

# Chem(Pro)<sup>2</sup>: the atlas of *chemoproteomic* probes labelling human proteins

Songsen Fu<sup>1,2,†</sup>, Zhen Chen<sup>3,†</sup>, Zhiming Luo<sup>2</sup>, Meiyun Nie<sup>2</sup>, Tingting Fu<sup>3</sup>, Ying Zhou<sup>3</sup>, Qingxia Yang<sup>4,\*</sup>, Feng Zhu<sup>3,5,\*</sup> and Feng Ni<sup>1,2,\*</sup>

<sup>1</sup>Institute of Drug Discovery Technology, Ningbo University, Ningbo 315211, China

<sup>2</sup>LeadArt Biotechnologies Ltd., Ningbo 315201, China

<sup>3</sup>College of Pharmaceutical Sciences, The Second Affiliated Hospital, Zhejiang University School of Medicine, State Key Laboratory of Advanced Drug Delivery and Release Systems, Zhejiang University, Hangzhou 310058, China

<sup>4</sup>Zhejiang Provincial Key Laboratory of Precision Diagnosis and Therapy for Major Gynecological Diseases, Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China

<sup>5</sup>Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 330110, China

\*To whom correspondence should be addressed. Tel: +86 189 8946 6518; Fax: +86 571 8820 8444; Email: zhufeng@zju.edu.cn

Correspondence may also be addressed to Feng Ni. Email: nifeng@nbu.edu.cn

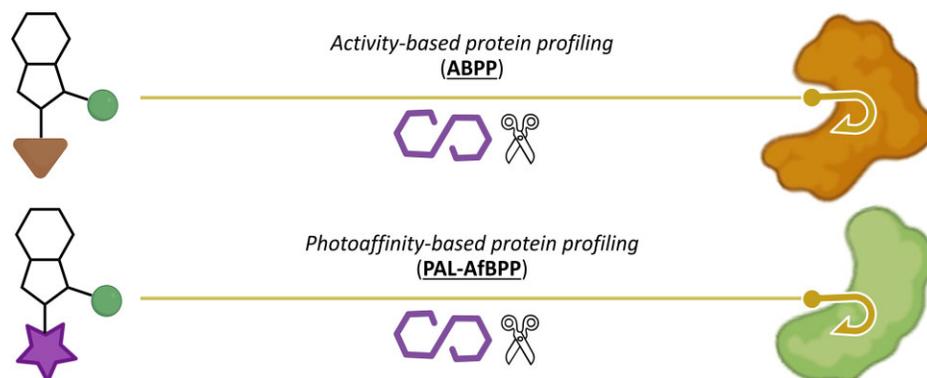
Correspondence may also be addressed to Qingxia Yang. Email: yangqx@zju.edu.cn

†The first two authors should be regarded as Joint First Authors.

## Abstract

*Chemoproteomic* probes (CPPs) have been widely considered as powerful molecular biological tools that enable the highly efficient discovery of both binding proteins and modes of action for the studied compounds. They have been successfully used to validate targets and identify binders. The design of CPP has been considered extremely challenging, which asks for the generalization using a large number of probe data. However, none of the existing databases gives such valuable data of CPPs. Herein, a database entitled 'Chem(Pro)<sup>2</sup>' was therefore developed to systematically describe the atlas of diverse types of CPPs labelling human protein in living cell/lysate. With the booming application of *chemoproteomic* technique and *artificial intelligence* in current chemical biology study, Chem(Pro)<sup>2</sup> was expected to facilitate the AI-based learning of interacting pattern among molecules for discovering innovative targets and new drugs. Till now, Chem(Pro)<sup>2</sup> has been open to all users without any login requirement at: <https://idrblab.org/chemprosquare/>

## Graphical abstract



## Introduction

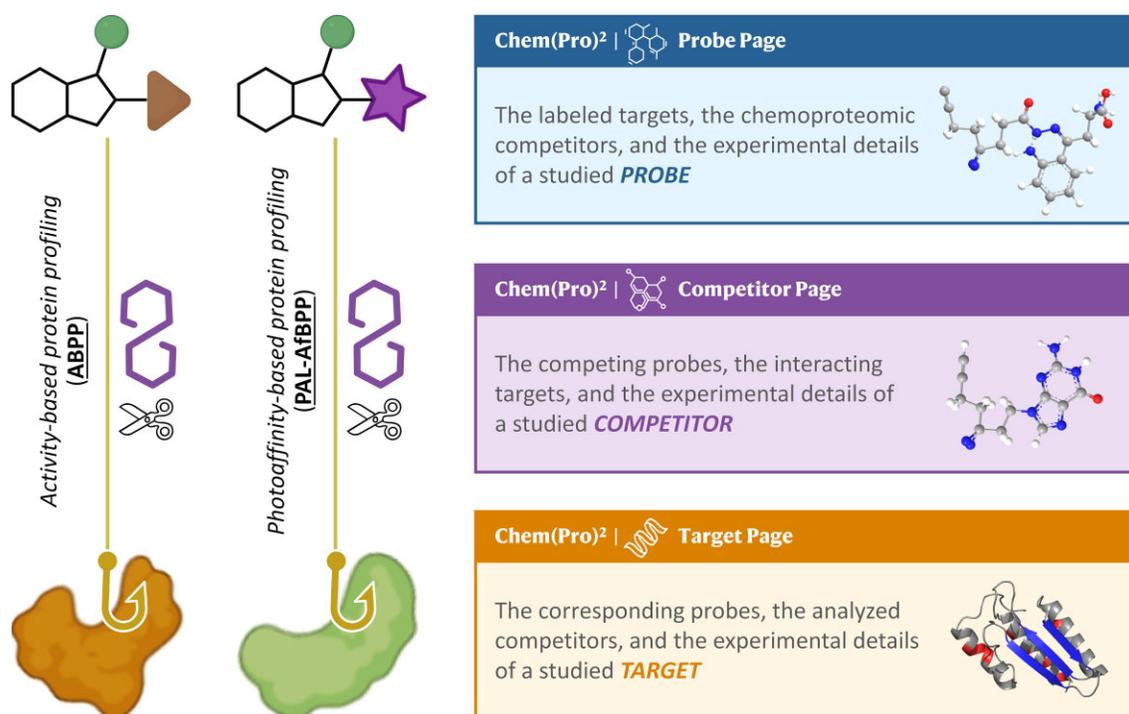
Chemoproteomic probes (CPPs) have been widely considered as a powerful molecular biological tool that enables the highly efficient discovery of both binding proteins and modes of action for studied compounds (1–3). They have been successfully applied in various contexts, including identifying/validating drug targets by mapping compound-protein interactions (4–8), screening lead compounds in early drug discovery (9–12)

and optimizing the efficacy/specificity of the identified leads (13–16). It is known that the design of CPP remains extremely challenging, which asks for the generalization based on a large number of probe data (17). All the studies highlight the demands for the valuable information of CPP, and it is therefore critical to collect the *big data* of CPP for promoting the AI-based learning of interacting patterns among molecules for discovering innovative targets and drugs (18–27).

Received: August 15, 2024. Revised: September 25, 2024. Editorial Decision: October 6, 2024. Accepted: October 11, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** The relationships among the three key components in our Chem(Pro)<sup>2</sup> database: CPPs, *competitors* and probe labelling targets. One CPP can label many targets, and specific competitor can compete with the CPP when binding to targets. As described on the left side, there were two types of probe-associated techniques: *activity-based protein profiling* (ABPP) and *photoaffinity-based protein profiling* (PAL-AfBPP). As demonstrated on the right side, the primary content in each of the Chem(Pro)<sup>2</sup> page (the pages for *Probe/Competitor/Target*) was also provided.

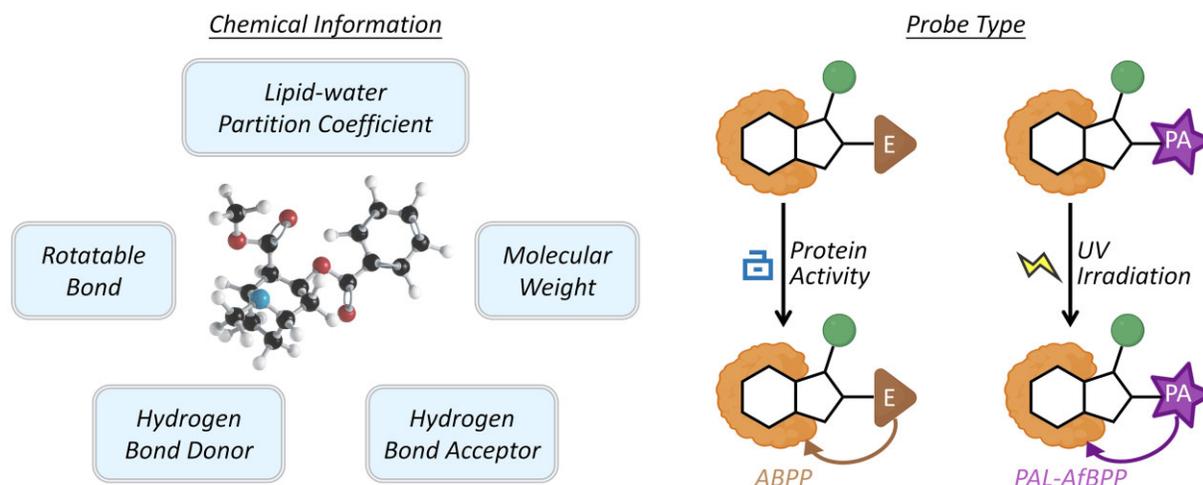
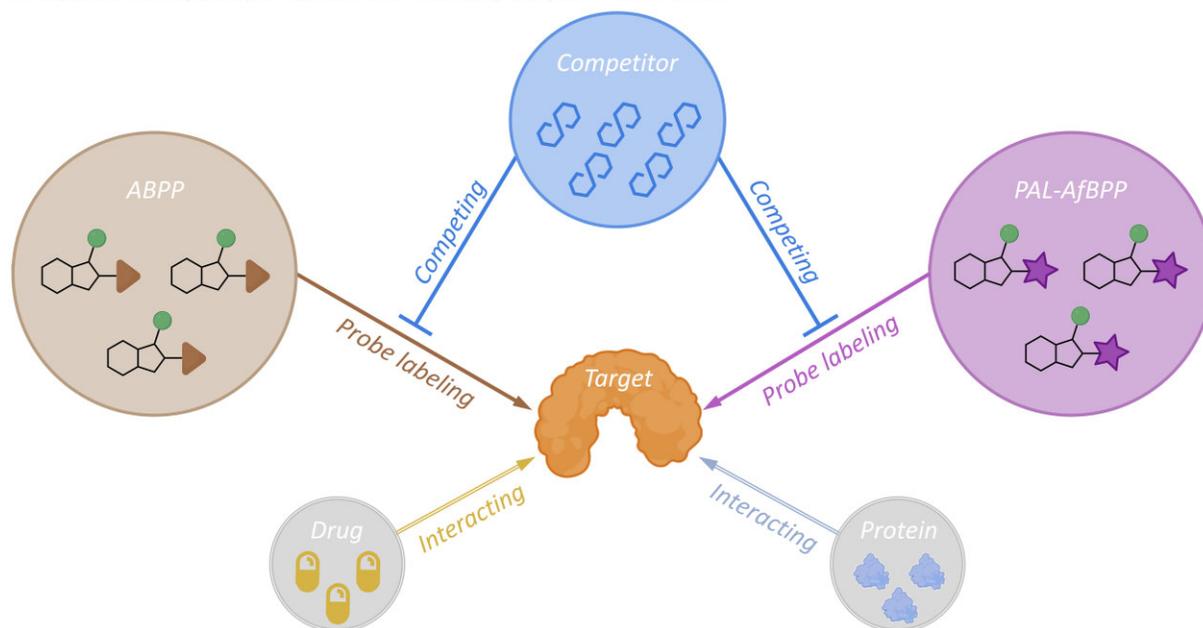
So far, a variety of databases have been developed to provide the interactions between compound and protein, including BindingDB (28), ChEMBL (29), CovPDB (30), GtoPdb (31), KLIFS (32), Chemical Probes Portal (33) and CovalentInDB (34). Only one database named CysDB (35) has been available for describing CPP data, which focuses on offering three CPPs covalently binding to the *cysteine* sites of 11 621 human proteins detected using the *chemoproteomic* technique, and 363 competitive compounds (namely, *competitors*) corresponding to the CPPs were accumulated using experimentally-validated data. Most of the existing databases have attracted broad interest from worldwide audiences and won substantial reputation from the relevant research community, which further highlights the great importance of the accumulation of these valuable CPP data.

However, it is known that the probes are designed to covalently bind to any amino acids (not just cysteine) of the human proteins, and the diversity of the binding amino acid is of great importance for the identification of the target/drug (36–42). Besides the covalent ones, photoaffinity probes are known to be appealing and promising to noncovalent drugs for the enrichment of targets from the whole cellular proteome (43,44), and it is critical to have the binding data of probes validated in living cell to represent the complexity of the biological system (45–48). To the best of our knowledge, none of the existing databases provided such valuable information as described above. Therefore, it is urgently needed to have a knowledge base that systematically describes the precious data of diverse types of *chemoproteomic probes* (CPPs) labelling the human proteins in living cells.

Herein, a database named Chem(Pro)<sup>2</sup>, which systematically describes the atlas of *chemoproteomic probes* (CPPs) la-

labelling human proteins in living cells was therefore developed. First, CPPs were collected based on comprehensive literature reviews in PubMed, which led to a total of 603 CPPs (133 covalent and 470 photoaffinity), 1016 *competitors* of CPPs and 14 250 binding proteins in the human genome (including 4649 enzymes, 1357 channel/transporters, 225 GPCRs and so on). Second, all interactions between CPPs and their labelling human proteins were experimentally discovered based on a total of 118 living cell types (such as HEK293T, HCT-116, PaTu-8988t, MDA-MB-231, HeLa) from 22 healthy/disease organs (including kidney, colon, uterus, breast, pancreas, and liver). Third, the information on the probe's binding sites in human protein was comprehensively collected from the literatures and explicitly provided in Chem(Pro)<sup>2</sup>, which resulted in a total of 135 486 binding sites in 13 938 human proteins labelled by 160 CPPs. Finally, the binding ratios between 524 CPPs and 863 *competitors* (which challenge the labelling of CPPs on proteins) were collected, which brought about 2 118 636 records of binding ratio experimentally validated based on 106 living cell types (including K562, Hep-G2, Ramos, SH-SY5Y and KYSE410).

All in all, Chem(Pro)<sup>2</sup> is unique in systematically providing diverse types of *chemoproteomic probes* (both the covalent and photoaffinity ones) that experimentally label human proteins in living cells. Its provided data were fully cross-linked to many established molecular biological knowledge bases, such as UniProt (49), PubChem (50), TTD (51), DrugMAP (52) and ChEMBL (29). With the booming applications of *chemoproteomic techniques* in chemical biology research, our database is therefore expected to attract broad interests from the relevant research community. The Chem(Pro)<sup>2</sup> is now freely accessible by all users at: <https://idrblab.org/chemprosquare/>.

**A** Explicit Description for Each Chemoproteomic Probe in Chem(Pro)<sup>2</sup>**B** Explicit Description for Each Probe-labelling Target in Chem(Pro)<sup>2</sup>

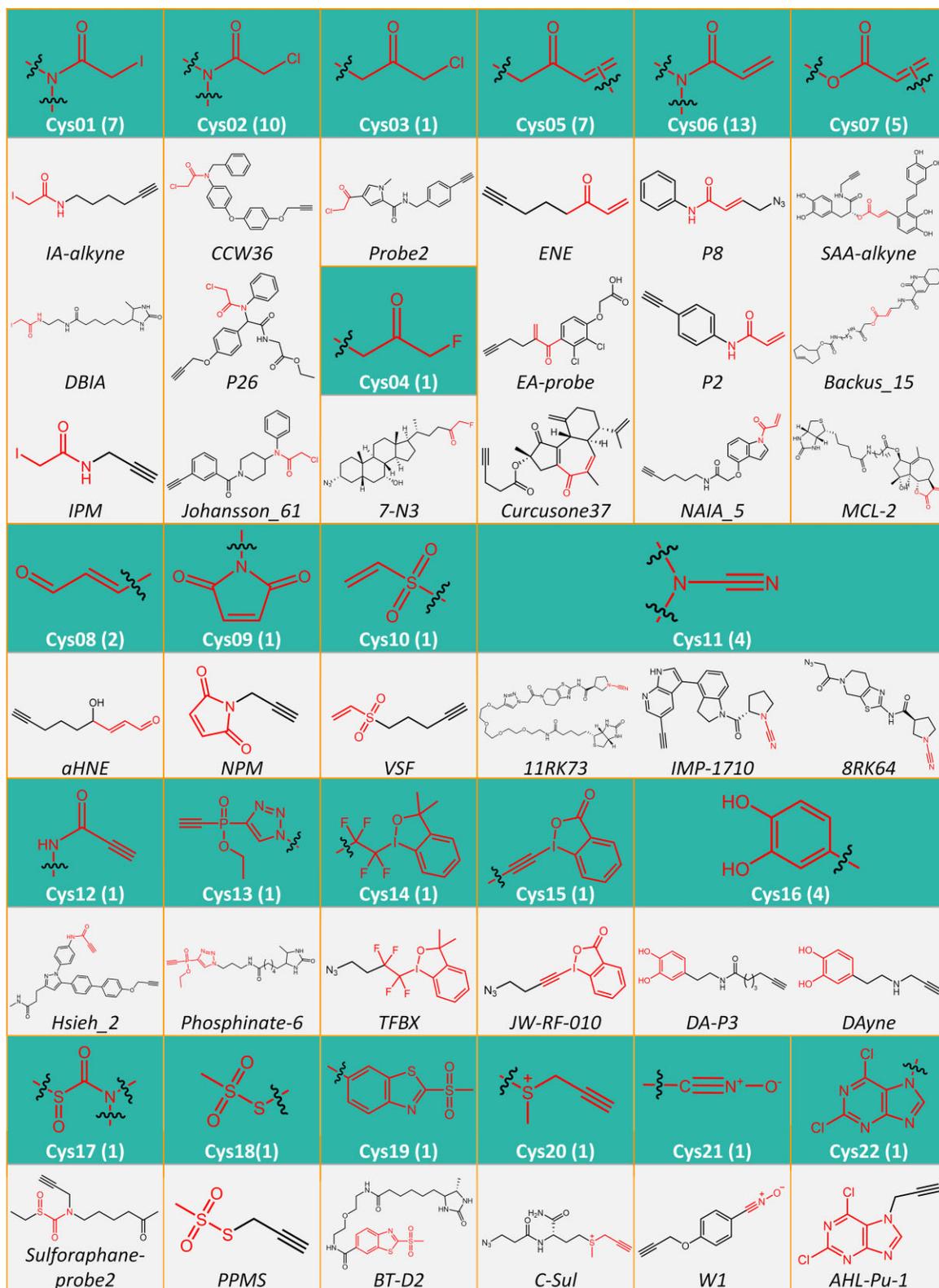
**Figure 2.** Detailed information offered in Chem(Pro)<sup>2</sup> for describing the *chemoproteomic* probes (CPPs) and their labelling targets. **(A)** explicit descriptions for each CPP in Chem(Pro)<sup>2</sup> database, which included various chemical data (such as lipid-water partition coefficient, molecular weight, hydrogen bond donor, hydrogen bond acceptor, and rotatable bond) and diverse probe techniques (ABPP and PAL-AfBPP). **(B)** explicit descriptions for every probe-labelling target in Chem(Pro)<sup>2</sup>, which described the competing mechanism of *competitors* underlying each probe technique, and a variety of drugs and proteins that interacted with the studied probe-labelling target.

**Factual content and data retrieval****Data collection for chemoproteomic probes (CPPs), competitors and targets**

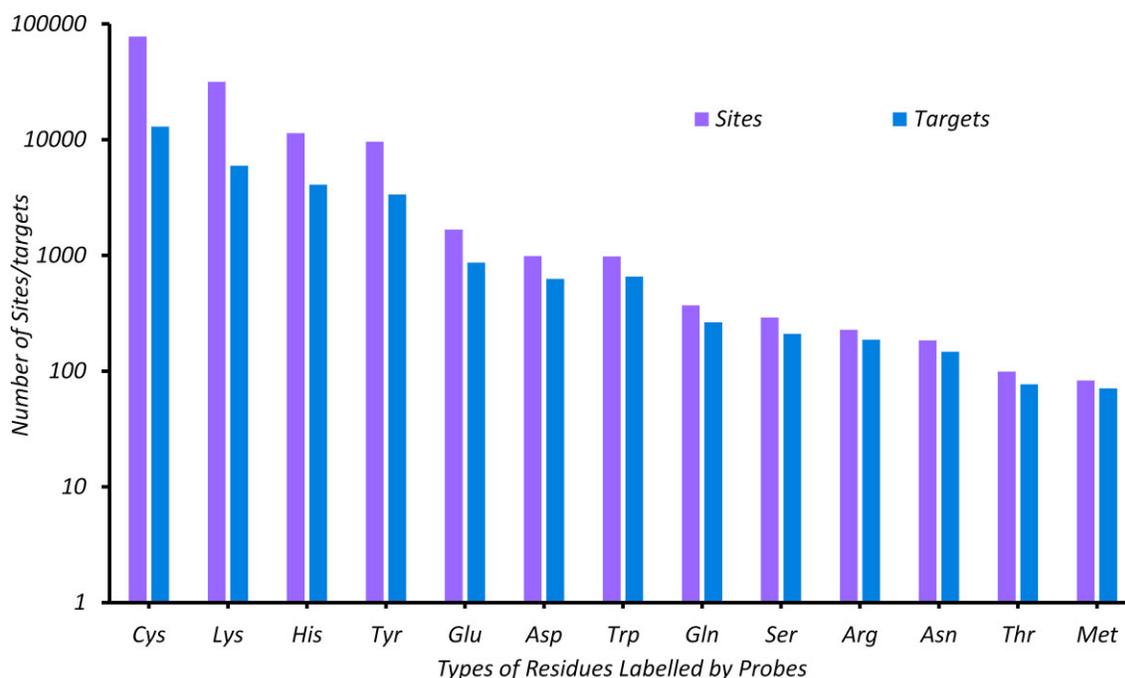
The data for Chem(Pro)<sup>2</sup> were collected through the following procedures. Particularly, a literature review was conducted using keywords and combinations such as ‘chemoproteomics + probes’, ‘chemoproteomics + targets’, ‘activity-based protein profiling’, ‘affinity-based protein profiling’, ‘photoaffinity labelling’ and ‘click chemistry’ in PubMed and PRIDE (53). Entries detailing reported probes along with their corresponding target were documented. As a result, Chem(Pro)<sup>2</sup> established a comprehensive *chemoproteomic atlas* that integrated a diverse series of CPPs with their labelling

targets and *competitors*. Experimentally, each studied CPP can label multiple targets and be competed by several competitors, which established the relationship among the three main components in Chem(Pro)<sup>2</sup> database: CPP, *competitor*, and target.

As depicted on the left side of Figure 1, there were two typical types of probe-protein profiling technique: *activity-based protein profiling* (ABPP (54)) and *photoaffinity-based protein profiling* (PAL-AfBPP (55)). The probes in Chem(Pro)<sup>2</sup> were specifically designed for labeling and enriching the target protein profiles in proteomics analyses and the key differences between probes in Chem(Pro)<sup>2</sup> and the Chemical Probes Portal was shown in Supplementary Table S1. Moreover, the main contents in the Chem(Pro)<sup>2</sup> pages describing those three components were briefly provided (the right side of Figure 1),



**Figure 3.** The atlas of the activity-based protein profiling (ABPP) probe labelling cysteine. There were 22 types of ABPP probes under this category (from Cys01 to Cys22, highlighted in green background, the number in bracket denoted the total amount of ABPP probes within certain probe type). Under each probe type, both structure and name of the representative probes were provided, and the corresponding probe warhead was highlighted in red color. A square frame in orange was used to highlight the specific information for each cysteine-labelling probe type.



**Figure 4.** Different numbers of probe-labelled sites and targets according to the various types of residues labelled by all probes described by Chem(Pro)<sup>2</sup>. Among all existing amino acids, a total of thirteen amino acids were identified in this study to be labelled by at least one *chemoproteomic* probe, which made Chem(Pro)<sup>2</sup> covering the most comprehensive types of probe-labelled amino acids among available databases. Particularly, the *cysteine* was the most popular one labelled by probe, while the *methionine* was the one with the least numbers of sites and targets.

which included the labeled targets, *chemoproteomic competitors*, and experimental detail of the studied probe, the competing probes, interacting targets, and experimental detail of a studied competitor, and the corresponding probes, analyzed competitors, and experimental detail of the studied target. In Chem(Pro)<sup>2</sup>, readers could readily retrieve the information of the three main components and their one-to-many relationship, and 603 CPPs (133 covalent and 470 photoaffinity), 1016 *competitors* and 14 250 protein targets (such as 4649 enzymes and 1357 channel/transporters) were finally included into Chem(Pro)<sup>2</sup>.

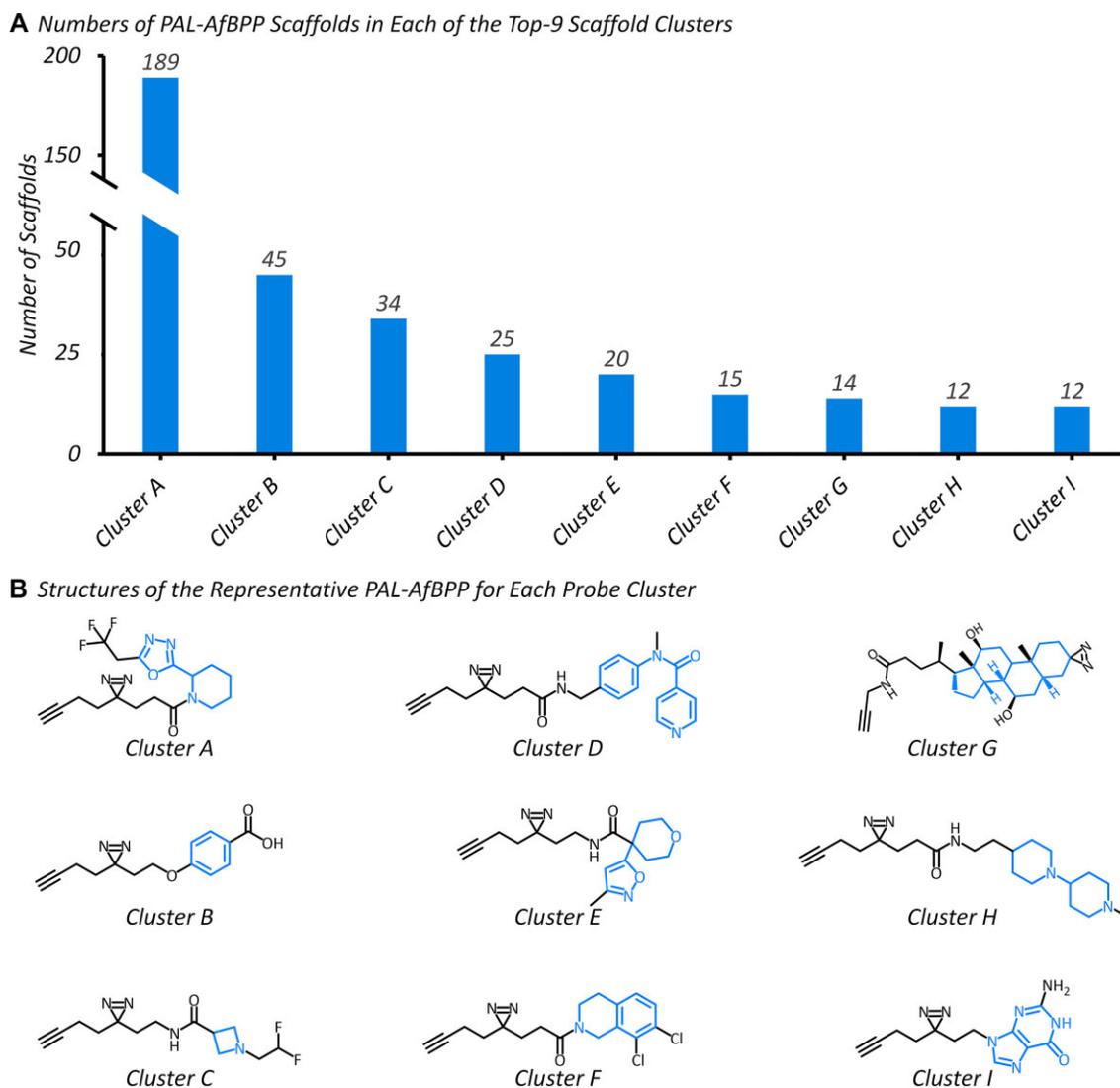
Furthermore, a variety of detailed information was incorporated into our Chem(Pro)<sup>2</sup> to describe the *chemoproteomic* probes and their labelling target. As depicted in Figure 2A, *chemoproteomic* probes in Chem(Pro)<sup>2</sup> were systematically described by various chemical data (such as lipid-water partition coefficients, molecular weight, rotatable bond, hydrogen bond donor and hydrogen bond acceptor) and probe techniques (ABPP and PAL-AfBPP). Meanwhile, the probe-labelling targets were also explicitly introduced and shown in Figure 2B. Chem(Pro)<sup>2</sup> introduced all interactions and labelling processes involving a studied probe, and the competing mechanisms of *competitors* underlying different probe techniques were also shown. Moreover, a large number of drugs (2033 approved, 2379 clinical trial, 3588 preclinical/patented and 9631 investigative ones) and 10 533 proteins interacting with the probe-labelling targets were also collected.

### Activity-based protein profiling (ABPP) probes labelling the protein targets

ABPP is a well-used *chemoproteomic* technique initially developed to monitor enzyme activities in complex proteomes

(56–63). Various types of enzymes have been targeted using ABPP probes, which include proteases, kinases, phosphatases, glycosidases, oxidoreductases, and so on (64–67). To overcome the limitation of traditional protein profiling techniques, broad-spectrum ABPP probes with highly reactive warheads have been designed to label the intrinsically reactive residues across various protein families (68). For example, the iodoacetamide-derived probe ‘IA-alkyne’ labelled over 1000 functional cysteines in human cells (69), which included a list of hyper-reactive cysteines associated with enzyme’s functional sites, highlighting the versatility of broad-spectrum probes. In addition to the large number of probes targeting cysteine (70), other amino acid residues, such as lysine, tyrosine, aspartic acid, glutamic acid, histidine, tryptophan and methionine, have also been frequently adopted in current *chemoproteomic* studies (71–74). The key features of the probes included (a) a chemically reactive warhead that covalently binds to a specific residue/cofactor on the labelled target, (b) a reporter tag for detecting/purifying the labelled targets and/or (c) a binding group that fine-tunes target-labelling specificity and minimizing undesirable interactions between the warhead and the tag.

In Chem(Pro)<sup>2</sup>, a comprehensive collection of ABPP probes along with their labelled targets was provided. As one of the most popular labelling residues, cysteine was reported to be the only one covalently labelled by marketed drugs, which highlighted its dominant role in *chemoproteomic* research. In this study, the atlas of all the cysteine-labelled ABPP probes collected to Chem(Pro)<sup>2</sup> was systematically described in Figure 3, which resulted in 22 types of ABPP probe (from Cys01 to Cys22, highlighted in green background, the numbers in brackets indicated the total number of collected probes in the corresponding probe type). In

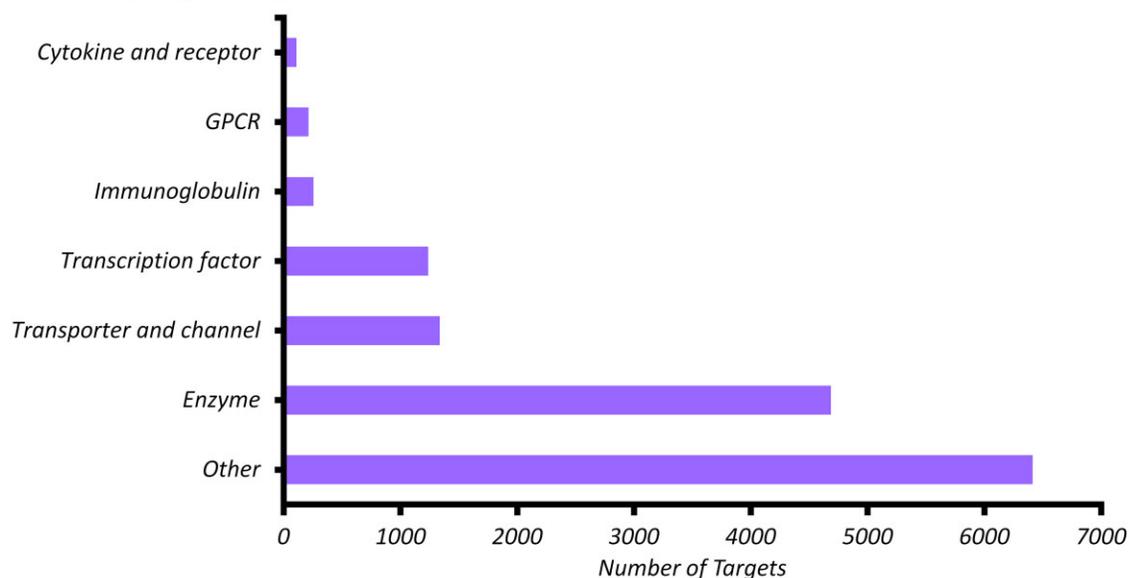
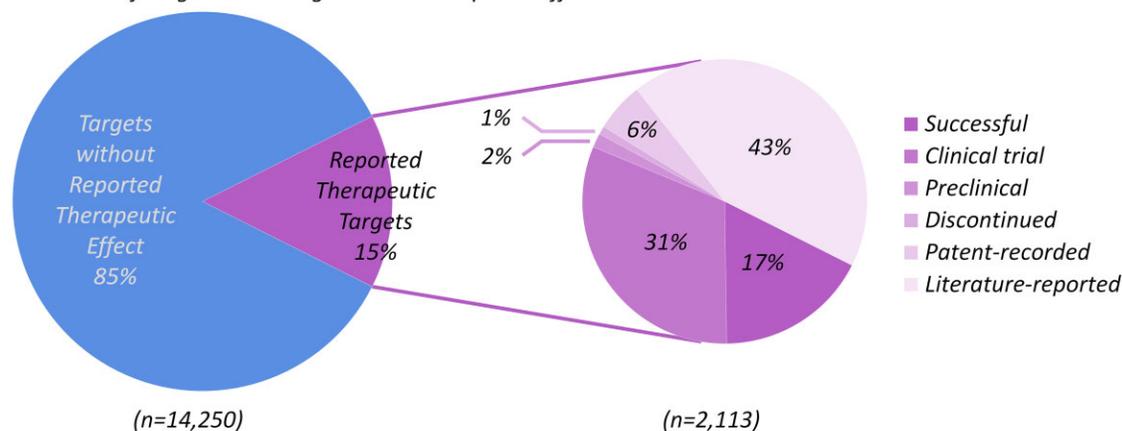


**Figure 5.** The number and structure of *photoaffinity-based protein profiling* (PAL-AfBPP) probe scaffold in different scaffold clusters. **(A)** the number of PAL-AfBPP scaffolds in each of the top-9 scaffold cluster. **(B)** the structure of the representative PAL-AfBPP probe for each probe cluster, and the corresponding scaffold was highlighted in blue color.

each probe type, both the structure and name of the representative ABPP probes were illustrated, with the corresponding warhead highlighted in red color. A number of square frames in orange were used to highlight certain data for each cysteine-labelled probe type. Particularly, the commonly-used warheads (colored in red) included electron-deficient haloalkane or electron-deficient alkene, leading to the development of diverse ABPP probes, such as *ENE* (17), *AZ-9* (36), *IA-alkyne* (69), *IPM* (75), *DBIA* (76,77) and *NAIA\_5* (78). Other warheads incorporated many chemotypes: (a) electron-deficient alkyne probes (such as *Hsieh\_2* (79), and *phosphinate-6* (80)), (b) benziodoxole probes (such as *TFBX* (81) and *JW-RF-010* (82)), (c) catechol probes (such as *DA-P3* (83), and *DAyne* (84)), (d) sulfonium probes (such as *C-Sul* (85)), (e) nitrile oxide probes (such as *W1* (86)) and so on.

Additionally, many CPPs in Chem(Pro)<sup>2</sup> database were discovered to label the post-translational modification of cysteine. As provided in [Supplementary Figure S1](#), these

CPPs belonged to four different classes: 4-hydroxy-2-nonenal-modified cysteine (*Cys-HNE*), acrolein-modified cysteine (*Cys-Acrolein*), sulfenylation of cysteine (*Cys-SOH*) and sulfination of cysteine (*Cys-SO<sub>2</sub>H*). Taking the type of *Cys-SOH* as an example, there were three types: (a) Wittig reagent probes (such as *WYneX* (87)), (b) benzothiazine probes (such as *BTD* (88)), (c) dimedone probes (such as *DYn-2* (89)). Except for cysteine and its post-translational modifications, the ABPP probes for other amino acids were also collected to Chem(Pro)<sup>2</sup>. As shown in [Supplementary Figure S2](#), the atlas of *activity-based protein profiling* (ABPP) probes labelling the residues beyond cysteine and its post-translational modification was systematically described, with detailed information was explicitly discussed in the corresponding figure legend. All in all, Chem(Pro)<sup>2</sup> provided 135 486 ABPP probe-labelled sites relevant to 13 amino acids in 13 938 human proteins (as shown in Figure 4), which made it the most comprehensive repository of probe-labelled amino acids among existing knowledge bases.

**A** Numbers of Targets in Different Protein Functional Classes**B** Statistics of Targets according to Their Therapeutic Effects and Clinical Trial Statuses

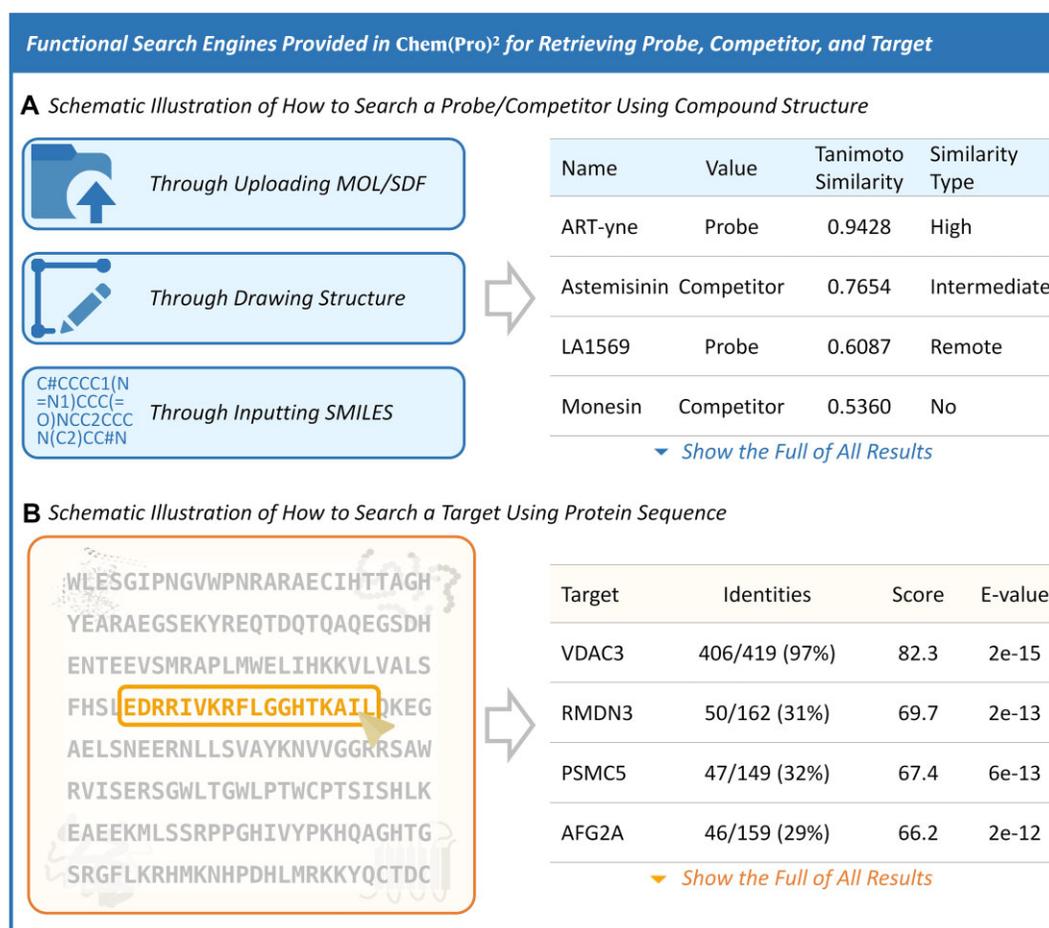
**Figure 6.** A systematic classification of probe-labelled targets. **(A)** the numbers of targets among a variety of protein functional classes. **(B)** the percentages of targets according to their therapeutic effects and clinical trial status. Particularly, about 15% of the probe-labelled targets had already been previously reported as ‘therapeutic target’, which could be further classified into successful, clinical trial, preclinical, discontinued, patent-recorded and literature-reported ones.

### Photoaffinity-based protein profiling (PAL-AfBPP) probes labelling targets

Most FDA-approved drugs bind non-covalently to protein targets. Tools for evaluating reversible interactions between small molecules and proteins are invaluable in this context. To address this need, the *affinity-based protein profiling* (AfBPP) strategy incorporated with *photoaffinity labelling* (PAL) (90) was developed. PAL-AfBPP probe (commonly known as the *photoaffinity-based protein profiling* probe) has become a versatile tool for mapping the reversible interactions between small molecules and proteins in living cells, aiding in the identification of protein targets for various molecules (such as lead compounds and natural product (91–93)), which significantly expanded the spectrum of ‘ligandable’ proteins (94–97). A typical PAL-AfBPP probe consists of three essential components: a scaffold that facilitates the probe’s approach to the targeted proteins, a photoreactive group that forms a covalent bond with the target upon exposure to radiation, and a

reporter tag for subsequent detection or enrichment. Typical photoreactive groups included aryl azides, diazo ketones, diazirines, benzophenones and aryl tetrazoles, with diazirines being favored due to high reactivity and compact size. While the type and attached sites of the photoreactive group impacted labelling profile of scaffold (98,99), the overall landscape of protein labelling was predominantly determined by the diversity of scaffold structure. Till now, a large number of PAL-AfBPP probes and their labelled proteins (95,96) have been generated within the research community, making it an opportune time to construct a comprehensive database, which would give substantial insight into the design of innovative *chemoproteomic* probes and the discovery of new therapeutics.

The latest Chem(Pro)<sup>2</sup> contained 488 PAL-AfBPP probes labelling 4 864 human proteins. Given the importance of scaffold diversity, the structure variety of all accumulated scaffolds was further investigated. First, RDKit package (100) was uti-



**Figure 7.** The functional search engines provided in Chem(Pro)<sup>2</sup> for retrieving probe, competitor, and target. **(A)** a schematic illustration of how to search a probe/competitor based on user-defined compound structures. Chem(Pro)<sup>2</sup> engine could output the compound similarity by matching the user-defined one to all the probe/competitor structures in our database, and a variety of similarity degrees were calculated (High Similarity, Intermediate Similarity, and Remote Similarity). **(B)** a schematic illustration of how to search probe-labelled target using user-defined protein sequence. Chem(Pro)<sup>2</sup> engine could calculate the sequence similarity by aligning the user-defined sequence against all target sequences collected to our database, and similarity levels were measured using the BLAST *E-values* and *Identities* as the quantitative similarity scores.

lized to generate *Murcko scaffolds* and compute *Morgan fingerprints* for all probes, and the similarity among probes was calculated using *Tanimoto* coefficients (101). Second, based on these similarity scores, hierarchical clustering of all probes was performed using a linkage strategy to calculate the distances among clusters with the threshold set to 0.85 (102). Third, within each cluster, the representative scaffolds were discovered by calculating the average fingerprint and identifying the scaffold closest to this average. Finally, a total of 35 clusters were generated, and the numbers of PAL-AfBPP scaffolds in each of the top-9 scaffold clusters were provided in Figure 5A. Moreover, the structures of the representative PAL-AfBPP probe scaffolds within those top-9 clusters were described in Figure 5B, which demonstrated considerable structure diversity among all the collected probes.

#### Target diversity in functional and clinical relevance

Chem(Pro)<sup>2</sup> collected a total of 14 250 probe-labelling targets. To have a deeper understanding of these targets, both functional and clinical relevance diversities of the labelled protein targets were explicitly depicted in Figure 6A and B, respectively. Particularly, the targets fell in different protein functional classes, such as enzymes,

transporters/channels, transcription factors, immunoglobulins, GPCRs, and cytokines/cytokine receptors (as demonstrated in Figure 6A). Furthermore, all probe-labelling targets were mapped to a well-known database TTD (103), which described the statistics of those targets based on their therapeutic effects and clinical trial status (as illustrated in Figure 6B). Specifically, 15% of those targets were reported therapeutic ones, among which 17% and 31% were the ones of approved (successful) and clinical trial drugs, respectively. For the targets of clinical importance, those established CPPs could play critical role in evaluating target occupancy for drugs associated with preclinical or clinical targets, as well as facilitating ligand screening for literature-reported target. For those targets yet to enter the clinical trials, the CPPs in Chem(Pro)<sup>2</sup> could also be proactively employed for promoting the studies of ligand screening, thereby capturing the opportunities for lead/drug discovery.

#### Functional search engines for retrieving the probe, competitor, and target

A variety of search engines were constructed in this study and provided online in our Chem(Pro)<sup>2</sup> for retrieving the information of probe, competitor and target. Chem(Pro)<sup>2</sup> has

developed a keyword-based search engine that allows users to search for drugs, competitors, and targets across the entire database. For example, in the ‘Search for Probe’ interface, users can directly search for probes by name (e.g. IA alkyne, AZ-9, A-DA), or retrieve a list of probes associated with specific targets (e.g. SLC25A20) or competitors (e.g. JZ128). What’s more, Figure 7A schematically illustrates the way to search for a probe/competitor based on the user-defined compound structures. Particularly, the provided engine could output the compound similarities by matching the user-defined one to all probe/competitor structures in Chem(Pro)<sup>2</sup> database, and a variety of similarity degrees could then be calculated, and classified into different levels of similarity (High Similarity, Intermediate Similarity and Remote Similarity). Meanwhile, Figure 7B further provided the way to search probe-labelled targets based on the user-defined protein sequence. Chem(Pro)<sup>2</sup> engine could calculate sequence similarities by aligning the user-defined sequence against all target sequences collected to the database, and similarity levels were measured using the BLAST *E-values* and *Identities* as the quantitative similarity scores (104). All in all, by integrating such sophisticated measurements, our database was expected to streamline the search procedure by offering powerful and user-friendly interfaces that aimed at supporting efficient data retrieval and scientific data analysis.

## Data availability

Chem(Pro)<sup>2</sup> is open to all users without login requirement at: <https://idrblab.org/chemprosquare/>.

## Supplementary data

Supplementary Data are available at NAR Online.

## Funding

National Natural Science Foundations of China [42 388 101, 22 304 091, 82 373 790, 22 220 102 001, U1909208]; Ningbo Top Talents Project [215-432094250]; Scientific Research Grant of Ningbo University [215-432000282]; Fundamental Research Funds for the Central University [2018QNA7023]; Natural Science Foundations of Zhejiang Province [LR21H300001]; National Key R&D Program of China [2022YFC3400501]; Key R&D Programs of the Zhejiang Province [2020C03010]; Double Top-Class University [181 201\*194 232 101]; Leading Talents of the ‘Ten Thousand Plan’ National High-Level Talent Support Plan of China; Alibaba-Zhejiang University Joint Research Centers of Future Digital Healthcare; Westlake Laboratory (Westlake Lab of Life Science & Biomedicine); Alibaba Cloud; Information Technology Center of Zhejiang University. Funding for open access charge: Scientific Research Grant of Ningbo University [215-432000282].

## Conflict of interest statement

None declared.

## References

1. Kulkarni,R.A., Bak,D.W., Wei,D., Bergholtz,S.E., Briney,C.A., Shrimp,J.H., Alpsy,A., Thorpe,A.L., Bavari,A.E., Crooks,D.R.,

- et al. (2019) A chemoproteomic portrait of the oncometabolite fumarate. *Nat. Chem. Biol.*, **15**, 391–400.
2. Chen,Y., Craven,G.B., Kamber,R.A., Cuesta,A., Zheresh,S., Moroz,Y.S., Bassik,M.C. and Taunton,J. (2023) Direct mapping of ligandable tyrosines and lysines in cells with chiral sulfonyl fluoride probes. *Nat. Chem.*, **15**, 1616–1625.
3. Xue,W., Fu,T., Deng,S., Yang,F., Yang,J. and Zhu,F. (2022) Molecular mechanism for the allosteric inhibition of the human serotonin transporter by antidepressant escitalopram. *ACS Chem. Neurosci.*, **13**, 340–351.
4. Chen,X., Wang,Y., Ma,N., Tian,J., Shao,Y., Zhu,B., Wong,Y.K., Liang,Z., Zou,C. and Wang,J. (2020) Target identification of natural medicine with chemical proteomics approach: probe synthesis, target fishing and protein identification. *Signal Transduct. Target Ther.*, **5**, 72.
5. Bakker,A.T., Kotsogianni,I., Mirenda,L., Straub,V.M., Avalos,M., van den Berg,R., Florea,B.I., van Wezel,G.P., Janssen,A.P.A., Martin,N.I., et al. (2023) Chemical proteomics reveals antibiotic targets of oxadiazolones in MRSA. *J. Am. Chem. Soc.*, **145**, 1136–1143.
6. Zeng,X., Yang,X., Fan,J., Tan,Y., Ju,L., Shen,W., Wang,Y., Wang,X., Chen,W., Ju,D., et al. (2021) MASI: microbiota-active substance interactions database. *Nucleic Acids Res.*, **49**, D776–D782.
7. Nguyen,D.D., Xiao,T., Wang,M. and Wei,G.W. (2017) Rigidity strengthening: a mechanism for protein-ligand binding. *J. Chem. Inf. Model.*, **57**, 1715–1721.
8. Mou,M., Pan,Z., Zhou,Z., Zheng,L., Zhang,H., Shi,S., Li,F., Sun,X. and Zhu,F. (2023) A transformer-based ensemble framework for the prediction of protein-protein interaction sites. *Research (Wash D C)*, **6**, 0240.
9. Boike,L., Henning,N.J. and Nomura,D.K. (2022) Advances in covalent drug discovery. *Nat. Rev. Drug Discov.*, **21**, 881–898.
10. Castellón,J.O., Ofori,S., Burton,N.R., Julio,A.R., Turmon,A.C., Armenta,E., Sandoval,C., Boatner,L.M., Takayoshi,E.E., Faragalla,M., et al. (2024) Chemoproteomics identifies state-dependent and proteoform-selective caspase-2 inhibitors. *J. Am. Chem. Soc.*, **146**, 14972–14988.
11. Fu,J., Zhang,Y., Wang,Y., Zhang,H., Liu,J., Tang,J., Yang,Q., Sun,H., Qiu,W., Ma,Y., et al. (2022) Optimization of metabolomic data processing using NOREVA. *Nat. Protoc.*, **17**, 129–151.
12. Yang,Q., Li,B., Chen,S., Tang,J., Li,Y., Li,Y., Zhang,S., Shi,C., Zhang,Y., Mou,M., et al. (2021) MMEASE: online meta-analysis of metabolomic data by enhanced metabolite annotation, marker selection and enrichment analysis. *J. Proteomics*, **232**, 104023.
13. Kavanagh,M.E., Horning,B.D., Khattry,R., Roy,N., Lu,J.P., Whitby,L.R., Ye,E., Brannon,J.C., Parker,A., Chick,J.M., et al. (2022) Selective inhibitors of JAK1 targeting an isoform-restricted allosteric cysteine. *Nat. Chem. Biol.*, **18**, 1388–1398.
14. Donovan,K.A., Ferguson,F.M., Bushman,J.W., Eleuteri,N.A., Bhunia,D., Ryu,S., Tan,L., Shi,K., Yue,H., Liu,X., et al. (2020) Mapping the degradable kinase provides a resource for expedited degrader development. *Cell*, **183**, 1714–1731.
15. Xue,W., Yang,F., Wang,P., Zheng,G., Chen,Y., Yao,X. and Zhu,F. (2018) What contributes to serotonin-norepinephrine reuptake inhibitors’ dual-targeting mechanism? The key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation. *ACS Chem. Neurosci.*, **9**, 1128–1140.
16. Zhang,Y., Zhou,Y., Zhou,Y., Yu,X., Shen,X., Hong,Y., Zhang,Y., Wang,S., Mou,M., Zhang,J., et al. (2024) TheMarker: a comprehensive database of therapeutic biomarkers. *Nucleic Acids Res.*, **52**, D1450–D1464.
17. He,J.X., Fei,Z.C., Fu,L., Tian,C.P., He,F.C., Chi,H. and Yang,J. (2022) A modification-centric assessment tool for the performance of chemoproteomic probes. *Nat. Chem. Biol.*, **18**, 904–912.

18. Nguyen,D.D. and Wei,G.W. (2019) AGL-score: algebraic graph learning score for protein-ligand binding scoring, ranking, docking, and screening. *J. Chem. Inf. Model.*, **59**, 3291–3304.
19. Kawamata,Y., Ryu,K.A., Hermann,G.N., Sandahl,A., Vantourout,J.C., Olow,A.K., Adams,L.A., Rivera-Chao,E., Roberts,L.R., Gnam,S., *et al.* (2023) An electroaffinity labelling platform for chemoproteomic-based target identification. *Nat. Chem.*, **15**, 1267–1275.
20. Offensperger,F., Tin,G., Duran-Frigola,M., Hahn,E., Dobner,S., Ende,C.W.A., Strohbach,J.W., Rukavina,A., Brennstetter,V., Ogilvie,K., *et al.* (2024) Large-scale chemoproteomics expedites ligand discovery and predicts ligand behavior in cells. *Science*, **384**, eadk5864.
21. Wang,S., Ju,D. and Zeng,X. (2024) Mechanisms and clinical implications of human gut microbiota-drug interactions in the precision medicine era. *Biomedicines*, **12**, 194.
22. Zeng,X., Zhao,H., Li,Y., Fan,J., Sun,Y., Wang,S., Wang,Z., Song,P. and Ju,D. (2015) Targeting hedgehog signaling pathway and autophagy overcomes drug resistance of BCR-ABL-positive chronic myeloid leukemia. *Autophagy*, **11**, 355–372.
23. Lu,M.K., Yin,J.Y., Zhu,Q., Lin,G.L., Mou,M.J., Liu,F.Y., Pan,Z.Q., You,N.X., Lian,X.C., Li,F.C., *et al.* (2023) Artificial intelligence in pharmaceutical sciences. *Engineering*, **27**, 37–69.
24. Zheng,L., Shi,S., Lu,M., Fang,P., Pan,Z., Zhang,H., Zhou,Z., Zhang,H., Mou,M., Huang,S., *et al.* (2024) AnnoPRO: a strategy for protein function annotation based on multi-scale protein representation and a hybrid deep learning of dual-path encoding. *Genome Biol.*, **25**, 41.
25. Sun,X., Zhang,Y., Li,H., Zhou,Y., Shi,S., Chen,Z., He,X., Zhang,H., Li,F., Yin,J., *et al.* (2023) DRESIS: the first comprehensive landscape of drug resistance information. *Nucleic Acids Res.*, **51**, D1263–D1275.
26. Li,F., Yin,J., Lu,M., Yang,Q., Zeng,Z., Zhang,B., Li,Z., Qiu,Y., Dai,H., Chen,Y., *et al.* (2022) ConSIG: consistent discovery of molecular signature fromOMIC data. *Brief. Bioinform.*, **23**, bbac253.
27. Yang,Q., Li,B., Tang,J., Cui,X., Wang,Y., Li,X., Hu,J., Chen,Y., Xue,W., Lou,Y., *et al.* (2020) Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. *Brief. Bioinform.*, **21**, 1058–1068.
28. Gilson,M.K., Liu,T., Baitaluk,M., Nicola,G., Hwang,L. and Chong,J. (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.*, **44**, D1045–D1053.
29. Mendez,D., Gaulton,A., Bento,A.P., Chambers,J., De Veij,M., Félix,E., Magarinos,M.P., Mosquera,J.F., Mutowo,P., Nowotka,M., *et al.* (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.*, **47**, D930–D940.
30. Gao,M., Moumbock,A.F.A., Qaseem,A., Xu,Q. and Günther,S. (2022) CovPDB: a high-resolution coverage of the covalent protein-ligand interactome. *Nucleic Acids Res.*, **50**, D445–D450.
31. Harding,S.D., Armstrong,J.F., Faccenda,E., Southan,C., Alexander,S.P.H., Davenport,A.P., Spedding,M. and Davies,J.A. (2024) The IUPHAR/BPS guide to PHARMACOLOGY in 2024. *Nucleic Acids Res.*, **52**, D1438–D1449.
32. Kooistra,A.J., Kanev,G.K., van Linden,O.P., Leurs,R., de Esch,I.J. and de Graaf,C. (2016) KLIFS: a structural kinase-ligand interaction database. *Nucleic Acids Res.*, **44**, D365–D371.
33. Antolin,A.A., Sanfelice,D., Crisp,A., Villasclaras Fernandez,E., Mica,I.L., Chen,Y., Collins,I., Edwards,A., Müller,S., Al-Lazikani,B., *et al.* (2023) The Chemical Probes Portal: an expert review-based public resource to empower chemical probe assessment, selection and use. *Nucleic Acids Res.*, **51**, D1492–D1502.
34. Du,H., Gao,J., Weng,G., Ding,J., Chai,X., Pang,J., Kang,Y., Li,D., Cao,D. and Hou,T. (2021) CovalentInDB: a comprehensive database facilitating the discovery of covalent inhibitors. *Nucleic Acids Res.*, **49**, D1122–D1129.
35. Boatner,L.M., Palafox,M.F., Schweppe,D.K. and Backus,K.M. (2023) CysDB: a human cysteine database based on experimental quantitative chemoproteomics. *Cell Chem. Biol.*, **30**, 683–698.
36. Ma,N., Hu,J., Zhang,Z.M., Liu,W., Huang,M., Fan,Y., Yin,X., Wang,J., Ding,K., Ye,W., *et al.* (2020) 2H-azirine-based reagents for chemoselective bioconjugation at carboxyl residues inside live cells. *J. Am. Chem. Soc.*, **142**, 6051–6059.
37. Lv,S., Xu,F., Fan,Y., Ding,K. and Li,Z. (2023) Cyclopropenone, cyclopropeniminium ion, and cyclopropenethione as novel electrophilic warheads for potential target discovery of triple-negative breast cancer. *J. Med. Chem.*, **66**, 2851–2864.
38. Li,S., Zhang,P., Xu,F., Hu,S., Liu,J., Tan,Y., Tu,Z., Sun,H., Zhang,Z.M., He,Q.Y., *et al.* (2022) Ynamide electrophile for the profiling of ligandable carboxyl residues in live cells and the development of new covalent inhibitors. *J. Med. Chem.*, **65**, 10408–10418.
39. Yang,T., Cuesta,A., Wan,X., Craven,G.B., Hirakawa,B., Khamphavong,P., May,J.R., Kath,J.C., Lapek,J.D. Jr., Niessen,S. *et al.* (2022) Reversible lysine-targeted probes reveal residence time-based kinase selectivity. *Nat. Chem. Biol.*, **18**, 934–941.
40. Hahn,H.S., Toroitich,E.K., Borne,A.L., Brulet,J.W., Libby,A.H., Yuan,K., Ware,T.B., McCloud,R.L., Ciancone,A.M. and Hsu,K.L. (2020) Global targeting of functional tyrosines using sulfur-triazole exchange chemistry. *Nat. Chem. Biol.*, **16**, 150–159.
41. Li,F., Zhou,Y., Zhang,Y., Yin,J., Qiu,Y., Gao,J. and Zhu,F. (2022) POSREG: proteomic signature discovered by simultaneously optimizing its reproducibility and generalizability. *Brief. Bioinform.*, **23**, bbac040.
42. Li,Y.H., Li,X.X., Hong,J.J., Wang,Y.X., Fu,J.B., Yang,H., Yu,C.Y., Li,F.C., Hu,J., Xue,W.W., *et al.* (2020) Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs. *Brief. Bioinform.*, **21**, 649–662.
43. Li,T., Wang,A., Zhang,Y., Chen,W., Guo,Y., Yuan,X., Liu,Y. and Geng,Y. (2024) Chemoproteomic profiling of signaling metabolite fructose-1,6-bisphosphate interacting proteins in living cells. *J. Am. Chem. Soc.*, **146**, 15155–15166.
44. Martín-Acosta,P., Meng,Q., Klimek,J., Reddy,A.P., David,L., Petrie,S.K., Li,B.X. and Xiao,X. (2022) A clickable photoaffinity probe of betulinic acid identifies tropomyosin as a target. *Acta Pharm. Sin. B*, **12**, 2406–2416.
45. Yang,F., Jia,G., Guo,J., Liu,Y. and Wang,C. (2022) Quantitative chemoproteomic profiling with data-independent acquisition-based mass spectrometry. *J. Am. Chem. Soc.*, **144**, 901–911.
46. Spradlin,J.N., Zhang,E. and Nomura,D.K. (2021) Reimagining druggability using chemoproteomic platforms. *Acc. Chem. Res.*, **54**, 1801–1813.
47. Lian,X., Zhang,Y., Zhou,Y., Sun,X., Huang,S., Dai,H., Han,L. and Zhu,F. (2024) SingPro: a knowledge base providing single-cell proteomic data. *Nucleic Acids Res.*, **52**, D552–D561.
48. Zhang,Y., Sun,H., Lian,X., Tang,J. and Zhu,F. (2023) ANPELA: significantly enhanced quantification tool for cytometry-based single-cell proteomics. *Adv. Sci. (Weinh)*, **10**, e2207061.
49. UniProt, C. (2023) UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
50. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B., *et al.* (2021) PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.*, **49**, D1388–D1395.
51. Zhou,Y., Zhang,Y., Zhao,D., Yu,X., Shen,X., Zhou,Y., Wang,S., Qiu,Y., Chen,Y. and Zhu,F. (2024) TTD: therapeutic target database describing target druggability information. *Nucleic Acids Res.*, **52**, D1465–D1477.
52. Li,F., Yin,J., Lu,M., Mou,M., Li,Z., Zeng,Z., Tan,Y., Wang,S., Chu,X., Dai,H., *et al.* (2023) DrugMAP: molecular atlas and pharma-information of all drugs. *Nucleic Acids Res.*, **51**, D1288–D1299.

53. Perez-Riverol, Y., Bai, J.W., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A. and Eisenacher, M. (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.*, **50**, D543–D552.
54. McGregor, N.G.S., de Boer, C., Foucart, Q.P.O., Beenakker, T., Offen, W.A., Codee, J.D.C., Willems, L.I., Overkleeft, H.S. and Davies, G.J. (2023) A multiplexing activity-based protein-profiling platform for dissection of a native bacterial xyloglucan-degrading system. *ACS Cent. Sci.*, **9**, 2306–2314.
55. Wang, W.J., Ling, Y.Y., Shi, Y., Wu, X.W., Su, X., Li, Z.Q., Mao, Z.W. and Tan, C.P. (2024) Identification of mitochondrial ATP synthase as the cellular target of Ru-polypyridyl-beta-carboline complexes by affinity-based protein profiling. *Natl. Sci. Rev.*, **11**, nwae234.
56. Lin, Z., Wang, X., Bustin, K.A., Shishikura, K., McKnight, N.R., He, L., Suci, R.M., Hu, K., Han, X., Ahmadi, M., et al. (2021) Activity-based hydrazine probes for protein profiling of electrophilic functionality in therapeutic targets. *ACS Cent. Sci.*, **7**, 1524–1534.
57. Liu, Y., Patricelli, M.P. and Cravatt, B.F. (1999) Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 14694–14699.
58. Greenbaum, D., Medzihradsky, K.F., Burlingame, A. and Bogoy, M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.*, **7**, 569–581.
59. Bustin, K.A., Abbas, A., Wang, X., Abt, M.C., Zackular, J.P. and Matthews, M.L. (2023) Characterizing metabolic drivers of *Clostridioides difficile* infection with activity-based hydrazine probes. *Front. Pharmacol.*, **14**, 1074619.
60. Zhang, Y., Liu, X., Li, F., Yin, J., Yang, H., Li, X., Liu, X., Chai, X., Niu, T., Zeng, S., et al. (2024) INTEDE 2.0: the metabolic roadmap of drugs. *Nucleic Acids Res.*, **52**, D1355–D1364.
61. Yang, Q., Wang, Y., Zhang, Y., Li, F., Xia, W., Zhou, Y., Qiu, Y., Li, H. and Zhu, F. (2020) NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. *Nucleic Acids Res.*, **48**, W436–W448.
62. Tang, J., Fu, J., Wang, Y., Li, B., Li, Y., Yang, Q., Cui, X., Hong, J., Li, X., Chen, Y., et al. (2020) ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief. Bioinform.*, **21**, 621–636.
63. Yin, J., Chen, Z., You, N., Li, F., Zhang, H., Xue, J., Ma, H., Zhao, Q., Yu, L., Zeng, S., et al. (2024) VARIDT 3.0: the phenotypic and regulatory variability of drug transporter. *Nucleic Acids Res.*, **52**, D1490–D1502.
64. Dettling, S.E., Ahmadi, M., Lin, Z., He, L. and Matthews, M.L. (2020) Discovery of electrophiles and profiling of enzyme cofactors. *Curr. Protoc. Chem. Biol.*, **12**, e86.
65. Cravatt, B.F., Wright, A.T. and Kozarich, J.W. (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.*, **77**, 383–414.
66. Nomura, D.K., Dix, M.M. and Cravatt, B.F. (2010) Activity-based protein profiling for biochemical pathway discovery in cancer. *Nat. Rev. Cancer*, **10**, 630–638.
67. Yin, J., Li, F., Zhou, Y., Mou, M., Lu, Y., Chen, K., Xue, J., Luo, Y., Fu, J., He, X., et al. (2021) INTEDE: interactome of drug-metabolizing enzymes. *Nucleic Acids Res.*, **49**, D1233–D1243.
68. Cravatt, B.F. (2023) Activity-based protein profiling - finding general solutions to specific problems. *Isr. J. Chem.*, **63**, e202300029.
69. Weerapana, E., Wang, C., Simon, G.M., Richter, F., Khare, S., Dillon, M.B., Bachovchin, D.A., Mowen, K., Baker, D. and Cravatt, B.F. (2010) Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature*, **468**, 790–795.
70. Xiao, W.D., Chen, Y. and Wang, C. (2023) Quantitative chemoproteomic methods for reactive cysteinome profiling. *Isr. J. Chem.*, **63**, e202200100.
71. Ciancone, A.M., Hosseinibarkoobe, S., Bai, D.L., Borne, A.L., Ferris, H.A. and Hsu, K.L. (2022) Global profiling identifies a stress-responsive tyrosine site on EDC3 regulating biomolecular condensate formation. *Cell Chem. Biol.*, **29**, 1709–1720.
72. Brulet, J.W., Borne, A.L., Yuan, K., Libby, A.H. and Hsu, K.L. (2020) Liganding functional tyrosine sites on proteins using sulfur-triazole exchange chemistry. *J. Am. Chem. Soc.*, **142**, 8270–8280.
73. Gehringer, M. and Laufer, S.A. (2019) Emerging and re-emerging warheads for targeted covalent inhibitors: applications in medicinal chemistry and chemical biology. *J. Med. Chem.*, **62**, 5673–5724.
74. Hillebrand, L., Liang, X.J.J., Serafim, R.A.M. and Gehringer, M. (2024) Emerging and re-emerging warheads for targeted covalent inhibitors: an update. *J. Med. Chem.*, **67**, 7668–7758.
75. Fu, L., Li, Z., Liu, K., Tian, C., He, J., He, J., He, F., Xu, P. and Yang, J. (2020) A quantitative thiol reactivity profiling platform to analyze redox and electrophile reactive cysteine proteomes. *Nat. Protoc.*, **15**, 2891–2919.
76. Kuljanin, M., Mitchell, D.C., Schweppe, D.K., Gikandi, A.S., Nusinow, D.P., Bulloch, N.J., Vinogradova, E.V., Wilson, D.L., Kool, E.T., Mancias, J.D., et al. (2021) Reimagining high-throughput profiling of reactive cysteines for cell-based screening of large electrophile libraries. *Nat. Biotechnol.*, **39**, 630–641.
77. Yang, K., Whitehouse, R.L., Dawson, S.L., Zhang, L., Martin, J.G., Johnson, D.S., Paulo, J.A., Gygi, S.P. and Yu, Q. (2023) Accelerating multiplexed profiling of protein-ligand interactions: high-throughput plate-based reactive cysteine profiling with minimal input. *Cell Chem. Biol.*, **31**, 565–576.
78. Koo, T.Y., Lai, H., Nomura, D.K. and Chung, C.Y. (2023) N-acryloylindole-alkyne (NAIA) enables imaging and profiling new ligandable cysteines and oxidized thiols by chemoproteomics. *Nat. Commun.*, **14**, 3564.
79. Hsieh, I.S., Gopula, B., Chou, C.C., Wu, H.Y., Chang, G.D., Wu, W.J., Chang, C.S., Chu, P.C. and Chen, C.S. (2019) Development of novel irreversible pyruvate kinase M2 inhibitors. *J. Med. Chem.*, **62**, 8497–8510.
80. Stieger, C.E., Park, Y., Geus, M.A.R., Kim, D.J., Huhn, C., Slenczka, J.S., Ochtrup, P., Mächler, J.M., Süßmuth, R.D., Broichhagen, J., et al. (2022) DFT-guided discovery of ethynyl-triazolyl-phosphinates as modular electrophiles for chemoselective cysteine bioconjugation and profiling. *Angew. Chem. Int. Ed Engl.*, **61**, e202205348.
81. Abegg, D., Tomanik, M., Qiu, N., Pechalrieu, D., Shuster, A., Commare, B., Togni, A., Herzon, S.B. and Adibekian, A. (2021) Chemoproteomic profiling by cysteine fluoroalkylation reveals myrocin G as an inhibitor of the nonhomologous end joining DNA repair pathway. *J. Am. Chem. Soc.*, **143**, 20332–20342.
82. Abegg, D., Frei, R., Cerato, L., Prasad Hari, D., Wang, C., Waser, J. and Adibekian, A. (2015) Proteome-wide profiling of targets of cysteine reactive small molecules by using ethynyl benziodoxolone reagents. *Angew. Chem. Int. Ed Engl.*, **54**, 10852–10857.
83. Weigert, M.A., Meighen-Berger, K.M., Hacker, S.M., Feige, M.J. and Sieber, S.A. (2023) A chemical probe unravels the reactive proteome of health-associated catechols. *Chem. Sci.*, **14**, 8635–8643.
84. Hurben, A.K., Erber, L.N., Tretyakova, N.Y. and Doran, T.M. (2021) Proteome-wide profiling of cellular targets modified by dopamine metabolites using a bio-orthogonally functionalized catecholamine. *ACS Chem. Biol.*, **16**, 2581–2594.
85. Wang, R., Yang, D.Y., Tian, T., An, Y.H., Wan, C., Chang, Q., Liang, M.C., Hou, Z.F., Wang, Y., Zhang, L., et al. (2022) Low-toxicity sulfonium-based probes for cysteine-specific profiling in live cells. *Anal. Chem.*, **94**, 4366–4372.
86. Liu, Y., Liu, J., Zhang, X., Guo, C., Xing, X., Zhang, Z.M., Ding, K. and Li, Z. (2023) Oxidant-induced bioconjugation for protein labeling in live cells. *ACS Chem. Biol.*, **18**, 112–122.

87. Shi, Y.L., Fu, L., Yang, J. and Carroll, K.S. (2021) Wittig reagents for chemoselective sulfenic acid ligation enables global site stoichiometry analysis and redox-controlled mitochondrial targeting. *Nat. Chem.*, **13**, 1140–1150.
88. Fu, L., Jung, Y.G., Tian, C.P., Ferreira, R.N., Cheng, R.F., He, F.C., Yang, J. and Carroll, K.S. (2023) Nucleophilic covalent ligand discovery for the cysteine redoxome. *Nat. Chem. Biol.*, **19**, 1309–1319.
89. Paulsen, C.E., Truong, T.H., Garcia, F.J., Homann, A., Gupta, V., Leonard, S.E. and Carroll, K.S. (2011) Peroxide-dependent sulfenylation of the EGFR catalytic site enhances kinase activity. *Nat. Chem. Biol.*, **8**, 57–64.
90. Homan, R.A., Lapek, J.D., Woo, C.M., Niessen, S., Jones, L.H. and Parker, C.G. (2024) Photoaffinity labelling with small molecules. *Nat. Rev. Methods Primers*, **4**, 30.
91. Smith, E. and Collins, I. (2015) Photoaffinity labeling in target-and binding-site identification. *Future Med. Chem.*, **7**, 159–183.
92. Pan, S., Zhang, H., Wang, C., Yao, S.C. and Yao, S.Q. (2016) Target identification of natural products and bioactive compounds using affinity-based probes. *Nat. Prod. Rep.*, **33**, 612–620.
93. Halloran, M.W. and Lumb, J.P. (2019) Recent applications of diazirines in chemical proteomics. *Chem.*, **25**, 4885–4898.
94. Gao, K., Nguyen, D.D., Tu, M. and Wei, G.W. (2020) Generative network complex for the automated generation of drug-like molecules. *J. Chem. Inf. Model.*, **60**, 5682–5698.
95. Parker, C.G., Galmozzi, A., Wang, Y., Correia, B.E., Sasaki, K., Joslyn, C.M., Kim, A.S., Cavallaro, C.L., Lawrence, R.M., Johnson, S.R., *et al.* (2017) Ligand and target discovery by fragment-based screening in human cells. *Cell*, **168**, 527–541.
96. Wang, Y., Dix, M.M., Bianco, G., Remsberg, J.R., Lee, H.Y., Kalocsay, M., Gygi, S.P., Forli, S., Vite, G., Lawrence, R.M., *et al.* (2019) Expedited mapping of the ligandable proteome using fully functionalized enantiomeric probe pairs. *Nat. Chem.*, **11**, 1113–1123.
97. Zheng, L., Shi, S., Sun, X., Lu, M., Liao, Y., Zhu, S., Zhang, H., Pan, Z., Fang, P., Zeng, Z., *et al.* (2024) MoDAFold: a strategy for predicting the structure of missense mutant protein based on AlphaFold2 and molecular dynamics. *Brief. Bioinform.*, **25**, bbae006.
98. Conway, L.P., Jadhav, A.M., Homan, R.A., Li, W., Rubiano, J.S., Hawkins, R., Lawrence, R.M. and Parker, C.G. (2021) Evaluation of fully-functionalized diazirine tags for chemical proteomic applications. *Chem. Sci.*, **12**, 7839–7847.
99. Tan, Y., Fu, S., Yang, T., Xie, Y., Shen, G., Yan, J., Zhao, Y. and Ni, F. (2022) Evaluation of site-diversified, fully functionalized diazirine probes for chemical proteomic applications. *Chem. Commun.*, **58**, 9532–9535.
100. Bemis, G.W. and Murcko, M.A. (1996) The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.*, **39**, 2887–2893.
101. Bajusz, D., Racz, A. and Heberger, K. (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.*, **7**, 20.
102. Dunkel, M., Fullbeck, M., Neumann, S. and Preissner, R. (2006) SuperNatural: a searchable database of available natural compounds. *Nucleic Acids Res.*, **34**, D678–D683.
103. Zhou, Y., Zhang, Y.T., Lian, X.C., Li, F.C., Wang, C.X., Zhu, F., Qiu, Y.Q. and Chen, Y.Z. (2022) Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res.*, **50**, D1398–D1407.
104. Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezuk, Y., *et al.* (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.*, **41**, W29–W33.