Current Synthetic and Systems Biology

Hasan MM et al., Curr Synthetic Sys Biol 2018, 6:1
DOI: 10.4172/2332-0737.1000137

Commentary Open Access

Computational Modeling of Lysine Post-Translational Modification: An Overview

Md. Mehedi Hasan[1]*, Mst. Shamima Khatun[2], and Hiroyuki Kurata[1,3]

[1]*Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan*
[2]*Department of Statistics, Laboratory of Bioinformatics, Rajshahi University-6205, Bangladesh*
[3]*Biomedical Informatics R&D Center, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan*

Commentary

Living organisms have a magnificent ordered and complex structure. In regulating the cellular functions, post-translational modifications (PTMs) are critical molecular measures. They alter protein conformation, modulating their activity, stability and localization. Up to date, more than 300 types of PTMs are experimentally discovered *in vivo* and *in vitro* pathways [1,2]. Major and common PTMs are methylation, ubiquitination, succinylation, phosphorylation, glycosylation, acetylation, and sumoylation.

PTM is a biological mechanism common to both prokaryotic and eukaryotic organisms, which controls the protein functions and stability or the proteolytic cleavage of regulatory subunits and affects all aspects of cellular life. The PTM of a protein can also determine the cell signaling state, turnover, localization, and interactions with other proteins [3]. Therefore, the analysis of proteins and their PTMs are particularly important for the study of heart disease, cancer, neurodegenerative diseases and diabetes [4,5]. Since the characterization of PTMs gets invaluable insight into the cellular functions in etiological processes, there are still challenges. Specifically, the major challenges in studying PTMs are the development of specific detection and purification methods.

The PTMs are categorized into several axes. The first one is grouped by the residue side-chains of modification sites. In this category, almost 15 of the 20 types of amino acid side-chains can undergo the modifications (Table 1) [6,7]. The second one is a fragment of coenzyme or co-substrate coupled to the protein and concomitant modification by chemical nature, including S-adenosylmethionine dependent methylation, acetyl CoA dependent acetylation, NAD-dependent ADP ribosylation, CoASH-dependent phosphopantetheinylation, ATP-dependent phosphorylation, and phosphoadenosinephosphosulfate (PAPS) -dependent sulfurylation. The third categorization of PTM is grouped by the hydrophobic residues for membrane localization. It has acquired various lipid modifications [prenylation, glycosyl phosphatidylinositol (GPI), palmitoylation anchor attachment, glypiation, farnesylation, geranylgeranylation]. However, many PTMs have biased and overlapped with the arrangement of other PTMs in surrounding amino acid sequences. This tendency is often embodied within a sequence motif. For example, it has been observed that nearly 60% class PTMs of protein succinylation sites are surrounded or overlapped with protein acetylation sites [8]. These PTMs can also affect the physicochemical properties of proteins, which can provide a mechanism for the dynamic regulation of molecular, self-assembly [9]. The PTMs have been found in the all types of proteins such as the structural proteins, plasma membrane receptors and nuclear transcription factors (Figure 1).

Lysine is one of the most frequently occurred PTM sites, which has important regulatory and functional consequences. In 1964, Allfrey et al. [10] observed that the gene expression can be regulated by covalently introducing methyl and acetyl groups on lysine residues in histones. Recently, some studies have discovered that lysine acts as a hot spot for PTMs, and a number of protein lysine modifications could occur in both histone and non-histone proteins [11,12]. For instance, lysine methylation in non-histone proteins can regulate the protein activity and protein structure stability [13]. In 2004, the Nobel Prize in Chemistry was awarded jointly to Aaron Ciechanover, Avram Hershko and Irwin Rose for the discovery of lysine ubiquitin-mediated protein degradation [14].

Moreover, in biological process, lysine can be modified by the primary glycolytic intermediate 1,3-biophosphoglycerate (1,3-BPG) through 3-phosphoglyceryl-lysine protein [15], whereas in glycolytic processes lysine glycation is involved [16]. A rapid progress in proteomic technologies have greatly accelerated the identification of lysine modifications proteins and the discovery of new lysine PTMs [11,12,17,18]. Therefore, it is urgently needed to know the function of these lysine PTMs, since the number of lysine PTMs have been greatly expanded to the research community. Moreover, it is also essential to create lysine PTM databases for researchers to store, query and manage the lysine PTM data.

Computational prediction of lysine PTM sites

Some kinds of PTMs, such as succinylation, ubiquitination, acylation, methylation, sumoylation and deamination, occurred on lysine residues. For the last several decades remarkable progresses have been carried in the identification and functional analysis of lysine PTMs in proteins. Lysine PTMs play a vital role in protein folding, protein function, and interactions with other proteins [19,20]. Due to the important biological functions of protein lysine PTMs, it is very important to analyze and understand the function of lysine PTMs.

The lysine PTMs of proteins have been identified by a variety of experimental techniques including the mass spectrometry (MS) [21,22], chromatin immunoprecipitation (ChIP)[23], liquid chromatography [24], radioactive chemical method [25], western blotting [26], and eastern blotting [24]. The MS technique is one of the mainstay routes in detecting PTMs in a high-throughput manner. The new MS and capillary liquid chromatography instrumentation have made revolutionary advance in enrichment strategies in our growing understanding of many PTMs [27]. A similar strategy of fragmentation

***Corresponding author:** Md. Mehedi Hasan, Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan, Tel: +81-(0)93-884-3061, E- mail: mehedicau@hotmail.com

Received February 10, 2018; **Accepted** February 12, 2018; **Published** February 15, 2018

Citation: Hasan MM, Khatun MS, Kurata H (2018) Computational Modeling of Lysine Post-Translational Modification: An Overview. Curr Synthetic Sys Biol 6: 137. doi:10.4172/2332-0737.1000137

Residues	Reactions	Example
Asp (D)	phosphorylation isomerization to isoaspartyl	Protein tyrosine phosphatases; response regulators in two- component systems
Glu (E)	Methylation,	Chemotaxis receptor proteins
	Carboxylation	γ-carboxyglutamyl residues in blood coagulation
	Polyglycination	Tubulin
	Polyglutamylation	Tubulin
Ser (S)	Phosphorylation	Protein serine kinases and phosphatases
	O-glycosylation	Notch O-glycosylation
	Phosphopantetheinylation	Fatty acid synthase
	Autocleavages	Pyruvamidyl enzyme formation
Thr (T)	Phosphorylation O-glycosylation	Protein threonine kinases/phosphatases
Tyr (Y)	Phosphorylation	Tyrosine kinases/phosphatases
	Sulfation	CCR5 receptor maturation
	Ortho-nitration	Inflammatory responses
	TOPA quinone	Amine oxidase maturation
His (H)	Phosphorylation aminocarboxypropylation	Sensor protein kinases in two-component regulatory systems
	N-methylation	Diphthamide formation / methyl CoM reductase
Lys (K)	N-methylation	Histone methylation
	N-acylation by acetyl, biotinyl, lipoyl, succinyl, ubiquityl groups	Histone acetylation; swinging-arm prosthetic groups; ubiquitin; sumo (small ubiquitin-like modifier) tagging of proteins
	C-hydroxylation	Collagen maturation
	Crotonylation	Histone lysine modification
	Pupylation	Prokaryotic ubiquitin like protein degradation protein
Cys (C)	S-hydroxylation (S-OH)	Sulfenate intermediates
	Disulfide bond formation	Protein in oxidizing environments
	Phosphorylation	Protein tyrosine phosphataseras isoforms
	S-acylation	ras isoforms
	S-prenylation protein splicing	Intein excisions
Met (M)	Oxidation to sulfoxide	Met sulfoxide reductase
Arg (R)	N-methylation	Histones
	N-ADP-ribosylation	the α-subunit of $G_{s\alpha}$
Asn (N)	N-glycosylation	N-glycoproteins
	N-ADP-ribosylation	Etheno-ADP-ribosylation
	Protein splicing	Intein excision step
Gln (Q)	Transglutamination	Protein cross-linking
Trp (T)	C-mannosylation	plasma-membrane proteins
Pro (P)	C-hydroxylation	collagen; hypoxia-inducible factor 1
Gly (G)	C-hydroxylation	C-terminal amide formation

Table 1: Specification of protein PTMs grouped by residue side-chains [6,7].

for PTM identification is the beam-type collision induced dissociation, also called higher energy collisional dissociation [28]. These types of fragmentation are characterized by higher activation energy. Most of the fragmentation methods of precursor ions are based on the radical anions or thermal electrons [29]. In general, the experimental analysis of PTMs often requires labor intensive sample preparations and hazardous or expensive chemical reagents. The substrate is separated from non-radioactive ATP by the kinase assay and generates radioactive waste. Since most of the radioactive substances show a short half-life, the fresh reagent must be frequently acquired. And sometimes, the substrate concentration of assay is often much higher than the expected substrate concentrations [30]. In the above discussion we can summarize that, the identification of PTMs by the experimental techniques is laborious, time-consuming and usually expensive.

In contrast with the traditional experimental methods, computational analysis of lysine PTMs has also been an attractive and alternative approach due to its accuracy, cost-effective and high-speed [31,32]. The computational methods are more efficient for identifying large-scale novel lysine PTM substrates. A summary of the prediction pipeline of lysine PTM is shown in Figure 2. The computational tools can narrow down the number of potentially candidates and rapidly generate useful information for investigating further experimental approach. So far the computational prediction of protein lysine PTMs has been an important research topic in the field of protein bioinformatics. Although the great progress has been made by employing various statistical learning approaches with numerous feature vectors, a problem is to obtain more accurate prediction. It needs rigorous features encoding methods, machine learning, and statistical analysis to predict lysine PTMs. Indeed, computational method development of lysine PTM site prediction has initiated since 2008 [33]. In the next section, we will introduce some existing databases for lysine PTMs.

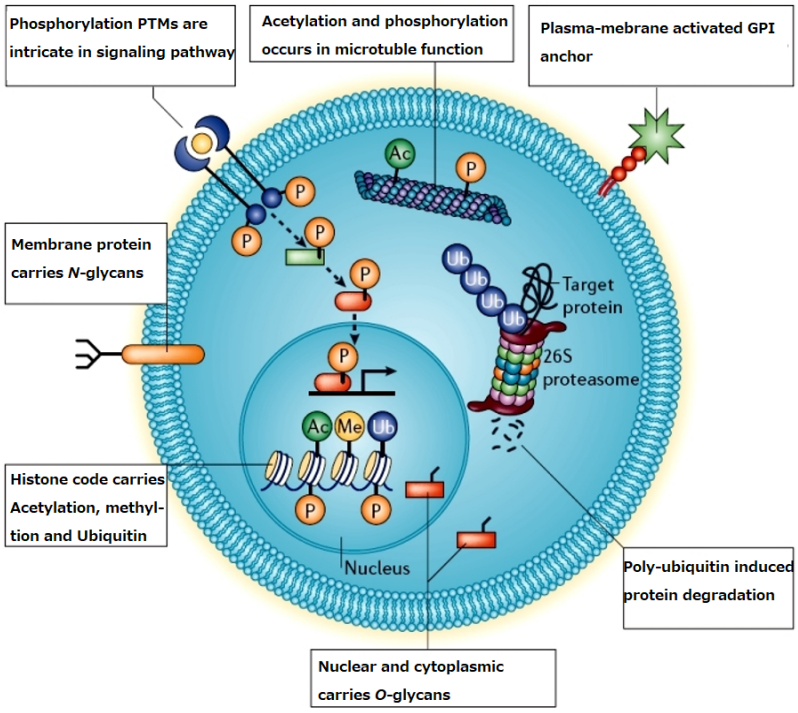

Figure 1: Important locations for different types of PTMs in a cell [68].

Databases of lysine PTM sites

Recently, rapid progresses in proteomic technologies have greatly accelerated the identification of well-characterized lysine PTM sites. Determinations of lysine PTM data with experimental technologies are also greatly extended. How to organize, store and update these data becomes an important issue. Up to now, a number of experimentally verified lysine PTM databases have been constructed (Table 2). For instance, the CPLM is a lysine PTM database that integrates abundant protein annotations [34]. In total, the CPLM database contained 45,748 lysine modification proteins with 189,919 experimentally verified lysine modification sites for 122 species (CPLM 1.0). It is expected that huge data will be generated from lysine PTMs in the future. Therefore, the diversity of protein lysine PTMs requires specialized databases to store them. Based on lysine PTM databases, many bioinformatics methods have been developed for analyzing the internal motif [35,36]. It is becoming a hot topic in the study of protein bioinformatics.

Feature for the computational prediction of lysine PTM sites

Feature mining is one of the most important steps for predicting lysine PTM sites. Appropriate features in the prediction model enable the accurate prediction of protein lysine PTMs. In general, the feature vectors refer to the characterization of the sequences and local structures around the protein functional sites. Ideally, the features can clearly distinguish PTM sites from other random sites. In the real world, however, the feature of protein functional sites can also exist on the non-functional sites of proteins. In the prediction PTM sites, this specific problem is particularly prominent due to the sequence diversity. For instance, some sequence motifs are very weak and not available with the sequence evolutionary information [37-42]. To address this problem, we can search PSI-BLAST [32,43,44] against the NCBI NR database to generate a profile (i.e., position-specific scoring matrix [45-50]. Such a sequence profile reflects the conservation and variation between protein sequences through the evolutionary information [37-39]. Moreover, to isolate the weak motifs from protein sequences, Hidden Markov models (HMMs) have been extensively used [51,52]. It can examine the unaligned sequences or a common motif within a set of unaligned sequences. HMM profiles can be automatically trained or estimated, from unaligned protein sequences [51].

In the prediction of lysine PTMs, researchers have made plenty of efforts for mining the protein lysine PTM characteristics. These characteristics might be suitable for a particular protein lysine PTM classification problem, thus mining new features is always an important task for lysine PTM prediction. The features are mostly encoded by three ways, namely based on the protein sequence, evolutionary, and structural information (Figure 2). In most cases, the features are extracted from protein sequences because the protein sequence data is more enthusiastically available than the protein structure data. In addition, the features based on protein sequences are often straightforward and the simplest features. For instance, the linear arrangement of residues directly depicts the flanking sequences of lysine PTM sites. In the linear arrangement of residues, the physicochemical amino acid index properties have also been widely used in the prediction of protein lysine PTM sites [53,54].

Algorithm of lysine PTM sites prediction

After determining the appropriate features, the next job is to select an appropriate machine learning algorithm to integrate these features for the prediction of protein lysine PTM sites. Generally, machine learning model used for building the trained model to test the novel dataset. It will improve the accuracy of the prediction if the prediction algorithm is appropriate. These prediction algorithms of lysine PTM sites can be classified into two categories, i.e., statistical probabilistic algorithms and machine learning algorithms. In the next sections, we will discuss some of the probabilistic and machine learning classifiers for lysine PTM prediction.

Database	Web-site	Description
dbPTM [7]	http://dbPTM.mbc.nctu.edu.tw/	dbPTM is an integrated resource database for protein PTM, which was collected from the biological databases in the public domain such as Swiss-Prot, PhosphoELM, O-GlycBase, and UniProtKB protein entries, etc.
SuccinSite [31]	http://systbio.cau.edu.cn/SuccinSite/	Hasan *et al.* created a lysine succinylation site database, which had integrated five succinylation families' dataset.
SysPTM2 [69]	http://lifecenter.sgst.cn/SysPTM/	Li *et al.* created a database SysPTM 2.0, which was integrated into two datasets, SysPTM-A and SysPTM-B. This database was collected from the public data resources and peer reviewed MS/MS literature, respectively.
CPLM [34]	http://cplm.biocuckoo.org	Liu *et al.* created a lysine modification database term as CPLM, which consisted 12 types of lysine PTM, including acetylation, butyrylation, crotonylation, glycation, malonylation, phosphoglycerylation, propionylation, ubiquitination, sumoylation, methylation, succinylation and pupylation.
PupDB [70]	http://cwtung.kmu.edu.tw/pupdb/	PupDB is a most popular database for protein pupylation sites, which was constructed by collecting the experimentally identified pupylated proteins and pupylation sites from the published studies. Until now, it is an updated database for lysine pupylation sites.
DbPTM 3.0[71]	http://dbptm.mbc.nctu.edu.tw/	Lu *et al.* developed an informative resource database called DpTM3.0 for PTM sites.
PhosphoSitePlus [72]	http://www.phosphosite.org	Hornbeck *et al.* also created a PhosphoSitePlus database from experimentally indentified PTM in human and mouse proteins. It has included phosphorylation, acetylation, ubiquitination, and methylation sites.
UbiProt [73]	http://ubiprot.org.ru/	The lysine ubiquitin-modified site Database (UbiProt) was integrated thousands of high-confidence *in vivo* identified lysine ubiqutination on the basis of mass spectrometry. It has also included the specific information of proteins, including the nature of protein, species (mostly yeast and humans), ubiquitin-modified feature, references and related links.
SCUD [74]	http://scud.kaist.ac.kr/	A special collection of yeast lysine ubiquitin protein and its corresponding enzyme database. SCUD in version 1.0 contains 11 E2 enzyme, 42 E3 enzyme, 20 DUB enzymes as well as 940 ubiquitinated substrate.
mUbiSiDa [75]	http://222.193.31.35:8000/About_ubiquitination.php	Chen et al. created an integrated bioinformatics resource for protein animal ubiquitylation sites database termed as mUbiSiDa.
hUbiquitome [76]	http://202.38.126.151/hmdd/hubi/	hUbiquitome database released by Peking University, which was included the experimental verification of human ubiquitin-associated proteins. In this database a total of 1 E1, 12 E2, and 138 E3 substrates were existed. The database is smaller, but the confidence level is higher.
E3Net [77]	http://pnet.kaist.ac.kr/e3net/	Korea institute of science and technology bioinformatics laboratory developed E3Net. It has been updating significantly and interfaces more friendlily. It has included the total 427 species of pan pigment of modified E3 and 4,896 real proteins information.

Table 2: Some popular databases for protein PTM sites.

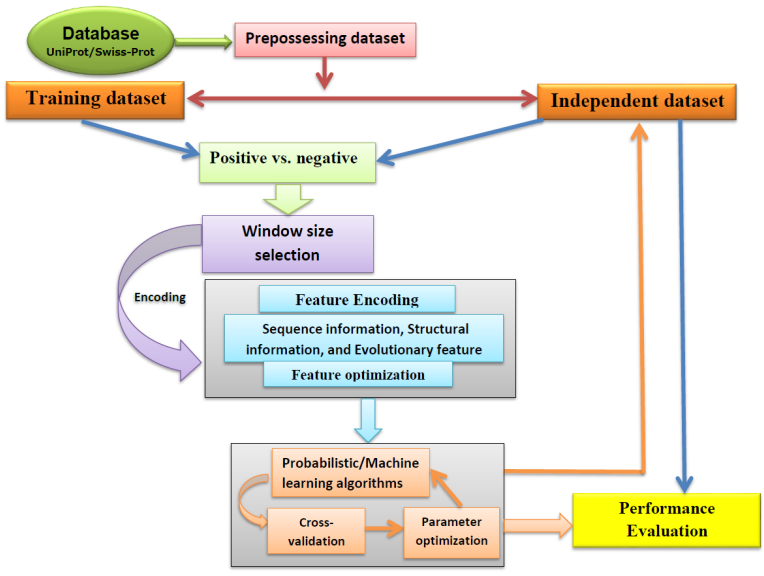

Figure 2: A flow-chart for computational prediction of protein lysine PTMs. Initially, the dataset was collected from public database. Then need to preprocess the collected datasets for making proper positive and negative samples. The encoded feature vectors were independently put into the probabilistic/machine learning models to produce independent prediction scores. Eventually, optimum performance scores were calculated by using cross-validation and parameter optimization, a confident cutoff was considered to identify the lysine PTM site.

Naïve Bayes

Naïve Bayes (NB) is a statistical probabilistic algorithm based on the statistical learning theory of Bayesian theorem [55]. The advantages of NB algorithm are very straightforward and high speed. In NB theorem, the posterior probability of a random event is the conditional probability, which is assigned after the relevant evidence is taken into account. The majority of biologists think that, for analyzing the biological data NB is an essential algorithm for analyzing biological [56]. Although, the NB models are much outlier affected and do not handle the noise datasets [57]. In lysine PTM prediction, the NB algorithm has been widely used [41].

Random forest

The random forest (RF) algorithm is a machine learning algorithm developed by Leo Breiman [58]. This model developed by using an ensemble of classification trees. RF has been widely used in lysine PTM prediction [31,32]. It was implemented as the RF package in R at https://cran.r-project.org/web/packages/randomForest/. RF is one of the most influential machine learning algorithm [59].

Support vector machine

Graft is an efficient machine learning algorithm, support vector machine (SVM) has been widely used in lysine PTM prediction [32]. In particular, the kernel radial basis function (RBF) with LIBSVM package (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) was used to train the classifiers [60]. For a given training vector $x_i \in R^n$, if the corresponding class label is $y_i \in \{-1, 1\}$, $i=1, 2, \ldots, L$, then the optimized SVM model is given by:

$$\left(\frac{1}{2}\right) w^T w + C\sum_{i=1}^{L}\zeta_i \quad (1)$$

Under the required condition,

$$y_i\left(w^T x_i + b\right) \geq 1 - \zeta_i \quad (2)$$

where , w is the weight vector to the hyperplane, C is the cost parameter, and ζ_i is the slack variable. The RBF kernel can be easily transformed to liner separation from high dimensional features. The commonly used RBF function can be defined as:

$$K\left(x_i, x_j\right) = \exp(-\gamma \mid x_i - x_j \mid^2) \quad (3)$$

where γ is the kernel parameter and $\gamma>0$, which determines how the samples are transformed to a high- dimensional space. The tuning parameters C and γ were maximized based on the training dataset by performing grid search.

Neural networks

In machine learning and cognitive science approaches, a neural network (NN) is a nonlinear statistical classifier that is able to distinguish complex relationships between two variables [61]. For example, multilayer perceptron (MLP) is one type of NN model. The MLP model has multiple layers, i.e., there are one or more nonlinear, hidden layers between the input and output layers. In the field of protein bioinformatics research, NNs have also a wide range of applications, such as protein functional sites prediction [62-64], protein secondary structure prediction [65,66] and tertiary structure prediction [67]. Common implementations of NNs software are SNNS (http://www.ra.cs.uni-tuebingen.de/SNNS/) and FANN (http://leenissen.dk/Fann/WP/) [68-77].

In this study, we have shown an overview of lysine PTM site prediction. The application and development for predicting lysine PTM sites is emerging as a promising field in protein bioinformatics research. Fundamentally, high-throughput omics techniques require rigorous computational analysis for more accurate prediction. Combining experimental and computational technologies for analyzing lysine PTMs dataset will certainly enhance our knowledge.

References

1. Witze ES, Old WM, Resing KA, Ahn NG (2007) Mapping protein post-translational modifications with mass spectrometry. Nat Methods 4: 798-806.
2. Liddy KA, White MY, Cordwell SJ (2013) Functional decorations: post-translational modifications and heart disease delineated by targeted proteomics. Genome Med 5: 20.
3. Xie L, Liu W, Li Q, Chen S, Xu M, et al. (2015) First succinyl-proteome profiling of extensively drug-resistant Mycobacterium tuberculosis revealed involvement of succinylation in cellular physiology. J Proteome Res 14: 107-119.
4. Yang M, Yang J, Zhang Y, Zhang W (2016) Influence of succinylation on physicochemical property of yak casein micelles. Food Chem 190: 836-842.
5. Rohira AD, Chen CY, Allen JR, Johnson DL (2013) Covalent small ubiquitin-like modifier (SUMO) modification of Maf1 protein controls RNA polymerase III-dependent transcription repression. J Biol Chem 288: 19288-19295.
6. Khoury GA, Baliban RC, Floudas CA (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 1.
7. Huang KY, Su MG, Kao HJ, Hsieh YC, Jhong JH, et al. (2016) dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. Nucleic Acids Res 44: D435-446.
8. Weinert BT, Scholz C, Wagner SA, Iesmantavicius V, Su D, et al. (2013) Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. Cell Rep 4: 842-851.
9. Merbl Y, Kirschner MW (2014) Post-translational modification profiling-a high-content assay for identifying protein modifications in mammalian cellular systems. Curr Protoc Protein Sci 77: 27 28 21-27 28 13.
10. Allfrey VG, Faulkner R, Mirsky AE (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA 51: 786-794.
11. Xie Z, Dai J, Dai L, Tan M, Cheng Z, et al. (2012) Lysine succinylation and lysine malonylation in histones. Mol Cell Proteomics 11: 100-107.
12. Zhang Z, Tan M, Xie Z, Dai L, Chen Y, et al. (2011) Identification of lysine succinylation as a new post-translational modification. Nat Chem Biol 7: 58-63.
13. Huang J, Berger SL (2008) The emerging field of dynamic lysine methylation of non-histone proteins. Curr Opin Genet Dev 18: 152-158.
14. Hershko A (2005) The ubiquitin system for protein degradation and some of its roles in the control of the cell-division cycle (Nobel lecture). Angew Chem Int Ed Engl 44: 5932-5943.
15. Moellering RE, Cravatt BF (2013) Functional lysine modification by an intrinsically reactive primary glycolytic metabolite. Science 341: 549-553.
16. Ansari NA, Moinuddin, Ali R (2011) Glycated lysine residues: a marker for non-enzymatic protein glycation in age-related diseases. Dis Markers 30: 317-324.
17. Cheng J, Yang Y, Fang J, Xiao J, Zhu T, et al. (2013) Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. J Biol Chem 288: 1329-1339.
18. Gareau JR, Reverter D, Lima CD (2012) Determinants of small ubiquitin-like modifier 1 (SUMO1) protein specificity, E3 ligase, and SUMO-RanGAP1 binding activities of nucleoporin RanBP2. J Biol Chem 287: 4740-4751.
19. Striebel F, Imkamp F, Sutter M, Steiner M, Mamedov A, et al. (2009) Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. Nat Struct Mol Biol 16: 647-651.
20. DeMartino GN (2009) Pupylation: Something old, something new, something borrowed, something glu. Trends Biochem Sci 34: 155-158.

21. Medzihradszky KF (2005) Peptide sequence analysis. Methods Enzymol 402: 209-244.

22. Agarwal KL, Kenner GW, Sheppard RC (1969) Feline gastrin. An example of peptide sequence analysis by mass spectrometry. J Am Chem Soc 91: 3096-3097.

23. Umlauf D, Goto Y, Feil R (2004) Site-specific analysis of histone methylation and acetylation. Methods Mol Biol 287: 99-120.

24. Welsch DJ, Nelsestuen GL (1988) Amino-terminal alanine functions in a calcium-specific process essential for membrane binding by prothrombin fragment 1. Biochemistry 27: 4939-4945.

25. Slade DJ, Subramanian V, Fuhrmann J, Thompson PR (2014) Chemical and biological methods to detect post-translational modifications of arginine. Biopolymers 101: 133-143.

26. Jaffrey SR, Erdjument-Bromage H, Ferris CD, Tempst P, Snyder SH (2001) Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. Nat Cell Biol 3: 193-197.

27. Doll S, Burlingame AL (2015) Mass spectrometry-based detection and assignment of protein posttranslational modifications. ACS Chem Biol 10: 63-71.

28. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci USA 101: 9528-9533.

29. Myers SA, Daou S, Affarel B, Burlingame A (2013) Electron transfer dissociation (ETD): the mass spectrometric breakthrough essential for O-GlcNAc protein site assignments-a study of the O-GlcNAcylated protein host cell factor C1. Proteomics 13: 982-991.

30. Basu A, Rose KL, Zhang J, Beavis RC, Ueberheide B, et al. (2009) Proteome-wide prediction of acetylation substrates. Proc Natl Acad Sci USA 106: 13785-13790.

31. Hasan MM, Yang S, Zhou Y, Mollah MN (2016) Succin-site: A computational tool for the prediction of protein succinylation sites by exploiting the amino acid patterns and properties. Mol Biosyst 12: 786-795.

32. Hasan MM, Zhou Y, Lu X, Li J, Song J, et al. (2015) Computational identification of protein pupylation sites by using profile-based composition of k-spaced amino acid pairs. PLoS One 10: e0129635.

33. Tung CW, Ho SY (2008) Computational identification of ubiquitylation sites from protein sequences. BMC Bioinformatics 9: 310.

34. Liu Z, Wang Y, Gao T, Pan Z, Cheng H, et al. (2014) CPLM: A database of protein lysine modifications. Nucleic Acids Res 42: D531-536.

35. Choudhary C, Mann M (2010) Decoding signalling networks by mass spectrometry-based proteomics. Nat Rev Mol Cell Biol 11: 427-439.

36. Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, et al. (2011) Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell 44: 325-340.

37. Gobel U, Sander C, Schneider R, Valencia A (1994) Correlated mutations and residue contacts in proteins. Proteins 18: 309-317.

38. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286: 295-299.

39. Dekker JP, Fodor A, Aldrich RW, Yellen G (2004) A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. Bioinformatics 20: 1565-1572.

40. Hasan MM, Khatun MS, Mollah MNH, Yong C, Guo D (2017) A systematic identification of species-specific protein succinylation sites using joint element features information. Int J Nanomedicine 12: 6303-6315.

41. Hasan MM, Guo D, Kurata H (2017) Computational identification of protein S-sulfenylation sites by incorporating the multiple sequence features information. Mol Biosyst 13: 2545-2550.

42. Hasan MM, Khatun MS (2018) Prediction of protein post-translational modication sites: An overview. Ann Proteom Bioinform 2: 049-057.

43. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

44. Hasan MM, Khatun MS (2017) Recent progress and challenges for protein pupylation sites prediction. EC Proteom Bioinform 2.1: 36-45.

45. Passerini A, Punta M, Ceroni A, Rost B, Frasconi P (2006) Identifying cysteines and histidines in transition-metal-binding sites using support vector machines and neural networks. Proteins 65: 305-316.

46. Youn E, Peters B, Radivojac P, Mooney SD (2007) Evaluation of features for catalytic residue prediction in novel folds. Protein Sci 16: 216-226.

47. Sharma A, Rastogi T, Bhartiya M, Shasany AK, Khanuja SP (2007) Type 2 diabetes mellitus: Phylogenetic motifs for predicting protein functional sites. J Biosci 32: 999-1004.

48. Vandermarliere E, Martens L (2013) Protein structure as a means to triage proposed PTM sites. Proteomics 13: 1028-1035.

49. Ren J, Wen L, Gao X, Jin C, Xue Y, et al. (2008) CSS-Palm 2.0: an updated software for palmitoylation sites prediction. Protein Eng Des Sel 21: 639-644.

50. Liu Z, Cao J, Ma Q, Gao X, Ren J, et al. (2011) GPS-YNO2: computational prediction of tyrosine nitration sites in proteins. Mol Biosyst 7: 1197-1204.

51. Hughey R, Krogh A (1996) Hidden Markov models for sequence analysis: An extension and analysis of the basic method. Comput Appl Biosci 12: 95-107.

52. Yoon BJ (2009) Hidden Markov models and their applications in biological sequence analysis. Curr Genomics 10: 402-415.

53. Zhao X, Ning Q, Chai H, Ma Z (2015) Accurate *in silico* identification of protein succinylation sites using an iterative semi-supervised learning technique. J Theor Biol 374: 60-65.

54. Xu HD, Shi SP, Wen PP, Qiu JD (2015) SuccFind: a novel succinylation sites online prediction tool via enhanced characteristic strategy. Bioinformatics.

55. Amirkhah R, Farazmand A, Gupta SK, Ahmadi H, Wolkenhauer O, et al. (2015) Naive Bayes classifier predicts functional microRNA target interactions in colorectal cancer. Mol Biosyst 11: 2126-2134.

56. Rani P, Pudi V (2008) RBNBC: Repeat based naive bayes classifier for biological sequences. Icdm 2008: Eighth Ieee International Conference on Data Mining, Proceedings: 989-994.

57. David J. Hand KY (2001) Idiot's bayes: Not so stupid after all? International Statistical Review / Revue Internationale de Statistique 69: 385-398.

58. Breiman L (2001) Random forests. Machine Learning 45: 5-32.

59. Fern NDMCE (2014) Do we need hundreds of classifiers to solve real world classification problems? Journal of Machine Learning 15: 3133–3181.

60. Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. ACM transactions on intelligent systems and technology p: 2.

61. Fukushima K (1975) Cognitron: A self-organizing multi-layered neural network. Biol Cybern 20: 121-136.

62. Tang YR, Chen YZ, Canchaya CA, Zhang Z (2007) GANNPhos: A new phosphorylation site predictor based on a genetic algorithm integrated neural network. Protein Eng Des Sel 20: 405-412.

63. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 4: 1633-1649.

64. Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, et al. (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. Bioinformatics 25: 2537-2543.

65. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292: 195-202.

66. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404-405.

67. Bienkowska JR, Dalgin GS, Batliwalla F, Allaire N, Roubenoff R, et al. (2009) Convergent random forest predictor: Methodology for predicting drug response from genome-scale data applied to anti-TNF response. Genomics 94: 423-432.

68. Adams J (2008) The proteome: Discovering the structure and function of proteins. Nature Education: 1(3):6.

69. Li J, Jia J, Li H, Yu J, Sun H, et al. (2014) SysPTM 2.0: An updated systematic resource for post-translational modification. Database (Oxford) 2014: bau025.

70. Tung CW (2012) PupDB: A database of pupylated proteins. BMC Bioinformatics 13: 40.

71. Lu CT, Huang KY, Su MG, Lee TY, Bretana NA, et al. (2013) DbPTM 3.0: An informative resource for investigating substrate site specificity and functional

association of protein post-translational modifications. Nucleic Acids Res 41: D295-305.

72. Shiromizu T, Adachi J, Watanabe S, Murakami T, Kuga T, et al. (2013) Identification of missing proteins in the neXtProt database and unregistered phosphopeptides in the Phosphositeplus database as part of the Chromosome-centric Human Proteome Project. J Proteome Res 12: 2414-2421.

73. Chernorudskiy AL, Garcia A, Eremin EV, Shorina AS, Kondratieva EV, et al. (2007) UbiProt: A database of ubiquitylated proteins. BMC Bioinformatics 8: 126.

74. Lee WC, Lee M, Jung JW, Kim KP, Kim D (2008) SCUD: *Saccharomyces cerevisiae* ubiquitination database. BMC Genomics 9: 440.

75. Chen T, Zhou T, He B, Yu H, Guo X, et al. (2014) mUbiSiDa: A comprehensive database for protein ubiquitination sites in mammals. PLoS One 9: e85744.

76. Du Y, Xu N, Lu M, Li T (2011) hUbiquitome: A database of experimentally verified ubiquitination cascades in humans. Database, Oxford, UK. bar055.

77. Han Y, Lee H, Park JC, Yi GS (2012) E3Net: A system for exploring E3-mediated regulatory networks of cellular functions. Mol Cell Proteomics 11: O111 014076.