-linked glycosylation: specificity and function.

Imanishi T, et al. Structure of the human IgG-Fc fragment in solution and its complex
with a human IgG1 Fc. J Immunol 2000;164:4178–84.

Kagiffy S, et al. Structural insights into the human IgG-Fc glycoforms reveals a correlation
between glycosylation and structural integrity. J Mol Biol 2003;325:979–89.

Ogeyanski V, et al. Structural characterization of a mutated, ADCC-enhanced
human Fc fragment. Mol Immunol 2008;45:1872–82.

Mitsunaka T, et al. Structure of improved therapeutic antibodies on defucosylation of their Fc glycans. Cells Genes 2011;16:1071–80.

Ferrarese A, et al. Chemical dissection of the APC Repeat 3 multistep phosphoryla-
tion and its complex with fragment B of protein A from Staphylococcus aureus.

Bosques CJ, et al. The interplay of glycosylation and disulphide formation in human
protein folding, form and function. Trends Biochem Sci 2006;31:597–607.

Humphrey SJ, et al. Dynamic adipocytokine phosphotyrosine regulates Akt directly
or through mTORC2. Cell Metab 2013;17:1009–17.

Chaudhuri R, et al. PhosphOrtholog: a web-based tool for cross-species mapping of
post-translational modifications using short sequence motifs. Nucleic Acids Res 2003;31:3635–41.

Mannick JB, et al. Three-dimensional structure of the Stat3 signaling kinase.

Olsen JV, et al. Quantitative phosphoproteomics reveals widespread full phosphor-
ylation site occupancy during mitosis. Sci Signal 2010;3:ra3-ra3.

Manning BD, et al. Akt signaling functions. Cell 2002;120:467–480.

Kuang Y-G, et al. Computational prediction of N-linked glycosylation incorporat-
ing structural properties and patterns. Bioinformatics 2012;28:2249–55.

Li F, et al. GlycoMine: a machine learning-based approach for predicting N- and C-
linked glycosylation in the human proteome. Bioinformatics 2015;btu852.

Gupta R, et al. Prediction of glycosylation across the human proteome and the correlation
between structural homology and glycosylation. Mol Biol Cell 2013;24:4733–43.

Chen CW, et al. A molecular redox switch on p21ras structural basis for the nitric
oxide-p21ras interaction. J Biol Chem 1999;274:43233–6.

Hanneder J, et al. Redox regulatory and anti-apoptotic functions of thioredoxin depend
on S-nitrosylation at cysteine 69. Nat Cell Biol 2002;4:743–49.

Li S, et al. Regulation of protein tyrosine phosphatase 1B in intact cells by S-
nitrosothiols. Arch Biochem Biochem Physiol 2003;410:269–79.

Timmer C-M. A single N-linked glycosylation site is required for the activation of the
olfactory cyclic nucleotide-gated channel by NO. J Biol Chem 2000;275:
15315–41.

Eu JP, et al. The skeletal muscle calcium release channel: coupled O2 sensor and NO
signaling functions. Cell 2000;102:499–508.

Derakhshian B, et al. Unbiased identification of cysteine S-nitrosylation sites on pro-	eins. Nat Protoc 2007;2:1685–91.

Tsang AH, et al. S-nitrosylation of XIAP compromises neuronal survival in
Parkinson’s disease. Proc Natl Acad Sci U S A 2006;103:2404–9.

Chen Y-Y, et al. Mass spectrometry-based analyses for identifying and charac-
terizing cysteine S-nitrosylation sites on proteins. Biochemistry 2002;41:13570–70.

Aranda E, et al. Nitric oxide and cancer: the emerging role of S-nitrosylation. Curr
Med Chem 2012;19:2204–21.

Yao D, et al. Nitrosative stress linked to sporadic Parkinson’s disease: S-
nitrosylation of parkin regulates its E3 ubiquitin ligase activity. Proc Natl
Acad Sci U S A 2004;101:10884–88.

Ushara T, et al. S-nitrosylated protein-disulphide isomerase links protein
misfolding to neurodegeneration. Nature 2006;441:513–7.

Cho D-H, et al. S-nitrosylation of Drp1 mediates C-amiﬁd-related mitochondrial
fission and neuronal injury. Science 2009;324:102–5.

Schonhoff CM, et al. S-nitrosothiol depletion in amyotrophic lateral sclerosis. Proc
Natl Acad Sci U S A 2006;103:2404–9.

Chen Y-Y, et al. Mass spectrometry-based analyses for identifying and charac-
terizing S-nitrosylation of protein tyrosine phosphatases. Methods 2007;42:
243–49.

Šinkjařík D, et al. S-nitrosylation of histone deacetylase 2 induces chromatin remodel-
ing in neurons. Nature 2008;455:411–5.

Fang K, et al. Reductive assays for S-nitrosothiols: implications for measurements
of the p21ras NO-glycoengineered SimpleCells with homogeneous GalNAc O-
glycosylation allow iso-

c tructural conformation and its complex with fragment B of protein A from Staphylococcus aureus.

Deisenhofer J. Crystallographic refinement and atomic models of a human Fc frag-
ment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and
2.8-Å resolution. Biochemistry 1981;20:2361–70.
Shao J, et al. Reversing histone methylation sites. Nature 2005;436:1103–7.

Bedford MT, et al. Arginine methylation: an emerging regulator of protein function. Genome Biol 2005;6:45–55.

Boisvert F-M, et al. Protein interfaces in signaling regulated by arginine methylation. Sci STKE 2005;2005:1–10.

Cheng X, et al. Structural and sequence motifs of protein (histone) methylation enzymes. Annu Rev Biophys Biomol Struct 2007;36:473–507.

Fackelmayer FO. Protein arginine methyltransferases: guardians of the Arg? Trends Biochem Sci 2005;30:666–71.

Lee DT, et al. Role of protein methylation in regulation of transcription. Endocrin Rev 2005;26:147–70.

Martin C, et al. The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol 2005;6:38–49.

Lapko VN, et al. Modulation of human Ipa1/i34-crystals includes suppression of gp91(phox), phagocytosis, and transmigration. Protein Sci 2005;14:45–54.

Xie B, et al. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT5 negatively affects Tat interactions with both cyclin T1 and the Tat transactivation region. J Virol 2007;81:4226–34.

Cheng D, et al. The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. Mol Cell 2007;25:71–83.

Torres-Padilla M-E, et al. Histone arginine methylation regulates pluripotency in embryonic stem cells. Genes Dev 2007;32:146–51.

Aletta JM, et al. Protein methylation: a signal event in post-translational modifications. J Biol Chem 2004;279:20691–7.

Barski A, et al. High-resolution profiling of histone modifications in the human genome. Cell 2007;129:823–37.

Porras-Yakushi TR, et al. Yeast ribosomal/cytochrome c SET domain methyltransferase subfamily identification of Rplt3ab methyltransferase sites and recognition motifs. J Biol Chem 2007;282:7688–76.

Shi S-P, et al. PMeS: prediction of methylation sites based on enhanced feature optimization. Bioinformatics 2004;20:1199–206.

Subramanian C, et al. Ku70 acetylation mediates neuroblastoma cell death induced by histone deacetylase inhibitors. Proc Natl Acad Sci U S A 2005;102:4842–7.

Kamata M, et al. N-acetylation of yeast ribosomal proteins and its effect on protein synthesis. J Proteome Res 2011;14:341–41.

Kurdistani SK, et al. Histone acetylation and deacetylation in yeast. Nat Rev Mol Cell Biol 2003;4:276–84.

Kuo M-L, et al. Arginine methyltransferase HMT regulates the AraC/pvR element in the R6/2 mouse model of Huntington’s disease. Neurobiol Dis 2007;25:297–308.

Chen X, et al. The diverse superfamily of lysine acetyltransferases and their roles in inflammation and other diseases. Nucleic Acids Res 2004;32:959–76.

Bannister AJ, et al. Reversing histone methylation. Nature 2005;436:1103–15.

Xie B, et al. Replication of human immunodeficiency virus type 1 Tat transactivation region. J Virol 2007;81:4226–34.

Dong D, et al. Histone methylation mediated by the nuclear amine oxidase hOFG. Nat Cell Biol 2004;6:199–45.

Peters AH, et al. Loss of the Swi39 histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell 2001;107:323–37.

Black JC, et al. Chromatin landscape: methylation beyond transcription. Epigenetics 2011;6:9–15.

Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007;448:533–60.

Santos-Rosa H, et al. Active genes are tri-methylated at K4 of histone H3. Nature 2002;419:407–11.

Black JC, et al. Histone lysine methylation dynamics: establishment, regulation, and biological role. Mol Cell 2012;48:401–507.

Roberts JD. DNA methylation and human disease. Nat Rev Genet 2005;6:597–610.

Cedar H, et al. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 2009;10:295–304.

Aleta JM, et al. Protein methylation: a signal event in post-translational modifications. Trends Biochem Sci 1998;23:89–91.

Paik WK, et al. Historical review: the field of protein methylation. Trends Biochem Sci 2007;32:146–52.

Chen K, et al. Expression of nitric oxide related enzymes in coronary heart disease. Annu Rev Biophys Biomol Struct 2005;34:267–89.

Shao J, et al. Computational identification of protein methylation sites through bio-profile Bayes feature selection. PLoS One 2009;4:e4920.

Grant JE, et al. Post-translational modifications in the rat lumbar spinal cord in ex-perimenterially induced chronic pain. BMC Neurosci 2005;6:39.

Xie B, et al. Replication of human immunodeficiency viruses engineered with heterologous Tat-transactivation response element interactions. J Virol 2003;77:1984–91.

Shen DM, et al. Incorporating structural characteristics for identification of protein methylation sites. J Comput Chem 2009;30:1532–43.

Shi S-P, et al. PMS: prediction of methylation sites based on enhanced feature encoding scheme. PLoS One 2012;7:e38772.

Lee T-Y, et al. Identification and characterization of lysine-methylated sites on histones and non-histone proteins. Comput Biol Chem 2014;50:11–8.

Wen P-P, et al. Accurate in silico prediction of species-specific methylation sites based on information gain feature optimization. Bioinformatics 2016;32:3107–15.

Cisneros MA, et al. Acetylation and deacetylation of non-histone proteins. Gene 2005;363:15–23.

Kouzardes T. Acetylation: a regulatory modification to rival phosphorylation? EMBO J 2000;19:1176–9.

Polevoda B, et al. No-terminal acetylation of eukaryotic proteins. J Biol Chem 2000;275:36479–82.

Polevoda B, et al. The diversity of acetylated proteins. Genome Biol 2002;3:31.

Yang JX. The diverse superfamily of lysine acetyltransferases and their roles in inflammation and other diseases. Nucleic Acids Res 2004;32:959–76.

Bannister AJ, et al. Acetylation of importin-α nuclear import factors by CBP/p300. Curr Biol 2000;10:467–70.

Brunet A, et al. Stress-dependent regulation of FOXO transcription factors by the N-cadherin/Insulin receptor signaling pathway. Proc Natl Acad Sci U S A 2005;102:1682–74.

Koh ET, et al. Glycosylation affects the activity of the human plasma coagulation factor VIII. J Biol Chem 2000;275:40857–63.

Dombrecht JP, et al. Histone deacetylation 6 inhibition compensates for the transport defect in Huntington’s disease by increasing tubulin acetylation. J Neurosci 2007;27:3571–83.

Li A, et al. Prediction of N-terminal acetylation sites. Bioinformatics 2005;21:189–96.

Iyer A, et al. Lysine acetylation in obesity, diabetes and metabolic disease. Immunol Metastasis 2008;25:183–92.

Polevoda B, et al. The diversity of acetylated proteins. Genome Biol 2002;3:31.

Yang JX. The diverse superfamily of lysine acetyltransferases and their roles in inflammation and other diseases. Nucleic Acids Res 2004;32:959–76.

Bannister AJ, et al. Acetylation of importin-α nuclear import factors by CBP/p300. Curr Biol 2000;10:467–70.

Brunet A, et al. Stress-dependent regulation of FOXO transcription factors by the N-cadherin/Insulin receptor signaling pathway. Proc Natl Acad Sci U S A 2005;102:1682–74.

Koh ET, et al. Glycosylation affects the activity of the human plasma coagulation factor VIII. J Biol Chem 2000;275:40857–63.

Dombrecht JP, et al. Histone deacetylation 6 inhibition compensates for the transport defect in Huntington’s disease by increasing tubulin acetylation. J Neurosci 2007;27:3571–83.

Li A, et al. Prediction of N-terminal acetylation sites. Bioinformatics 2005;21:189–96.

Iyer A, et al. Lysine acetylation in obesity, diabetes and metabolic disease. Immunol Metastasis 2008;25:183–92.

Polevoda B, et al. The diversity of acetylated proteins. Genome Biol 2002;3:31.

Yang JX. The diverse superfamily of lysine acetyltransferases and their roles in inflammation and other diseases. Nucleic Acids Res 2004;32:959–76.

Bannister AJ, et al. Acetylation of importin-α nuclear import factors by CBP/p300. Curr Biol 2000;10:467–70.

Brunet A, et al. Stress-dependent regulation of FOXO transcription factors by the N-cadherin/Insulin receptor signaling pathway. Proc Natl Acad Sci U S A 2005;102:1682–74.

Koh ET, et al. Glycosylation affects the activity of the human plasma coagulation factor VIII. J Biol Chem 2000;275:40857–63.

Dombrecht JP, et al. Histone deacetylation 6 inhibition compensates for the transport defect in Huntington’s disease by increasing tubulin acetylation. J Neurosci 2007;27:3571–83.
[271] Müller S, et al. SUMO, ubiquitin's mysterious cousin. Nat Rev Mol Cell Biol 2001;2:202–13.
[272] Seeler JS, et al. Nuclear and unclear functions of SUMO. Nat Rev Mol Cell Biol 2003;4:690–9.
[273] Lee L, et al. SUMO and Alzheimer's disease. Neuromolecular Med 2013;15:720–36.
[274] Zhao J. SUMoylation regulates diverse biological processes. Cell Mol Life Sci 2007;64:3017–13.
[275] Li M, et al. SUMO wrestling with type 1 diabetes. J Mol Med 2005;83:504–13.
[276] Blomster HA, et al. Novel proteomics strategy brings insight into the prevalence of SUMO-2 target sites. Mol Cell Proteomics 2009;8:1382–90.
[277] Veerbeek JM, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. PLoS One 2014;9:e87987.
[278] Impens F, et al. Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli. Proc Natl Acad Sci 2014;111:12432–7.
[279] Matic I, et al. Site-specific identification of SUMO-2 targets in cells reveals an inverted SUMOylation motif and a hydrophobic cluster SUMOylation motif. Mol Cell 2010;39:641–52.
[280] Zhao Q, et al. GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. Nucleic Acids Res 2014;42:W325–30.
[281] Beauchair G, et al. JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. Bioinformatics 2015;31:3483–91.
[282] Chen Y-Z, et al. SUMOhydro: a novel method for the prediction of sumoylation sites and hydrophobic properties. PLoS One 2012;7:e39105.
[283] Xu H-D, et al. Systematic characterization and prediction of post-translational modification cross-talk. Mol Syst Biol 2013;9:714.
[284] Strahl BD, et al. The language of covalent histone modifications. Nature 2000;403:41–5.
[285] Meek DW, et al. Isoform-specific determinants in the HP1 binding to histone 3: insights from molecular simulations. Amino Acids 2010;38:1571–81.
[286] Graefel C, et al. Force field parameters for the simulation of modified histone tails. J Comput Chem 2010;31:2434–51.
[287] Woods RJ, et al. Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. J Phys Chem B 1995;99:3832–46.
[288] Zhao S, et al. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 2008;29:1859–65.
[289] Schmid N, et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J 2011;40:843–56.
[290] Soares T, et al. Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme. J Biomol NMR 2004;30:407–22.
[291] Marrink SJ, et al. The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 2007;111:7812–24.
[292] de Jong DH, et al. Molecular view on protein sorting into liquid-ordered membrane domains mediated by gangliosides and lipid anchors. Faraday Discuss 2013;161:347–63.
[293] Stewart-Jones GB, et al. Trimeric HIV-1-Env structures define glycan shields from clades A, B, and G. Cell 2016;165:813–26.
[294] Olausson BE, et al. Molecular dynamics simulations reveal specific interactions of post-translational palmitoyl modifications with rhodopsin in membranes. J Am Chem Soc 2012;134:4324–31.
[295] Lukkarja AK, et al. Palmitoylated calnexin is a key component of the ribosome-translocon complex. EMBO J 2012;31:1823–35.
[296] Zhou Q, et al. Molecular insights into the membrane-associated phosphatidylinositol 4-kinase IIα. J Mol Biol 2007;368:190–202.
[297] Zoete V, et al. SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 2011;32:2359–68.
[298] Vanquelef E, et al. RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. Nucleic Acids Res 2011;39:W511–7.
[299] Malde AK, et al. An automated force field topology builder (ATB) and repository: version 1.0. J Chem Theory Comput 2011;7:4026–37.
[300] van Eerden FJ, et al. Characterization of thylakoid lipid membranes from cyanobacteria and higher plants by molecular dynamics simulations. Biochim Biophys Acta Membr 2015;1848:1319–30.
[301] Ingoldsson H, et al. Lipid organization of the plasma membrane. J Am Chem Soc 2014;136:16545–9.
[302] Audagnotto M, et al. Effect of the synaptic plasma membrane on the stability of the amyloid precursor protein homodimer. J Phys Chem Lett 2016;7:3572–8.
[303] van Eerden FJ, et al. Molecular dynamics of Photosystem II embedded in the thylakoid membrane. J Phys Chem B 2015;119:12432–40.