Journal of Genetics and Genomics 44 (2017) 243–250

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

PLMD: An updated data resource of protein lysine modifications

Haodong Xu, Jiaqi Zhou, Shaofeng Lin, Wankun Deng, Ying Zhang, Yu Xue*

Key Laboratory of Molecular Biophysics of Ministry of Education, College of Life Science and Technology and the Collaborative Innovation Center for Brain Science, Huazhong University of Science and Technology, Wuhan 430074, China

A R T I C L E I N F O

Article history:
Received 12 December 2016
Received in revised form
26 January 2017
Accepted 30 March 2017
Available online 3 May 2017

Keywords:
Post-translational modification (PTM)
Protein lysine modification (PLM)
Ubiquitination
Acetylation
PTM crosstalk

A B S T R A C T

Post-translational modifications (PTMs) occurring at protein lysine residues, or protein lysine modifications (PLMs), play critical roles in regulating biological processes. Due to the explosive expansion of the amount of PLM substrates and the discovery of novel PLM types, here we greatly updated our previous studies, and presented a much more integrative resource of protein lysine modification database (PLMD). In PLMD, we totally collected and integrated 284,780 modification events in 53,501 proteins across 176 eukaryotes and prokaryotes for up to 20 types of PLMs, including ubiquitination, acetylation, sumoylation, methylation, succinylation, malonylation, glutarylation, glycation, formylation, hydroxylation, butyrylation, propionylation, crotonylation, pupylation, neddylation, 2-hydroxyisobutyrylation, phosphoglycerylation, carboxylation, lipoylation and biotinylation. Using the data set, a motif-based analysis was performed for each PLM type, and the results demonstrated that different PLM types preferentially recognize distinct sequence motifs for the modifications. Moreover, various PLMs synergistically orchestrate specific cellular biological processes by mutual crosstalks with each other, and we totally found 65,297 PLM events involved in 90 types of PLM co-occurrences on the same lysine residues. Finally, various options were provided for accessing the data, while original references and other annotations were also present for each PLM substrate. Taken together, we anticipated the PLMD database can serve as a useful resource for further researches of PLMs. PLMD 3.0 was implemented in PHP + MySQL and freely available at http://plmd.biocuckoo.org.

1. Introduction

If not all, most of cellular functions are controlled by cell signaling pathways involving the proteins which are frequently modified by reversible post-translational modifications (PTMs) that dynamically coordinate the signaling networks (Mann and Jensen, 2003; Mertins et al., 2013; Morris et al., 2015; Strzyz, 2016). Among the major types of amino acids that can be modified, positively charged lysine residues play important roles in regulating protein functions, while the neutralization of the charge frequently brings enormous influences on the substrate proteins (Olsen et al., 2004). Accordingly, PTMs occurring at specific lysine residues in proteins, or protein lysine modifications (PLMs), play critical roles in regulating a broad spectrum of biological processes (Shaid et al., 2013; Choudhary et al., 2014; Huang et al., 2015; Hendriks and Vertegaal, 2016). In general, specific lysine residues undergo different PLM processes 1) by addition of small molecule functional groups, which occurs in acetylation (Choudhary et al., 2014), methylation (Lanouette et al., 2014), succinylation (Zhang et al., 2011), malonylation (Xie et al., 2012), glutarylation (Tan et al., 2014), butyrylation (Chen et al., 2007), propionylation (Chen et al., 2007), crotonylation (Tan et al., 2011), and biotinylation (Sharp et al., 2006), 2) by covalent linkage of some protein modifiers, which are critical in ubiquitination (Shaid et al., 2013), sumoylation (Lamoliatte et al., 2014), and pupylation (Poulsen et al., 2010) as well as NEDD8-mediated neddylation (Rabut and Peter, 2008), or 3) by non-enzymatic attachment of sugar molecules such as glycation (Goodman, 2013).

Recently, rapid progresses of proteomic methods (e.g., high-throughput liquid chromatography-mass spectrometry (LC-MS) techniques) and the introduction of pan-antibodies specific for distinct PLMs (Hendriks et al., 2014; Elia et al., 2015; Svinkina et al., 2015) have greatly promoted the identification of well-characterized PLMs as well as new ones. For example, using a monoclonal anti-Lys-ε-Gly-Gly (anti-diGly) antibody and a

---

* Corresponding author.
E-mail address: xueyu@hust.edu.cn (Y. Xue).

http://dx.doi.org/10.1016/j.jgg.2017.03.007

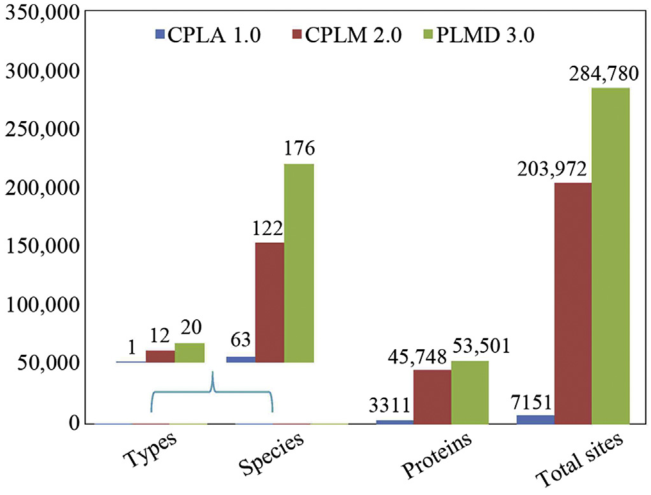


Fig. 1. Comparison of CPLA 1.0, CPLM 2.0 and PLMD 3.0. CPLA, Compendium of Protein Lysine Acetylation; CPLM, Compendium of Protein Lysine Modifications; PLMD, Protein Lysine Modification Database; Types, total lysine modification types; Species, total species lysine modification involved; Proteins, total lysine modification proteins; Total sites, total lysine modification sites.

polyclonal acetyl-lysine antibody, Elia et al. (2015) identified and quantified 33,500 ubiquitination and 16,740 acetylation sites, respectively. Also, Svinkina et al. (2015) identified over 10,000 acetylation sites in Jurkat cells with a mixture of anti-lysine acetylation antibodies. In addition, applying the combined fractional diagonal chromatography technology, Stes et al. (2014) identified more than 7500 endogenous ubiquitination sites in over 3300 proteins. Hendriks et al. (2014) also profiled a global sumoylation in human cells with high-resolution MS in a site-specific manner and totally identified over 4300 sumoylation sites in over 1600 proteins, and Weinert et al. (2013) identified almost 8000 succinylation sites from multiple species. In particular, using the state-of-the-art proteomic techniques in combination with chemical biology or biochemistry as validation tools, Dr. Yingming Zhao's group has identified a considerable number of novel PLMs such as succinylation, malonylation, propionylation crotonylation, glutarylation, and 2-hydroxyisobutyrylation, which has greatly advanced the identification and functional investigation of these PLMs (Chen et al., 2007; Xie et al., 2012; Dai et al., 2014; Tan et al., 2014). Since a flood of PLM sites have been uncovered, it has emerged to be a great challenge for the collection and integration of bulky PLM substrates and sites from different studies. Although several public databases, such as UniProt (UniProt Consortium, 2015), dbPTM (Lu et al., 2012), PhosphositePlus (Hornbeck et al., 2015), HPRD (Prasad et al., 2009) and SysPTM (Li et al., 2014), also compiled PLM information, they mainly focused on a general purpose of the collection of PTMs, and only a limited part of the identified PLM substrates and sites were included. Many newly identified PLM events still remain to be integrated.

In 2011, we developed a database of the compendium of protein lysine acetylation (CPLA) by manually collecting 7151 known acetylation sites in 3311 proteins (Liu et al., 2011). Later, we updated the CPLA 1.0 by extending acetylation to 12 types of PLMs, and renamed the database as the compendium of protein lysine modifications (CPLM 2.0) (Liu et al., 2014). In this study, we greatly improved our previous databases, and developed a much more comprehensive data resource of protein lysine modification database (PLMD). Compared to CPLA 1.0 and CPLM 2.0, PLMD 3.0 database has been greatly expanded in terms of modification types, species numbers, protein numbers and total modification sites. It contained 284,780 modification events in 53,501 proteins from 176 species for 20 types of PLMs (Fig. 1). Also, the detailed annotations of each protein entry together with the information of primary references were provided. Based on the PLMD data set, a motif-based analysis of sequence preferences was performed, and the most significantly over-represented sequence motifs were discovered around modification sites for 16 types of PLMs. Additionally, we detected 65,297 PLM events of 90 types of PLM co-occurrences on the same lysine residues, such as 24,487 acetylation-ubiquitination sites, 11,056 acetylation-succinylation sites, 5542 acetylation-malonylation sites, 4033 ubiquitination-succinylation sites, 3363 ubiquitination-sumoylation sites and 1992 succinylation-malonylation sites, and the results demonstrated that different types of PLMs prefer to crosstalk with each other. Taken together, PLMD 3.0 can service as an informative platform for

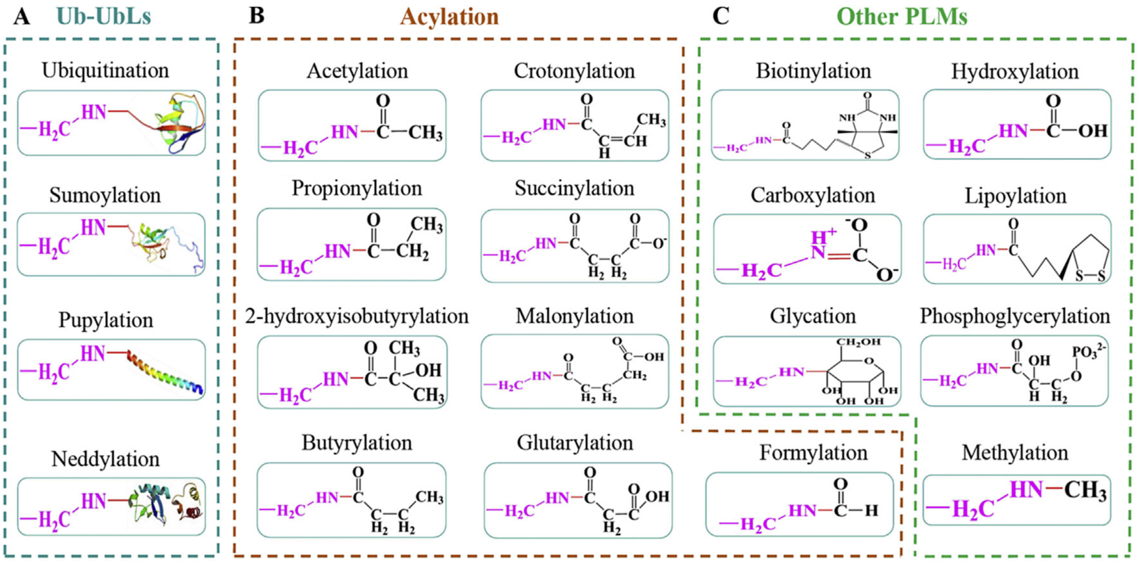


Fig. 2. Simplified diagram of 20 types of PLMs with molecular structures of ligands conjugated to lysine residues. **A**–**C**: Twenty types of PLMs classified into three categories: Ub-UbLs (**A**), acylation (**B**) and other PLMs (**C**). Ub-UbLs, ubiquitin and ubiquitin-like modifications.

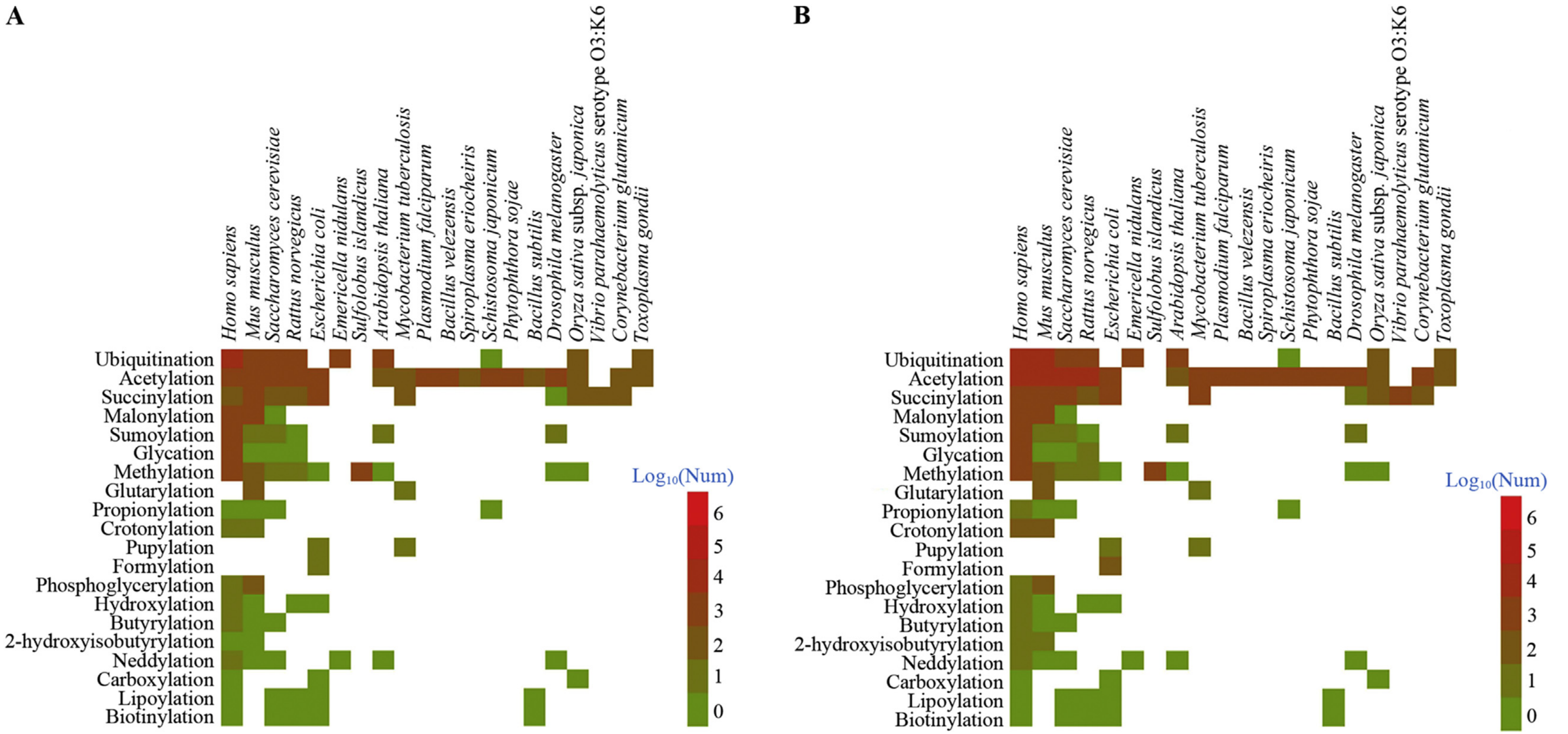
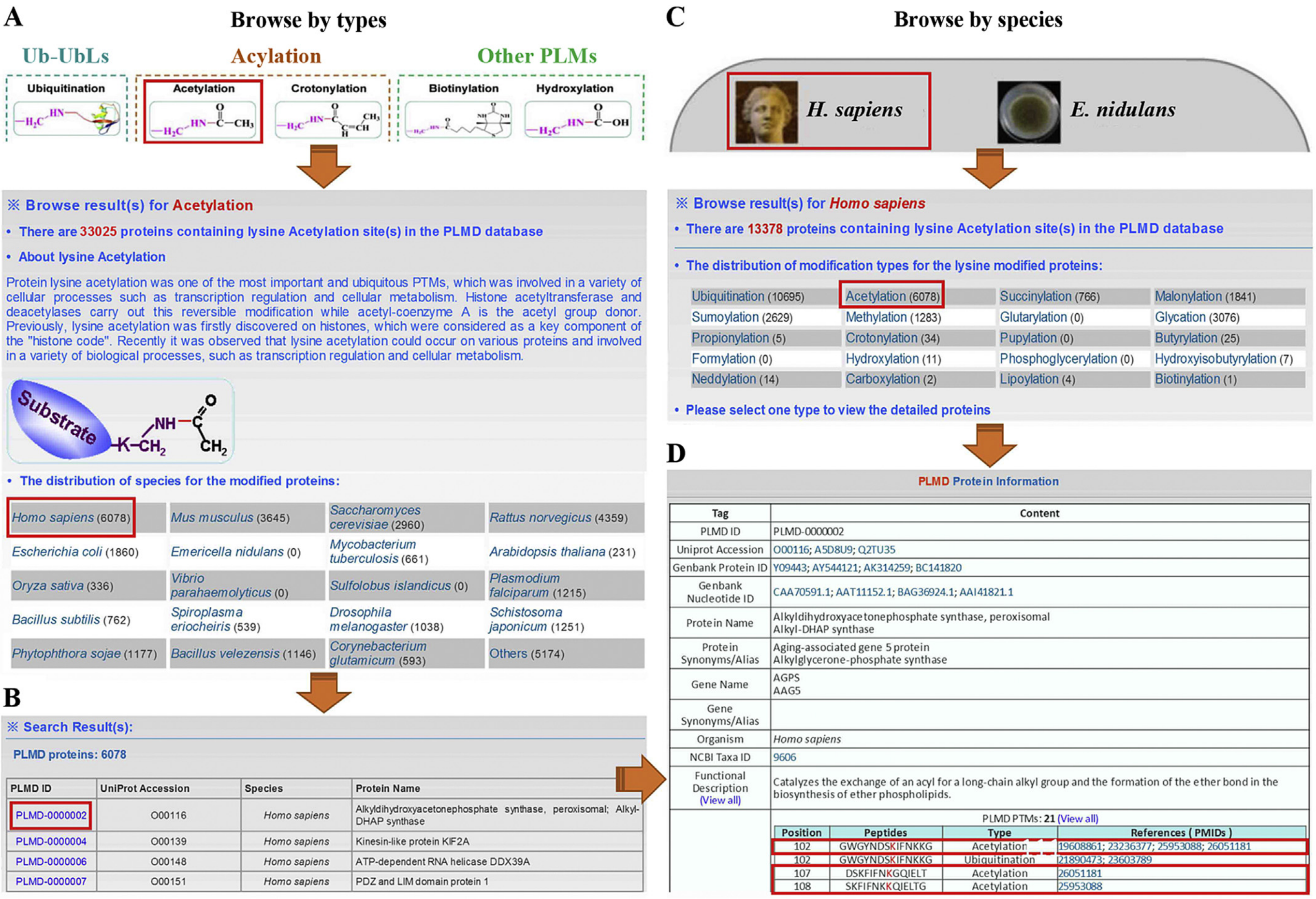


Fig. 3. The heatmaps for the distribution of protein or site number of different PLM types among different species. **A**: The heatmap for the distribution of substrate numbers. **B**: The heatmap for the distribution of modified lysine residue numbers.



※ Browse result(s) for Acetylation

- There are 33025 proteins containing lysine Acetylation site(s) in the PLMD database
- About lysine Acetylation

Protein lysine acetylation was one of the most important and ubiquitous PTMs, which was involved in a variety of cellular processes such as transcription regulation and cellular metabolism. Histone acetyltransferase and deacetylases carry out this reversible modification while acetyl-coenzyme A is the acetyl group donor. Previously, lysine acetylation was firstly discovered on histones, which were considered as a key component of the "histone code". Recently it was observed that lysine acetylation could occur on various proteins and involved in a variety of biological processes, such as transcription regulation and cellular metabolism.

- The distribution of species for the modified proteins:

Homo sapiens (6078)	*Mus musculus* (3645)	*Saccharomyces cerevisiae* (2960)	*Rattus norvegicus* (4359)
Escherichia coli (1860)	*Emericella nidulans* (0)	*Mycobacterium tuberculosis* (661)	*Arabidopsis thaliana* (231)
Oryza sativa (336)	*Vibrio parahaemolyticus* (0)	*Sulfolobus islandicus* (0)	*Plasmodium falciparum* (1215)
Bacillus subtilis (762)	*Spiroplasma eriocheiris* (539)	*Drosophila melanogaster* (1038)	*Schistosoma japonicum* (1251)
Phytophthora sojae (1177)	*Bacillus velezensis* (1146)	*Corynebacterium glutamicum* (593)	Others (5174)

※ Search Result(s):

PLMD proteins: 6078

PLMD ID	UniProt Accession	Species	Protein Name
PLMD-0000002	O00116	*Homo sapiens*	Alkyldihydroxyacetonephosphate synthase, peroxisomal; Alkyl-DHAP synthase
PLMD-0000004	O00139	*Homo sapiens*	Kinesin-like protein KIF2A
PLMD-0000006	O00148	*Homo sapiens*	ATP-dependent RNA helicase DDX39A
PLMD-0000007	O00151	*Homo sapiens*	PDZ and LIM domain protein 1

※ Browse result(s) for *Homo sapiens*

- There are 13378 proteins containing lysine Acetylation site(s) in the PLMD database
- The distribution of modification types for the lysine modified proteins:

Ubiquitination (10695)	Acetylation (6078)	Succinylation (766)	Malonylation (1841)
Sumoylation (2629)	Methylation (1283)	Glutarylation (0)	Glycation (3076)
Propionylation (5)	Crotonylation (34)	Pupylation (0)	Butyrylation (25)
Formylation (0)	Hydroxylation (11)	Phosphoglycerylation (0)	Hydroxyisobutyrylation (7)
Neddylation (14)	Carboxylation (2)	Lipoylation (4)	Biotinylation (1)

- Please select one type to view the detailed proteins

PLMD Protein Information

Tag	Content
PLMD ID	PLMD-0000002
Uniprot Accession	O00116; A5D8U9; Q2TU35
Genbank Protein ID	Y09443; AY544121; AK314259; BC141820
Genbank Nucleotide ID	CAA70591.1; AAT11152.1; BAG36924.1; AAI41821.1
Protein Name	Alkyldihydroxyacetonephosphate synthase, peroxisomal Alkyl-DHAP synthase
Protein Synonyms/Alias	Aging-associated gene 5 protein Alkylglycerone-phosphate synthase
Gene Name	AGPS AAG5
Gene Synonyms/Alias	
Organism	*Homo sapiens*
NCBI Taxa ID	9606
Functional Description (View all)	Catalyzes the exchange of an acyl for a long-chain alkyl group and the formation of the ether bond in the biosynthesis of ether phospholipids.

PLMD PTMs: **21** (View all)

Position	Peptides	Type	References (PMIDs)
102	GWGYNDSKIFNKKG	Acetylation	19608861; 23236377; 25953088; 26051181
102	GWGYNDSKIFNKKG	Ubiquitination	21890473; 23603789
107	DSKFIFNKGQIELT	Acetylation	26051181
108	SKFIFNKKQIELTG	Acetylation	25953088

Fig. 4. Two browse options of PLMD. **A** and **B**: Browse by PLM types. **C** and **D**: Browse by species.

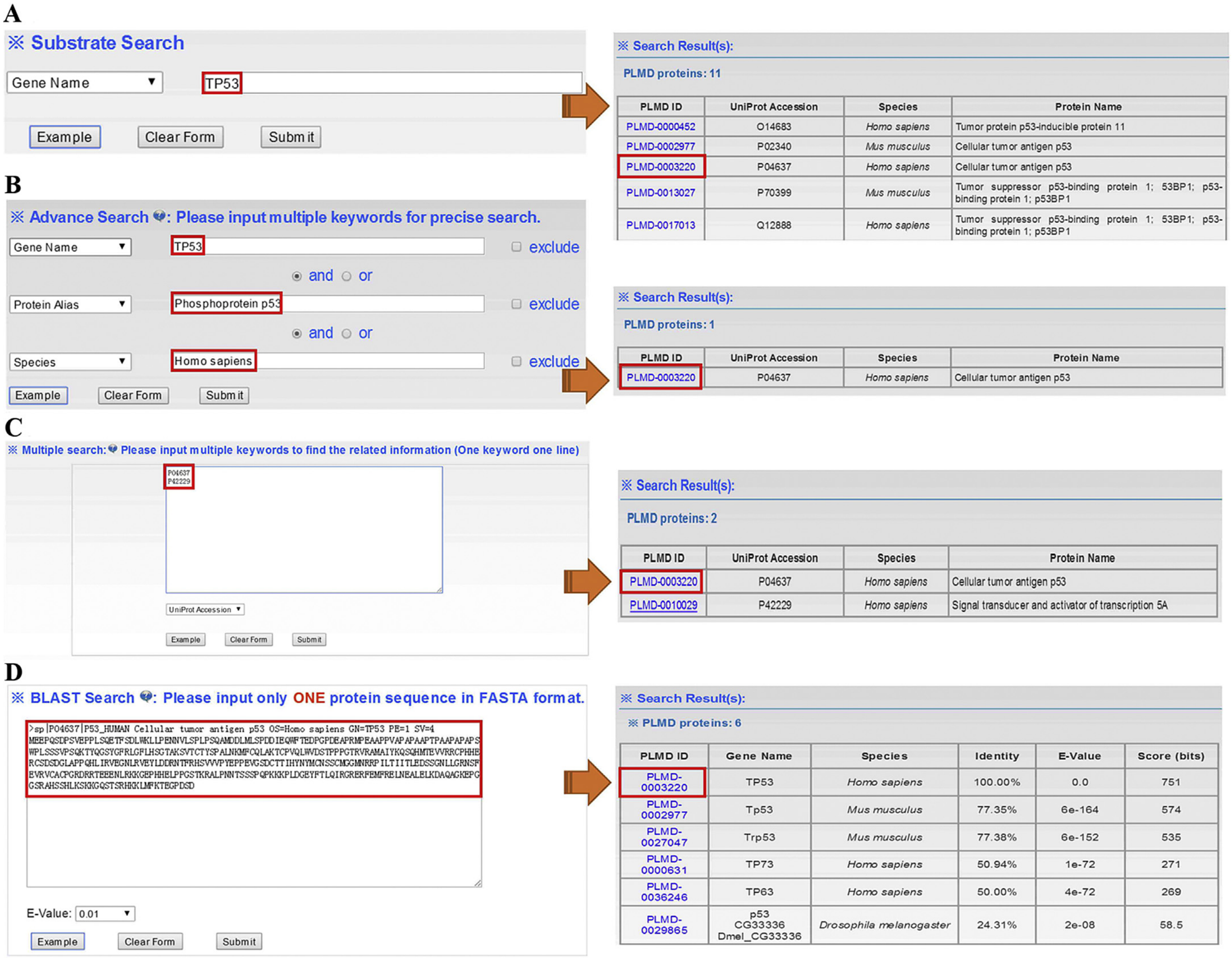
the community to access PLM information, and we anticipate that it could be a useful resource for further experimental or computational considerations.

2. Construction and content

We manually curated the searchable literature from PubMed to collect experimentally identified PLM substrates and sites by inputting keywords, including 'ubiquitination', 'acetylation', 'sumoylation', 'methylation', 'succinylation', 'malonylation', 'glutarylation', 'glycation', 'formylation', 'hydroxylation', 'butyrylation', 'propionylation', 'crotonylation', 'pupylation', 'neddylation', '2-hydroxyisobutyrylation', 'phosphoglycerylation', 'carboxylation', 'lipoylation' and 'biotinylation'. In order to provide a more comprehensive data resource for researchers, additional keywords such as 'ubiquitinated', 'acetyl', 'acetylated', 'SUMO', 'succinyl' and other related nomenclatures were adopted for collecting more data. All modified lysine residues were then mapped to the benchmark sequences retrieved from the UniProt database. In addition, more detailed annotations such as protein names, gene names and functional descriptions as well as sequence annotations of modified proteins were retrieved from UniProt and further integrated into the PLMD database for providing rich information. Also, primary references of PLM substrates and sites were offered to guarantee the reliability and quality of the database.

Totally, we found 284,780 PLM events occurring at 234,062 lysine residues of 53,501 proteins for 20 types of PLMs across 176 eukaryotes and prokaryotes (Table S1). Then we classified the PLMs into three categories: 1) four types of ubiquitin and ubiquitin-like modifications (Ub-UbLs), 2) nine types of acylations, and 3) seven types of other PLMs (Fig. 2). Among them, acylations and Ub-UbLs account for the vast majority of PLM events: the former possesses 141,276 (49.61%) acylation sites and the latter contains 130,194 (45.72%) sites. More specifically, ubiquitination (121,742 sites, 42.75%) and acetylation (111,253 sites, 39.07%), two extensively studied PLMs, still occupy a large proportion of all PLM sites with a growing number identified owing to their significant functional roles. However, it is noteworthy that the amount of newly identified lysine acylations is dramatically increasing. For example, the succinylation and malonylation sites have reached up to 18,593 (6.53%) and 9584 (3.37%), respectively. The rapid progress in the identification of these new lysine acylations is attributed to the advancement of proteomic techniques along with the emerging evidence that suggests that these new lysine acylations are important in regulating cellular metabolisms in both physiological and pathophysiological states (Hirschey and Zhao, 2015; Xu et al., 2015; Sadhukhan et al., 2016). However, only a small number of substrates could be detected regarding other new lysine acylations such as butyrylation, crotonylation, glutarylation and propionylation, which are mainly identified on histones. Moreover, for Ub-UbLs, the number of identified sumoylation sites is also increased rapidly, especially in the last two years.

Although plenty of PLMs were experimentally discovered across 176 species from our data, the amount of identified substrates is



A

Gene Name: TP53

PLMD proteins: 11

PLMD ID	UniProt Accession	Species	Protein Name
PLMD-0000452	O14683	*Homo sapiens*	Tumor protein p53-inducible protein 11
PLMD-0002977	P02340	*Mus musculus*	Cellular tumor antigen p53
PLMD-0003220	P04637	*Homo sapiens*	Cellular tumor antigen p53
PLMD-0013027	P70399	*Mus musculus*	Tumor suppressor p53-binding protein 1; 53BP1; p53-binding protein 1; p53BP1
PLMD-0017013	Q12888	*Homo sapiens*	Tumor suppressor p53-binding protein 1; 53BP1; p53-binding protein 1; p53BP1

B

Gene Name: TP53 ☐ exclude

● and ○ or

Protein Alias: Phosphoprotein p53 ☐ exclude

● and ○ or

Species: Homo sapiens ☐ exclude

PLMD proteins: 1

PLMD ID	UniProt Accession	Species	Protein Name
PLMD-0003220	P04637	*Homo sapiens*	Cellular tumor antigen p53

C

P04637
P42229

UniProt Accession

PLMD proteins: 2

PLMD ID	UniProt Accession	Species	Protein Name
PLMD-0003220	P04637	*Homo sapiens*	Cellular tumor antigen p53
PLMD-0010029	P42229	*Homo sapiens*	Signal transducer and activator of transcription 5A

D

```
>sp|P04637|P53_HUMAN Cellular tumor antigen p53 OS=Homo sapiens GN=TP53 PE=1 SV=4
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPS
WPLSSSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILTIITLEDSSGNLLGRNSF
EVRVCACPGRDRRTEEENLRKKGEPHHELPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
```

E-Value: 0.01

※ PLMD proteins: 6

PLMD ID	Gene Name	Species	Identity	E-Value	Score (bits)
PLMD-0003220	TP53	*Homo sapiens*	100.00%	0.0	751
PLMD-0002977	Tp53	*Mus musculus*	77.35%	6e-164	574
PLMD-0027047	Trp53	*Mus musculus*	77.38%	6e-152	535
PLMD-0000631	TP73	*Homo sapiens*	50.94%	1e-72	271
PLMD-0036246	TP63	*Homo sapiens*	50.00%	4e-72	269
PLMD-0029865	p53 CG33336 Dmel_CG33336	*Drosophila melanogaster*	24.31%	2e-08	58.5

Fig. 5. The search options of PLMD. **A**: 'Substrate Search' with one or multiple keywords. **B**: The 'Advanced Search' permits users to input up to three terms for query. **C**: The 'Multiple Search' permits users to enter multiple protein entries with a list of terms. **D**: The 'BLAST Search' allows users to find identical or homologous proteins with a protein sequence in FASTA format.

generally limited in most organisms. The distribution of PLM substrates and sites in 20 major species with >600 substrates were analyzed. For most PLM types, a considerable number of substrates (Fig. 3A) and sites (Fig. 3B) were identified in mammals, especially in *Homo sapiens*, *Mus musculus* and *Rattus norvegicus*. There are also more than half of the PLM types found in *Escherichia coli* and *Saccharomyces cerevisiae*. In addition, we observed that several types of Ub-UbLs were only exclusively identified in distinct species. For instance, ubiquitination, sumoylation and neddylation were only available in eukaryotes, while pupylation was only discovered in actinomyces.

3. Usage

Our database was developed in a user-friendly manner, and multiple options were provided for users to access the information of PLMs. Because of the variety of types and species of the modification data, two browse options including 'Browse by types' and 'Browse by species' were supplied in the database (Fig. 4). For convenience, PLMD allowed browsing 19 major species, while all the other species were denoted as 'Others'. Here we used lysine acetylation substrates from *H. sapiens* as an example to illustrate the usage of the browse options in PLMD. In the option of 'Browse by types', there were 20 simplified diagrams showing molecular structures of ligands conjugated to lysine residues, which were employed to represent the 20 types of PLMs and were divided into three categories including Ub-UbLs, acylation and other PLMs (Fig. 4A). First, by clicking on the 'Acetylation' button, a brief description of protein lysine acetylation and the distribution of acetylated proteins of different species were displayed. Then the acetylation substrates in *H. sapiens* could be listed through clicking on the '*Homo sapiens*' link (Fig. 4B). In the option of 'Browse by species', after clicking on the species diagram of *H. sapiens* (Fig. 4C), the distribution of lysine modified proteins of 20 types of PLMs in *H. sapiens* will be displayed, and then users can click on the link of 'Acetylation' to view the list of acetylated substrates in *H. sapiens*. The detailed information for any specified PLM protein could be accessed through the links in the list (Fig. 4D).

Besides above two options, PLMD also provided up to four implemented search options for users to query the database with

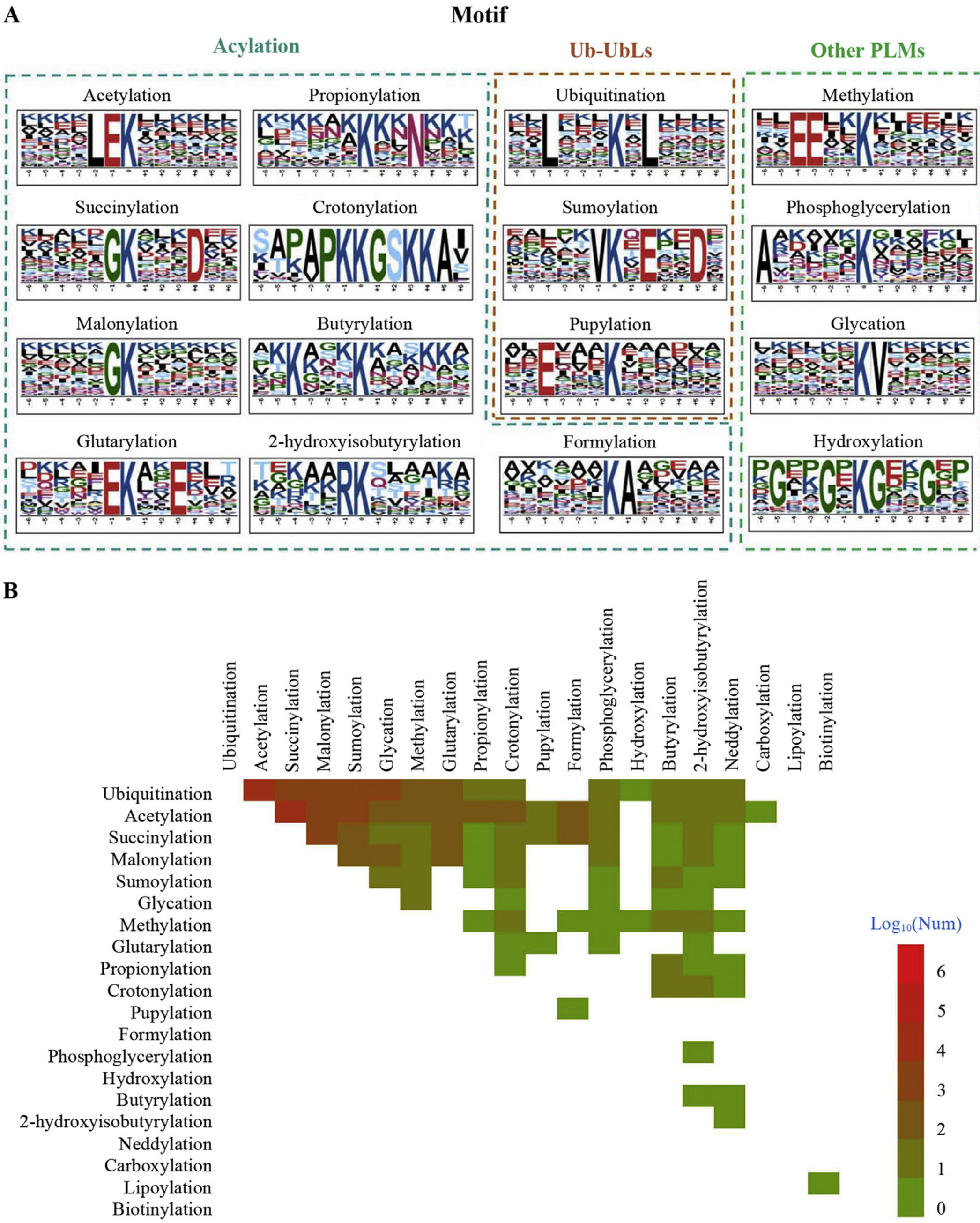


Fig. 6. The over-represented sequence motifs and potential crosstalks of different PLM types. **A**: Motif-based analyses of sequence preferences around PLM sites. Totally, 16 types of PLMs were found to process the most significantly over-represented sequence motifs. Due to the data limitation, the *P*-value of 2-hydroxyisobutyrylation was set as 0.001, and that of glutarylation and phosphoglycerylation was set as 0.0001. The other 13 types of PLMs were analyzed with a *P*-value < 0.00001. **B**: The heatmap for the distribution of co-occurrences of two different PLM types. The detailed numbers are provided in Table S2.

one or multiple keywords including ‘Substrate Search’ (Fig. 5A), ‘Advanced Search’ (Fig. 5B), ‘Multiple Search’ (Fig. 5C) and ‘BLAST Search’ (Fig. 5D). For the ‘Substrate Search’, users can input one or multiple keywords, e.g., using ‘TP53’ and selecting the ‘Gene Name’, the results will be shown in a tabular format with ‘PLMD ID’, ‘UniProt Accession’, ‘Species’ and ‘Protein Name’ (Fig. 5A). Furthermore, users can put up to three terms together by means of three operators of ‘and’, ‘or’ and ‘exclude’ to receive a more exhaustive acquirement via ‘Advanced Search’ (Fig. 5B). Also, ‘Multiple Search’ was present for retrieving multiple PLM proteins with a list of keywords (Fig. 5C). At last, with ‘BLAST Search’ option, users can find identical or homologous proteins with a protein sequence in the FASTA format by the application of NCBI BLAST packages (Fig. 5D) (Boratyn et al., 2013).

4. Discussion

The property of electron-rich and nucleophilic nature of the lysine chain makes it ideal for undergoing PTM reactions within multiple substrates (Shannon and Weerapana, 2015), which could affect the protein stability and activity and further regulate various physiological and biological functions. Moreover, an increasing number of studies also suggested that aberrances of PLMs are highly related to a considerable number of diseases (Greer and Shi, 2012; Flotho and Melchior, 2013; Morris et al., 2015; Sadoul and Khochbin, 2016). Recent advancement of proteomic techniques explosively expanded the number of PLM substrates and the discovery of novel PLMs. In view of the data accumulation, there is an urgent need to integrate these PLM sites together.

In this work, we updated the database of CPLM 2.0 into PLMD 3.0. The modification types, species numbers, protein numbers and total modification sites in PLMD 3.0 have been greatly expanded. Totally, it contained 284,780 PLM events in 53,501 substrates from 176 species for up to 20 types of PLMs, which then were classified into three categories including four types of Ub-UbLs, nine types of acylations and seven types of other PLMs.

In addition, PLM events occurring at specific positions of specific amino acids or peptides are catalyzed by a variety of corresponding enzymes (Xiong and Guan, 2012; Deng et al., 2016; Xu et al., 2016), so the flanking sequences around the modification sites of the same PLM types may possess a strong sequence conservation due to the enzyme specificity. Therefore, using the data set in PLMD combined with Motif-X (Schwartz and Gygi, 2005), we carried out a motif-based analysis of sequence preferences around modification sites for each PLM type. Sixteen types of PLMs were found to process the most significantly over-represented sequence motifs, including nine types of acylations, three types of Ub-UbLs and four types of other PLMs (Fig. 6A). Among them, the 2-hydroxyisobutyrylation was analyzed with $P < 0.001$ and glutarylation and phosphoglycerylation were analyzed with $P < 0.0001$ due to the data limitation, while other 13 types of PLMs were explored with $P < 0.00001$. For acylation motifs, the L and E residues were significantly enriched at -2 and -1 positions of the acetylation sequences, whereas N, K and R residues at $+3$, -4, and -1 positions are over-represented for lysine propionylation, butyrylation and 2-hydroxyisobutyrylation, respectively. Particularly, it was found that G residues were significantly enriched at -1 position for both succinylation and malonylation, suggesting that succinylation and malonylation incline to co-occur at the same position. For Ub-UbLs, LXXXKXL was discovered as the most predominant motif for ubiquitination. Meanwhile, KXE was found to be the over-represented motif for sumoylation, which was accordant with many reports that proteins conjugated by SUMO contain the basal KXE type motif (Geiss-Friedlander and Melchior, 2007). We also found the EEXXK is the most remarkable motif for lysine methylation. Due to the data limitation, we didn't find any over-represented motifs for the remaining four PLM types, including neddylation, carboxylation, lipoylation and biotinylation. In addition, some of the modification sites in the PLMD were reported by only one literature, while some were reported by multiple evidences. We checked modification events reported by multiple evidences and performed a motif analysis upon them (Fig. S1). From the results, for some PLMs, although the most significantly over-represented sequence motifs of identified sites reported multiple times are not fully consistent with the ones of all identified sites in the PLMD, those could be matched with the top five significantly over-represented sequence motifs of all identified PLM sites, which also indicated the reliability of our data sets.

Moreover, by competitively occurring at the same residue, multiple PLMs appear to act in combinatorial ways. From the collected data, we totally found 65,297 PLM events involved in 90 types of PLM co-occurrences. The distribution of co-occurrences of two different PLM types were pairwisely analyzed (Fig. 6B), and the detailed results were shown in Table S2. As the representatives of mostly abundant PLM crosstalks, there are 11,056 acetylation-succinylation, 5542 acetylation-malonylation and 1992 succinylation-malonylation sites in our results. Since lysine acylations depend on the similar acyl-CoA metabolic intermediates, such as acetyl-CoA (Ac-CoA), succinyl-CoA and malony-CoA (Hirschey and Zhao, 2015), it can be expected that they are more likely to occur in the same locations. Furthermore, a comprehensive analysis of the crosstalks among different PLMs in same tissues or cell lines was performed (Fig. S2). The result showed that the co-occurrences of multiple PLMs on the same lysine residue significantly occurred in majority of tissues or cell lines. Moreover, the *in situ* crosstalks between different PLMs in same tissues or cell lines are consistent with that in all identified PLM sites. Since multiple PLMs are significantly co-occurred, the preference of the crosstalk among different PLMs and the properties on the PLM crosstalk need further explorations.

In the future, PLMD will be continuously maintained to expand with the increasing availability of data in various resources as well as enhancements on the text mining algorithm which will enable PubMed to pinpoint and select more PLM substrates and sites from research articles. We anticipate that this updated database can provide a more useful resource for further computational or experimental studies.

Acknowledgments

This work was supported by grants from the National Basic Research Program (973 project; No. 2013CB933900), the Natural Science Foundation of China (Nos. 31671360 and J1103514), and the International Science & Technology Cooperation Program of China (No. 2014DFB30020).

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jgg.2017.03.007.

References

Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y., 2013. BLAST: a more efficient report with usability improvements. Nucleic Acids Res. 41, W29–W33.

Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S.C., Falck, J.R., Peng, J., Gu, W., Zhao, Y., 2007. Lysine propionylation and butyrylation are novel post-translational modifications in histones. Mol. Cell. Proteomics 6, 812–819.

Choudhary, C., Weinert, B.T., Nishida, Y., Verdin, E., Mann, M., 2014. The growing landscape of lysine acetylation links metabolism and cell signalling. Nat. Rev. Mol. Cell Biol. 15, 536–550.

Dai, L., Peng, C., Montellier, E., Lu, Z., Chen, Y., Ishii, H., Debernardi, A., Buchou, T., Rousseaux, S., Jin, F., 2014. Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. Nat. Chem. Biol. 10, 365–370.

Deng, W., Wang, Y., Ma, L., Zhang, Y., Ullah, S., Xue, Y., 2016. Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. Brief. Bioinform. bbw041.

Elia, A.E., Boardman, A.P., Wang, D.C., Huttlin, E.L., Everley, R.A., Dephoure, N., Zhou, C., Koren, I., Gygi, S.P., Elledge, S.J., 2015. Quantitative proteomic atlas of ubiquitination and acetylation in the DNA damage response. Mol. Cell 59, 867–881.

Flotho, A., Melchior, F., 2013. Sumoylation: a regulatory protein modification in health and disease. Annu. Rev. Biochem. 82, 357–385.

Geiss-Friedlander, R., Melchior, F., 2007. Concepts in sumoylation: a decade on. Nat. Rev. Mol. Cell Biol. 8, 947–956.

Goodman, C., 2013. Post-translational modifications: considering conditions. Nat. Chem. Biol. 9, 601–601.

Greer, E.L., Shi, Y., 2012. Histone methylation: a dynamic mark in health, disease and inheritance. Nat. Rev. Genet. 13, 343–357.

Hendriks, I.A., Vertegaal, A.C., 2016. A comprehensive compilation of SUMO proteomics. Nat. Rev. Mol. Cell Biol. 17, 581–595.

Hendriks, I.A., D'Souza, R.C., Yang, B., Verlaan-de Vries, M., Mann, M., Vertegaal, A.C., 2014. Uncovering global SUMOylation signaling networks in a site-specific manner. Nat. Struct. Mol. Biol. 21, 927–936.

Hirschey, M.D., Zhao, Y., 2015. Metabolic regulation by lysine malonylation, succinylation, and glutarylation. Mol. Cell. Proteomics 14, 2308–2315.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., Skrzypek, E., 2015. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 43, D512–D520.

Huang, H., Lin, S., Garcia, B.A., Zhao, Y., 2015. Quantitative proteomic analysis of histone modifications. Chem. Rev. 115, 2376–2418.

Lamoliatte, F., Caron, D., Durette, C., Mahrouche, L., Maroui, M.A., Caron-Lizotte, O., Bonneil, E., Chelbi-Alix, M.K., Thibault, P., 2014. Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. Nat. Commun. 5, 5409.

Lanouette, S., Mongeon, V., Figeys, D., Couture, J.F., 2014. The functional diversity of protein lysine methylation. Mol. Syst. Biol. 10, 724.

Li, J., Jia, J., Li, H., Yu, J., Sun, H., He, Y., Lv, D., Yang, X., Glocker, M.O., Ma, L., 2014. SysPTM 2.0: an updated systematic resource for post-translational modification. Database 2014, bau025.

Liu, Z., Cao, J., Gao, X., Zhou, Y., Wen, L., Yang, X., Yao, X., Ren, J., Xue, Y., 2011. CPLA 1.0: an integrated database of protein lysine acetylation. Nucleic Acids Res. 39 (Suppl. 1), D1029–D1034.

Liu, Z., Wang, Y., Gao, T., Pan, Z., Cheng, H., Yang, Q., Cheng, Z., Guo, A., Ren, J., Xue, Y., 2014. CPLM: a database of protein lysine modifications. Nucleic Acids Res. 42, D531–D536.

Lu, C.-T., Huang, K.-Y., Su, M.-G., Lee, T.-Y., Bretaña, N.A., Chang, W.-C., Chen, Y.-J., Chen, Y.-J., Huang, H.-D., 2012. DbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. Nucleic Acids Res. 41, D295–D305.

Mann, M., Jensen, O.N., 2003. Proteomic analysis of post-translational modifications. Nat. Biotechnol. 21, 255–261.

Mertins, P., Qiao, J.W., Patel, J., Udeshi, N.D., Clauser, K.R., Mani, D., Burgess, M.W., Gillette, M.A., Jaffe, J.D., Carr, S.A., 2013. Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat. Methods 10, 634–637.

Morris, M., Knudsen, G.M., Maeda, S., Trinidad, J.C., Ioanoviciu, A., Burlingame, A.L., Mucke, L., 2015. Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. Nat. Neurosci. 18, 1183–1189.

Olsen, J.V., Ong, S.-E., Mann, M., 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol. Cell. Proteomics 3, 608–614.

Poulsen, C., Akhter, Y., Jeon, A.H.W., Schmitt-Ulms, G., Meyer, H.E., Stefanski, A., Stühler, K., Wilmanns, M., Song, Y.H., 2010. Proteome–wide identification of mycobacterial pupylation targets. Mol. Syst. Biol. 6, 386.

Prasad, T.K., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., 2009. Human protein reference database–2009 update. Nucleic Acids Res. 37 (Suppl. 1), D767–D772.

Rabut, G., Peter, M., 2008. Function and regulation of protein neddylation. 'Protein modifications: beyond the usual suspects' review series. EMBO. Rep. 9, 969–976.

Sadhukhan, S., Liu, X., Ryu, D., Nelson, O.D., Stupinski, J.A., Li, Z., Chen, W., Zhang, S., Weiss, R.S., Locasale, J.W., 2016. Metabolomics-assisted proteomics identifies succinylation and SIRT5 as important regulators of cardiac function. Proc. Natl. Acad. Sci. U. S. A. 113, 4320–4325.

Sadoul, K., Khochbin, S., 2016. The growing landscape of tubulin acetylation: lysine 40 and many more. Biochem. J. 473, 1859–1868.

Schwartz, D., Gygi, S.P., 2005. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat. Biotechnol. 23, 1391–1398.

Shaid, S., Brandts, C., Serve, H., Dikic, I., 2013. Ubiquitination and selective autophagy. Cell Death Differ. 20, 21–30.

Shannon, D.A., Weerapana, E., 2015. Covalent protein modification: the current landscape of residue-specific electrophiles. Curr. Opin. Chem. Biol. 24, 18–26.

Sharp, J.S., Nelson, S., Brown, D., Tomer, K.B., 2006. Structural characterization of the E2 glycoprotein from Sindbis by lysine biotinylation and LC-MS/MS. Virology 348, 216–223.

Stes, E., Laga, M., Walton, A., Samyn, N., Timmerman, E., De Smet, I., Goormachtig, S., Gevaert, K., 2014. A COFRADIC protocol to study protein ubiquitination. J. Proteome Res. 13, 3107–3113.

Strzyz, P., 2016. Post-translational modifications: extension of the tubulin code. Nat. Rev. Mol. Cell Biol. 17, 609.

Svinkina, T., Gu, H., Silva, J.C., Mertins, P., Qiao, J., Fereshetian, S., Jaffe, J.D., Kuhn, E., Udeshi, N.D., Carr, S.A., 2015. Deep, quantitative coverage of the lysine acetylome using novel anti-acetyl-lysine antibodies and an optimized proteomic workflow. Mol. Cell. Proteomics 14, 2429–2440.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell 146, 1016–1028.

Tan, M., Peng, C., Anderson, K.A., Chhoy, P., Xie, Z., Dai, L., Park, J., Chen, Y., Huang, H., Zhang, Y., 2014. Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. Cell. Metab. 19, 605–617.

UniProt Consortium, 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212.

Weinert, B.T., Schölz, C., Wagner, S.A., Iesmantavicius, V., Su, D., Daniel, J.A., Choudhary, C., 2013. Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. Cell Rep. 4, 842–851.

Xie, Z., Dai, J., Dai, L., Tan, M., Cheng, Z., Wu, Y., Boeke, J.D., Zhao, Y., 2012. Lysine succinylation and lysine malonylation in histones. Mol. Cell. Proteomics 11, 100–107.

Xiong, Y., Guan, K.-L., 2012. Mechanistic insights into the regulation of metabolic enzymes by acetylation. J. Cell Biol. 198, 155–164.

Xu, H.-D., Shi, S.-P., Wen, P.-P., Qiu, J.-D., 2015. SuccFind: a novel succinylation sites online prediction tool via enhanced characteristic strategy. Bioinformatics 31, 3748–3750.

Xu, Y., Zhang, S., Lin, S., Guo, Y., Deng, W., Zhang, Y., Xue, Y., 2016. WERAM: a database of writers, erasers and readers of histone acetylation and methylation in eukaryotes. Nucleic Acids Res. 45, D264–D270.

Zhang, Z., Tan, M., Xie, Z., Dai, L., Chen, Y., Zhao, Y., 2011. Identification of lysine succinylation as a new post-translational modification. Nat. Chem. Biol. 7, 58–63.