Bioinformatics, 39(1), 2023, btac823
https://doi.org/10.1093/bioinformatics/btac823
Advance Access Publication Date: 22 December 2022
Original Paper

Databases and ontologies

PTMint database of experimentally verified PTM regulation on protein–protein interaction

Xiaokun Hong[1], Ningshan Li[2], Jiyang Lv[1], Yan Zhang [1], Jing Li [1]*, Jian Zhang [3]* and Hai-Feng Chen [1]*

[1]State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China, [2]Department of Bioinformatics and Biostatistics, SJTU-Yale Joint Center for Biostatistics and Data Science, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China and [3]Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Department of Pathophysiology, Shanghai Jiao-Tong University School of Medicine (SJTU-SM), Shanghai 200025, China

*To whom correspondence should be addressed.
Associate Editor: Peter Robinson

Received on May 25, 2022; revised on December 18, 2022; editorial decision on December 19, 2022; accepted on December 21, 2022

Abstract

Motivation: Post-translational modification (PTM) is an important biochemical process. which includes six most well-studied types: phosphorylation, acetylation, methylation, sumoylation, ubiquitylation and glycosylation. PTM is involved in various cell signaling pathways and biological processes. Abnormal PTM status is closely associated with severe diseases (such as cancer and neurologic diseases) by regulating protein functions, such as protein–protein interactions (PPIs). A set of databases was constructed separately for PTM sites and PPI; however, the resource of regulation for PTM on PPI is still unsolved.

Results: Here, we firstly constructed a public accessible database of PTMint (PTMs that are associated with PPIs) (https://ptmint.sjtu.edu.cn/) that contains manually curated complete experimental evidence of the PTM regulation on PPIs in multiple organisms, including *Homo sapiens*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Currently, the first version of PTMint encompassed 2477 non-redundant PTM sites in 1169 proteins affecting 2371 protein–protein pairs involving 357 diseases. Various annotations were systematically integrated, such as protein sequence, structure properties and protein complex analysis. PTMint database can help to insight into disease mechanism, disease diagnosis and drug discovery associated with PTM and PPI.

Availability and implementation: PTMint is freely available at: https://ptmint.sjtu.edu.cn/.

Contact: haifengchen@sjtu.edu.cn or jian.zhang@sjtu.edu.cn or jing.li@sjtu.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Post-translational modification (PTM) is an important biochemical process among several organisms. There are over 400 known PTM types, of which six types are well studied, including phosphorylation (Phos), acetylation (Ac), methylation (Me), sumoylation (Sumo), ubiquitylation (Ub) and glycosylation (Glyco). Most biological process and signaling pathway occur by interaction of two or more proteins (De Las Rivas and Fontanillo, 2010), which are regulated by PTM (Seet *et al.*, 2006). Abnormal PTM status on proteins could lead to severe diseases [such as Alzheimer's disease (Lau *et al.*, 2008), cancer (Gu *et al.*, 2013) and cardiovascular disease (Coxon *et al.*, 2012)] by regulating protein functions, such as protein–protein interactions (PPIs).

Thousands of PTM sites of several organisms have been identified owing to the advancements in mass spectrometry (MS) (Choudhary and Mann, 2010; Gao and Chen, 2021; Swietlik *et al.*, 2020). The abundant PTM data are stored in publicly available resources, such as PhosphoSitePlus (Hornbeck *et al.*, 2015), qPhos (Yu *et al.*, 2019), Phospho.ELM (Dinkel *et al.*, 2011), dbPTM (Li *et al.*, 2022), PLMD (Xu *et al.*, 2017) and the universal database, Uniprot (UniProt, 2012). However, the role of massive PTM sites in protein, especially on the PPIs is still remained to clarify.

Uniprot database offers the 'PTM/Processing' section to record the PTM sites and/or processing events. However, its relatively hard for inexperienced users to search and browse, and it also lacks the information of PTM onto the 3D structures. Another database, PhosphoSitePlus (Hornbeck *et al.*, 2015) provides PTM effects on PPIs based on literature mining with Linguamatics software (Bandy *et al.*, 2009), which inevitably includes much false positive results. To date, several tools and resources are emerged to predict the PTM functions. PTMfunc predicts the PTM effect based on the conservation in the domain (Beltrao *et al.*, 2012). PTMcode provided known and predicted functional associations between PTMs based on co-evolution theory (Minguez *et al.*, 2015). Another two different methods, one is Mechismo web server based on interface pair potentials (Betts *et al.*, 2015) and the other is FoldX software based on empirical forcefield (Schymkowitz *et al.*, 2005), which rely on interfacial PTM sites in practice, solely.

With all the above in consideration, we presented the PTMint database, a comprehensive experimentally verified PTM effects on PPIs, such as PTM types and sites, interaction proteins, detection methods, associated diseases and co-localization. Moreover, in order to facilitate the investigation of PTM roles, we combined the experimental evidence with sequence and structure annotation. This database will be helpful for researchers to explore the relationship among PTM, PPIs and diseases in sequence and structure aspects.

2 Materials and methods

2.1 Data sources

The workflow of the PTMint database construction was shown in Figure 1, including data collection and annotation. We defined the regulatory roles of PTM on PPIs (Betts *et al.*, 2015, 2017; Li *et al.*, 2012; Lin *et al.*, 2021; Seet *et al.*, 2006; Spektor and Rice 2009; Wang *et al.*, 2022): (i) Enhance: Increase affinity and (ii) Inhibit: Decrease affinity. We extracted the functional PTM sites and associated literature from Uniprot (UniProt, 2012), PTMD (Xu *et al.*, 2018), PTMfunc (Beltrao *et al.*, 2012) and PhosphoSitePlus (Hornbeck *et al.*, 2015) databases. We also downloaded the relevant literature using PubMed database by searching the following keywords and the combinations: *Homo sapiens*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, protein, bind, associate, enable, interact, interaction, inhibit, disable, prevent, dissociate, site, PTM, Phos, Ac, Me, Sumo, Ub and Glyco. Then, we checked the full text of the above nearly 3600 papers carefully to obtain complete experimental evidence, which included regulatory PTM sites and types, interacting proteins, detection methods, associated diseases and co-localization. Briefly, we examined the relationship between PTM and disease based on cellular or animal disease models in each literature. Then, based on the detection method of PPIs, we examined the protein interactions affected by PTM in the full text. By the way, we established the relationship among PTM, protein interactions and diseases.

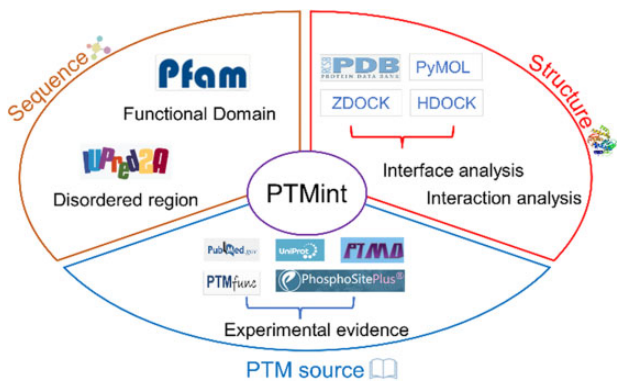

Fig. 1. The overall design and construction of PTMint database

2.2 Protein sequence analysis

We searched the Uniport database (UniProt, 2012) to obtain all canonical protein sequences. And the sequence window (upstream and downstream five residues around the PTM sites) was also extracted. The disorder propensity scores were calculated by IUPred2A (Meszaros *et al.*, 2018). Protein sequences were annotated using Pfam database to obtain functional domain information (Mistry *et al.*, 2021).

2.3 Protein structure

For individual proteins, we downloaded the full-length protein structures in bulk from AlphaFold Protein Structure Database (AlphaFold DB) (Varadi *et al.*, 2022). If the protein length was longer than 2700 amino acids, we used AlphaFold (version 2.1.2) (Jumper *et al.*, 2021) to predict the domain structures of long-length proteins, respectively. For protein complexes, all the paired protein sequences were mapped to the PDB database (Berman *et al.*, 2000) using blastp against pdbaa with e-value of 10^{-4}. The PDB entries were selected according to the following criteria: (i) The PTM sites existed on the structures. (ii) The protein name of matched sequence was same as query protein. (iii) The matched complexes with two chains were preferred. Due to the limited crystal structures, a large scale of protein–protein dockings was performed by molecular docking softwares. ZDOCK is a fast Fourier transform-based docking procedure for rigid proteins that searches for all possible binding modes in the translational and rotational space between two proteins and evaluates each pose using an energy-based scoring function (Pierce *et al.*, 2014). HDOCK is the hybrid docking algorithm of template modeling and free docking based on the docking program and allows the users to provide possible protein–protein binding sites to perform rapid protein–protein docking (Yan *et al.*, 2020). PyMOL performs molecular docking based on template alignment, which maximizes the retention of the docking mode of the original template structure (DeLano, 2010). For homology modeling, we used PyMOL (version 2.4.1) to obtain complex structure based on PDB template structure. For molecular docking, we used ZDOCK (version 3.0.2) and HDOCK (version 1.1) by using XL-MS data (cross-linking) through exhaustive curation of published literatures and predicted domain–domain interactions provided by INstruct database (Meyer *et al.*, 2013) as the docking constraints. For the protein–protein docking results provided by all softwares, we uniformly used the structure with the highest score in the software as the final docking complex structure.

2.4 Interaction analysis

Interaction assignment was handled with in-house software (Chen and Luo, 2007; Wang *et al.*, 2014). The hydrophobic interaction (HP) is defined when the mass centers of side chain for hydrophobic residues are closed within 6.5 Å. The charge–charge interaction within 11 Å plays a key role in protein/ligand-binding free energy (Qin *et al.*, 2010). Thus, the distance between the mass centers of charge residues is less than 11 Å, which was considered as electrostatic interaction (ELE). A hydrogen bond (HB) within the complex is defined when the distance of two polar heavy atoms is less than 3.5 Å and the bond angle is larger than 120°. We utilized the InterfaceResidues.py Python script created by Vertrees J (https://pymolwiki.org/) for complex interfaceResidues analysis. Briefly, this Python script splits the complex into two pieces for two interacting chain and then calculates the difference between the complex-based accessible surface areas and the chain-only-based accessible surface areas. If the value is greater than cutoff (the default is 1.0 Å2), the residues is marked as interfacial residue. The same process was handled for PTM sites to label interfacial PTM sites.

2.5 Secondary structure analysis

To obtain the property of the secondary structure of complex structures and PTM sites, the secondary structure content was calculated by Dictionary of Protein Secondary Structure algorithm (Kabsch and Sander, 1983) according to the residue-specific HBs in eight

secondary structures (π-helix, 3, (10)-helix, α-helix, β-bridge, β-sheet, turn, bend and coil). For simplification, we have classified into four types: (i) Helix: π-helix, 3, (10)-helix and α-helix; (ii) Sheet: β-bridge and β-sheet; (iii) Turn: turn; (iv) Loop: bend and coil.

2.6 Score calculation

Among the 20 basic amino acids in proteins, some amino acids are frequently PTM-modified, such as Serine (S), Threonine (T), Tyrosine (Y), Lysine (K) and Arginine (R). Serine and Threonine can be modified by Phos and Glyco. And Lysine can be modified by multiple modifications, such as Ac, Me, Sumo and Ub. To assess the important regulatory role of PTM in PPI networks *in vivo*, we introduced the importance score of PTM sites.

This importance score takes into account of the number of PTM types and interacting proteins and calculated with Equation (1).

$$\text{Score} = 2 \bullet \sum(\text{PTM} \bullet \text{N}). \quad (1)$$

In which N is the number of protein which specific PTM site regulates.

And the Score can be normalized with Equation (2).

$$\begin{aligned} &\text{Normalized\_score} = 1 - 1/\text{Score} \\ &\text{Normalized\_score} \in [0.5, 1) \end{aligned} \quad (2)$$

The normalized_score reflects the relative importance of specific regulatory PTM sites, which will increase as the number of PTM types or interaction proteins increase.

2.7 Database and web interface implementation

The web interfaces were implemented in Hyper Text Markup Language (HTML), JavaScript (JS) and Cascading Style Sheets. And the web frame was supported by Bootstrap v4 framework. Furthermore, 3Dmol.js plugin was employed to visualize protein 3D structures (Rego and Koes, 2015). And the PPI network was analyzed and visualized by ECharts plugin (Li *et al.*, 2018). Besides, all figures and tables in the website were performed in Python.

3 Results

3.1 Database and content

The current version of PTMint contains 2477 non-redundant PTM sites in 1169 proteins affecting 2371 protein–protein pairs involving 357 diseases. Uniport database provides the 425 records of PTM effect on PPI and 322 functional PTM sites. PTMD database provides the 45 records of PTM effect on PPI and 34 functional PTM sites. And PhosphositePlus guides us to search some reference literature based on text mining. In the two regulatory roles (Enhance and Inhibit), 'Enhance' has a bigger proportion, suggesting that PTM might tend to increase PPIs (Table 1). In our results, the top 1 of six main PTM types with the largest number is Phos (87%) (Fig. 2A). The main regulatory PTM sites are Serine (S), Threonine (T) and Tyrosine (Y), which have a cumulative proportion of 87.20% (Fig. 2B).

As shown in Supplementary Table S1, of all the experimentally verified regulatory PTM sites, 15.82% localizes in the protein–protein interfaces, suggesting a large proportion of non-interfacial sites

Table 1. Summary of PTM effects on PPIs

PTM types	PTM sites	Enhance	Inhibit
Phosphorylation	S, T, Y	2157	968
Acetylation	K	170	86
Methylation	K, R	81	19
Sumoylation	K	53	26
Ubiquitylation	K	60	11
Glycosylation	S, T	5	0

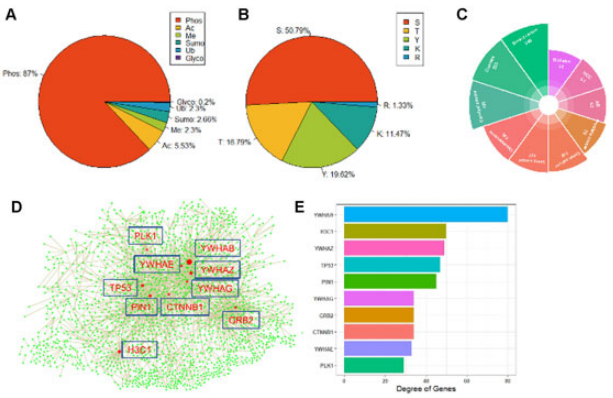

Fig. 2. The data in the PTMint database. Statistical analysis results in terms of PTM types (**A**) and PTM sites (**B**). (**C**) Top 10 diseases affected by PTM. (**D**) PTM-associated PPI network. Top 10 genes with highest degrees were highlighted in PPI (D) and visualized in bar plot (**E**)

can also regulate molecular interactions; 36.61% can be found in the functional domains, implicating the important biological role of PTM in modulating protein function. In the view of secondary structure, PTM sites tend to localize in the loop region rather than structured regions (Helix, Sheet and Turn) (Table S2). Based on results of score calculation, the score of most PTM sites (fraction: 74.54%) is 0.5, which indicates most site modified by one type of PTM, can modulate one PPI in the collected data. And the K10 site of H3C1 possesses the highest score of 0.97. Ac of K10 can regulate five different PPI, including BAZ1B, BRD7, CHD4, CRH and TRIM33. In addition, the Me of K10 can regulate 14 different PPIs, including AGO3, CBX1, CBX3, CDYL, CDYL2, CHAMP1, CHD4, DCAF8, HSFY1, KAT5, MAD2L2, POGZ, UHRF1 and ZNF470.

To further understand the intrinsic characteristics of PTM-modified proteins and interactor proteins, these all proteins were grouped into multiple classifications according to the biological function (Supplementary Table S3), mainly enzymes and transporters, suggesting these proteins with PTM participate in extensive biological processes and signaling pathways. In the database, there are total 2960 complex structures which 360 structures (fraction: 12.16%) come from PDB experimental structures, 203 structures (fraction: 6.86%) from homology modeling (PyMOL) and 2397 (fraction: 80.98%) structures from molecular docking (ZDOCK and HDOCK). According to the prior docking knowledge (XL-MS and domain–domain interaction), each complex was assigned a confidence value (High, Medium or Low), 32.26% of which were 'High' or 'Medium'.

The top 10 diseases affected by PTM were counted and shown in Figure 2C, which mainly included following three types: (i) cancers: breast cancer (number: 340), cancers (number: 323), cervical cancer (number: 302), osteosarcoma (number: 136), lung cancer (number: 127), colon cancer (number: 119), prostate cancer (number: 76) and hepatocellular carcinoma (HCC) (number: 61); (ii) Alzheimer's disease (AD) (number: 63); and (iii) diabetes (number: 52). We also constructed a PTM-associated PPI network (Fig. 2D). Nodes with high degree in PPI network represent the core nodes. The results showed that the top 10 genes with the highest degrees were: YWHAB (degree: 80), H3C1 (degree: 50), YWHAZ(degree: 49), TP53 (degree: 47), PIN1 (degree: 45), GRB2 (degree: 34), YWHAG (degree: 34), CTNNB1 (degree: 34), YWHAE (degree: 33) and PLK1 (degree: 29) (Fig. 2D and E), suggesting that these genes were disease-susceptible and potential drug targets.

3.2 Web search function

Quick search and advanced search were implemented on the homepage and 'Search' page, respectively. On the homepage, the user can directly search the database by inputting keyword (such as Gene, Uniprot, PTM, Effect and Organism) (Fig. 3B). Single or multiple filter conditions, such as Gene/Uniprot, Organism and PTM types can

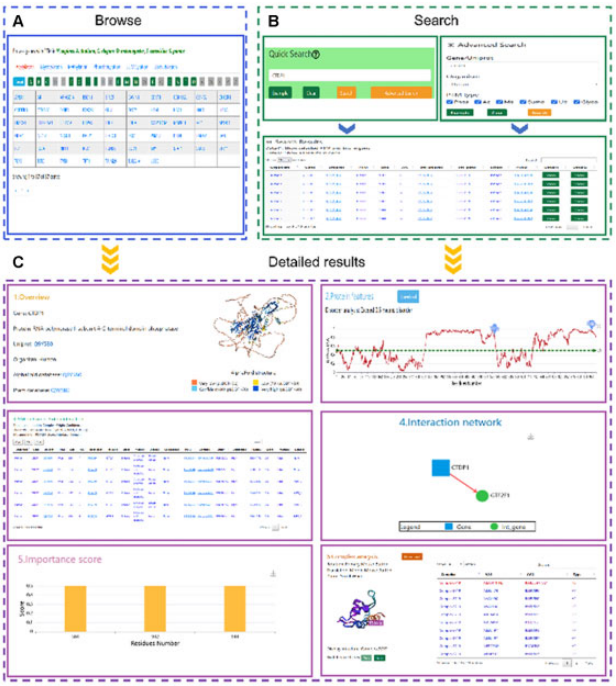

Fig. 3. The detailed information in PTMint database. (**A**) Browse function. (**B**) Search_function. (**C**) The returned search results using the CTDP1 as an example. The entry page consists of six major sections: (1) Protein overview. (2) Protein features. (3) PTM on PPIs. (4) Interaction network. Blue square: gene, Green circle: Int_gene, Red arrow: from gene to int_gene interaction. (5) Importance score of PTM sites and (6) complex analysis (A color version of this figure appears in the online version of this article)

be specified on the 'Search' page (Fig. 3B). Taking 'CTDP1' gene as an example, the searched results will be shown in a tabular format, including Organism, Gene, Uniprot, PTM, Site, AA, Int_uniprot, Int_gene, Effect, PMID and Detail (Fig. 3B). Hyperlinks for Uniprot and PubMed are provided. And the two types of detailed results can be shown by clicking the 'Show' buttons, respectively (Fig. 3B).

The typical result page consists of six main sections (Fig. 3C), including protein overview (such as organism, protein name, protein structure and protein domain information), protein features (disorder analysis), PTM on PPIs, interaction network, importance score of PTM sites and complex analysis. In 'PTM on PPIs' section, complete experimental evidence and structure, site annotation were integrated (such as whether PTM site localized on the interface or protein domain, complex origin), which can be saved or searched, easily. In 'Complex analysis' section, complex structures came from PDB database (Berman *et al.*, 2000) and local molecular dock. And PTM site and type were mapped onto the structure. And users can easily manipulate and switch structures. Furthermore, protein interactions (HB, HP and ELE) calculated by in-house software (Chen and Luo, 2007; Wang *et al.*, 2014) are shown in a tabular format. InterfaceResidues was also calculated to annotate the spatial location of PTM sites. 'Download' function was provided for users to download protein features and all complex information composed of complex structures, interfaceResidues and interaction. We also provided several external links, such as Uniport database (UniProt, 2012), AlphaFold database (Varadi *et al.*, 2022), Pfam database (Mistry *et al.*, 2021) and PubMed database by clicking underlined links.

3.3 Web browse function

The PTM types and genes were sorted and organized in alphabetical order, which allow the user to quickly obtain interested results (Fig. 3A).

3.4 Web download and help function

All data in the PTMint database can be downloaded in the 'Download' page, including PTM experimental evidence and protein structure information. And detailed instructions were available in the 'Help' page.

4 Discussion

To our knowledge, PTMint database is the first comprehensive database of experimental evidence of the PTM effects on PPIs, which not only includes complete experimental records, such as PTM types and sites, interacting proteins, detection methods, associated diseases and co-localization, but also integrates the according sequence and structure annotation (such as molecular dock and interaction analysis), systematically.

PTM level in proteins is controlled precisely based on a temporal and spatial context (Cohen, 2002; Hunter, 1995). And the same site might have different PTM types in various physiological states, such as cancer and hypoxia (Xu *et al.*, 2022). For example, Lysine (K) can selectively be acetylated, methylated or ubiquitylated. Serine (S) can be phosphorylated or glycosylated. In addition, we found the specific PTM type in a site of the protein, can regulate several proteins in our collected data. For example, phos-S289 in MDM4 can simultaneously induce MDM4-MDM2 and MDM4-p53 interactions (Wu *et al.*, 2012). Beta2 integrin Phos on Thr758 acts as a molecular switch to inhibit filamin binding and enhance the 14-3-3 protein binding to the integrin cytoplasmic domain (Takala *et al.*, 2008). Furthermore, 14-3-3 proteins, which contain a phosphoprotein-binding domains (PPBDs), can bind phos-T32 FOXO3 (Singh *et al.*, 2010), phos-S253 FOXO3 (Singh *et al.*, 2010), phos-T642 TBC1D4 (Ramm *et al.*, 2006), phos-S939 TSC2 (Cai *et al.*, 2006), phos-S981 TSC2 (Cai *et al.*, 2006) and phos-S99 BAD (Polzien *et al.*, 2009) in the PI3K-Akt signaling pathway. In order to assess the roles of regulatory sites, all regulatory sites of the protein are ranked according to the importance score, which calculated by PTM types and protein counts. Higher score means higher important role in disease process and potential drug targets. For example, the Y654 of CTNNB1 has a high score of 0.93, inhibiting its Phos by Imatinib offered a therapeutic value in patients with chronic myeloid leukemia (CML) (Coluccia *et al.*, 2007), which was in accord with PTM-associated PPI network results (Fig. 2D and E).

According to previous reports (Betts *et al.*, 2017; Shi *et al.*, 2001; Song *et al.*, 2008), PTM sites which located in the interface between two proteins, can regulate protein interactions. We supposed whether PTM sites which not located in the interface, can also enhance or inhibit interactions. Therefore, we analyzed all protein complex and interface amino acids. To our surprise, both interfacial and non-interfacial sites possess the regulatory roles (Supplementary Table S1). For example, phos-Y47 in Fe65 which not localizes in the interface, decreased Fe65 and RASD1 affinity by distal regulation (Lau *et al.*, 2008). Two PTM sites might be associated with crosstalk pattern based on spatial proximity (Brooks and Gu, 2003; Christensen *et al.*, 2005; Minguez *et al.*, 2015), so cooperation and antagonism among several PTM sites in two interacting proteins could be investigated by above labeled spatial location. Moreover, owing to above collected experimental evidence and structural annotation, a high-fidelity machine learning prediction method considering interface information and local microenvironment (Lu *et al.*, 2022) [such as partial charges, spatial location of carbon (C atom), nitrogen (N atom), oxygen (O atom), hydrogen (H atom) and sulfur (S atom), and solvent accessibility], which assessing PTM (such as Phos, Ac, Me, Sumo, Ub and Glyco) effects ('Enhance or 'Inhibit') on PPI (Betts *et al.*, 2015; Schymkowitz *et al.*, 2005) can be developed in the future. Although several algorithms has been developed to predict kinase-specific Phos sites (Wang *et al.*, 2020; Xue *et al.*, 2005), to predict Phos sites that specifically interact with phosphoprotein-binding domains (Guo *et al.*, 2020), a machine learning method to predict the PTM sites that govern PPIs in the view of PTM position, motif length and residues weights based on the sequence window we provided (upstream and downstream five

residues around the PTM site) could be a promising and challenge work.

Due to the limited crystal structures, several software and web-server offer the solution by molecular dock, such as ZDOCK (Pierce *et al.*, 2014), HDOCK (Yan *et al.*, 2020), ClusPro (Kozakov *et al.*, 2017) and HADDOCK (van Zundert *et al.*, 2016). In order to ensure the accuracy of docking results, a large scale of molecular docking was performed combined with experimental XL-MS data (cross-linking) and predicted domain–domain interactions. Furthermore, we also analyzed the interaction (HB, HP and ELE) to help researchers to better understand the PTM roles in disease. Changes of interactions (such as HB, HP and ELE) and structure (such as allostery, disorder-to-order, electrostatic potential and dynamic correlation network) induced by PTM can explain why PTM could regulate PPIs and disease progression (Devanand *et al.*, 2018; Lingyun and Ji-Bin, 2017). Therefore, our database will provide a valuable structural basis for further investigations, such as molecular dynamics simulation (MD) and Markov state models (MSMs). Furthermore, development of a specific forcefield for the simulation of protein–protein complex (Piana *et al.*, 2020) modified by multiple PTM types could be a valuable research filed. Although the statistical results were obtained based on our collected data (Table 1, Supplementary Tables S1–S3), but it might be potentially biased due to the limited PTM types and well-studied proteins.

PTMint database can be further improved in the following aspects. First, the current version of the database contains six main model organisms and PTM types. More organisms and PTM types will be added. Second, additional curations, such as PTM-targeted drugs, PTM expression analysis, PTM-associated survival analysis and association between PTM and mutation will be integrated. Third, we will replace the original predicted complex structures when newly high-resolution complex structures are released in the PDB database. In the future, we will continually maintain and update the PTMint database, when newly regulatory PTM sites are reported in the literature.

In conclusion, we developed PTMint, a comprehensive database of experimentally verified PTM effects on PPIs. We believed that this database should be a useful platform for biologists and bioinformaticians to explore PTM roles on disease development, diagnosis and drug discovery.

Funding

This work was supported by the Center for HPC at Shanghai Jiao Tong University, the National Key Research and Development Program of China [2020YFA0907700]; and the National Natural Science Foundation of China [21977068 and 32171242].

Conflict of Interest: none declared.

Data availability

All data in PTMint are freely accessible at https://ptmint.sjtu.edu.cn/ .

References

Bandy,J. *et al.* (2009) Mining protein-protein interactions from published literature using linguamatics I2E. *Methods Mol. Biol.*, **563**, 3–13.

Beltrao,P. *et al.* (2012) Systematic functional prioritization of protein post-translational modifications. *Cell*, **150**, 413–425.

Berman,H.M. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

Betts,M.J. *et al.* (2015) Mechismo: predicting the mechanistic impact of mutations and modifications on molecular interactions. *Nucleic Acids Res.*, **43**, e10.

Betts,M.J. *et al.* (2017) Systematic identification of phosphorylation-mediated protein interaction switches. *PLoS Comput. Biol.*, **13**, e1005462.

Brooks,C.L. and Gu,W. (2003) Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Curr. Opin. Cell Biol.*, **15**, 164–171.

Cai,S.L. *et al.* (2006) Activity of TSC2 is inhibited by AKT-mediated phosphorylation and membrane partitioning. *J. Cell Biol.*, **173**, 279–289.

Chen,H.F. and Luo,R. (2007) Binding induced folding in p53-MDM2 complex. *J. Am. Chem. Soc.*, **129**, 2930–2937.

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

Christensen,B. *et al.* (2005) Post-translationally modified residues of native human osteopontin are located in clusters: identification of 36 phosphorylation and five O-glycosylation sites and their biological implications. *Biochem. J.*, **390**, 285–292.

Cohen,P. (2002) Protein kinases–the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.*, **1**, 309–315.

Coluccia,A.M.L. *et al.* (2007) Bcr-Abl stabilizes beta-catenin in chronic myeloid leukemia through its tyrosine phosphorylation. *EMBO J.*, **26**, 1456–1466.

Coxon,C.H. *et al.* (2012) An investigation of hierachical protein recruitment to the inhibitory platelet receptor, G6B-b. *PLoS One*, **7**, e49543.

De Las Rivas,J. and Fontanillo,C. (2010) Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.*, **6**, e1000807.

DeLano,W.L. (2010) *PyMOL*. DeLano Scientific, San Carlos, CA, pp. 1–10.

Devanand,T. *et al.* (2018) Phosphorylation promotes binding affinity of Rap-Raf complex by allosteric modulation of switch loop dynamics. *Sci. Rep.*, **8**, 12976.

Dinkel,H. *et al.* (2011) Phospho.ELM: a database of phosphorylation sites-update 2011. *Nucleic Acids Res.*, **39**, D261–D267.

Gao,F.Y. and Chen,Y.M. (2021) Decoding protein phosphorylation signaling networks by mass spectrometry. *Sci. Bull.*, **66**, 2529–2541.

Gu,L. *et al.* (2013) Human DEAD box helicase 3 couples IkappaB kinase epsilon to interferon regulatory factor 3 activation. *Mol. Cell. Biol.*, **33**, 2004–2015.

Guo,Y.P. *et al.* (2020) GPS-PBS: a deep learning framework to predict phosphorylation sites that specifically interact with phosphoprotein-binding domains. *Cells*, **9**, 1266.

Hornbeck,P.V. *et al.* (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.

Hunter,T. (1995) Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell*, **80**, 225–236.

Jumper,J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kozakov,D. *et al.* (2017) The ClusPro web server for protein-protein docking. *Nat. Protoc.*, **12**, 255–278.

Lingyun,W. and Ji-Bin,P. (2017) Phosphorylation of KLHL3 at serine 433 impairs its interaction with the acidic motif of WNK4: a molecular dynamics study. *Protein Sci.*, **26**, 163–173.

Lau,K.F. *et al.* (2008) Dexras1 interacts with FE65 to regulate FE65-Amyloid precursor protein-dependent transcription. *J. Biol. Chem.*, **283**, 34728–34737.

Li,D. *et al.* (2018) ECharts: a declarative framework for rapid construction of web-based visualization. *Visual Informatics*, **2**, 136–146.

Li,X. *et al.* (2012) Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions. *J. Am. Chem. Soc.*, **134**, 1982–1985.

Li,Z. *et al.* (2022) dbPTM in 2022: an updated database for exploring regulatory networks and functional associations of protein post-translational modifications. *Nucleic Acids Res.*, **50**, D471–D479.

Lin,J. *et al.* (2021) A tri-functional amino acid enables mapping of binding sites for posttranslational-modification-mediated protein-protein interactions. *Mol. Cell.*, **81**, 2669–2681.e9.

Lu,H. *et al.* (2022) Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature*, **604**, 662–667.

Meszaros,B. *et al.* (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.

Meyer,M.J. *et al.* (2013) INstruct: a database of high-quality 3D structurally resolved protein interactome networks. *Bioinformatics*, **29**, 1577–1579.

Minguez,P. *et al.* (2015) PTMcode v2: a resource for functional associations of post-translational modifications within and between proteins. *Nucleic Acids Res.*, **43**, D494–D502.

Mistry,J. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
Piana,S. *et al.* (2020) Development of a force field for the simulation of single-chain proteins and protein-protein complexes. *J. Chem. Theory Comput.*, **16**, 2494–2507.
Pierce,B.G. *et al.* (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*, **30**, 1771–1773.
Polzien,L. *et al.* (2009) Identification of novel in vivo phosphorylation sites of the human proapoptotic protein BAD: pore-forming activity of BAD is regulated by phosphorylation. *J. Biol. Chem.*, **284**, 28004–28020.
Qin,F. *et al.* (2010) Induced fit or conformational selection for RNA/U1A folding. *RNA*, **16**, 1053–1061.
Ramm,G. *et al.* (2006) A role for 14-3-3 in insulin-stimulated GLUT4 translocation through its interaction with the RabGAP AS160. *J. Biol. Chem.*, **281**, 29174–29180.
Rego,N. and Koes,D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, **31**, 1322–1324.
Schymkowitz,J. *et al.* (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.
Seet,B.T. *et al.* (2006) Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.*, **7**, 473–483.
Shi,W.X. *et al.* (2001) Structural analysis of adenine phosphoribosyltransferase from *Saccharomyces cerevisiae*. *Biochemistry*, **40**, 10800–10809.
Singh,A. *et al.* (2010) Protein phosphatase 2A reactivates FOXO3a through a dynamic interplay with 14-3-3 and AKT. *Mol. Biol. Cell.*, **21**, 1140–1152.
Song,C.Y. *et al.* (2008) Serine 88 phosphorylation of the 8-kDa dynein light chain 1 is a molecular switch for its dimerization status and functions. *J. Biol. Chem.*, **283**, 4004–4013.
Spektor,T.M. and Rice,J.C. (2009) Identification and characterization of post-translational modification-specific binding proteins in vivo by mammalian tethered catalysis. *Proc. Natl. Acad. Sci. USA*, **106**, 14808–14813.
Swietlik,J.J. *et al.* (2020) Dissecting intercellular signaling with mass spectrometry-based proteomics. *Curr. Opin. Cell Biol.*, **63**, 20–30.
Takala,H. *et al.* (2008) Beta2 integrin phosphorylation on Thr758 acts as a molecular switch to regulate 14-3-3 and filamin binding. *Blood*, **112**, 1853–1862.
UniProt,C. (2012) Reorganizing the protein space at the universal protein resource (UniProt). *Nucleic Acids Res.*, **40**, D71–D75.
van Zundert,G.C.P. *et al.* (2016) The HADDOCK2.2 web server: user-Friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.*, **428**, 720–725.
Varadi,M. *et al.* (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.
Wang,C. *et al.* (2020) GPS 5.0: an update on the prediction of kinase-specific phosphorylation sites in proteins. *Genomics Proteomics Bioinformatics*, **18**, 72–80.
Wang,S. *et al.* (2022) Uncovering post-translational modification-associated protein-protein interactions. *Curr. Opin. Struct. Biol.*, **74**, 102352.
Wang,W. *et al.* (2014) New force field on modeling intrinsically disordered proteins.) *Chem. Biol. Drug Des.*, **84**, 253–269.
Wu,S. *et al.* (2012) Casein kinase 1alpha regulates an MDMX intramolecular interaction to stimulate p53 binding. *Mol. Cell. Biol.*, **32**, 4821–4832.
Xu,H. *et al.* (2018) PTMD: a database of human disease-associated post-translational modifications. *Genomics Proteomics Bioinformatics*, **16**, 244–251.
Xu,H.D. *et al.* (2017) PLMD: an updated data resource of protein lysine modifications. *J. Genet. Genomics*, **44**, 243–250.
Xu,P. *et al.* (2022) Erythrocyte transglutaminase-2 combats hypoxia and chronic kidney disease by promoting oxygen delivery and carnitine homeostasis. *Cell Metab.*, **34**, 299–316.e6.
Xue,Y. *et al.* (2005) GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res*, **33**, W184–W187.
Yan,Y.M. *et al.* (2020) The HDOCK server for integrated protein-protein docking. *Nat. Protoc.*, **15**, 1829–1852.
Yu,K. *et al.* (2019) qPhos: a database of protein phosphorylation dynamics in humans. *Nucleic Acids Res.*, **47**, D451–D458.