Cellular protein painting for structural and binding sites analysis via lysine reactivity profiling with o-phthalaldehyde

Zhenxiang Zheng[1], Ya Zeng[1], Kunjia Lai[1], Bin Liao[1], Pengfei Li[1], Chris Soon Heng Tan[1*]

1. Department of chemistry, College of science, Southern University of Science and Technology, Shenzhen, Guangdong, 518055, China PR.
*Correspondence: christan@sustech.edu.cn

Abstract

The three-dimensional structure and the molecular interaction of proteins determine their roles in many cellular processes. Chemical protein painting with protein mass spectrometry can identify ligand binding sites, changes in protein structural conformations and protein-protein interactions. Nevertheless, most current protein painting techniques identified protein targets and binding sites of drugs *in vitro* using cell lysate or purified protein. Here, we screened 11 kinds of membrane-permeable lysine-reactive chemical probes for intracellular covalent labeling of endogenous proteins, which reveals *ortho*-phthalaldehyde (OPA) as the most reactive probe in intracellular environment. An MS workflow was developed and coupled with a new data analysis strategy termed RAPID (Reactive Amino acid Profiling by Inverse Detection) to enhance detection sensitivity. RAPID-OPA successfully identified structural change induced by allosteric drug TEPP-46 on its target protein PKM2, which was applied to profile conformation change of the proteome occurring in cells during thermal denaturation. Application of RAPID-OPA on cells treated with geldanamycin successfully revealed its binding sites on target proteins. Thus, RAPID-OPA for cellular protein painting permits the identification of ligand-binding sites and detection of protein structural changes occurring in cells.

Introduction

Proteins perform cellular functions through their three-dimensional structures, local conformational features (e.g. catalytic sites), and the molecular interactions in which they are involved. Investigating protein structural conformations, protein-ligand interaction, and protein-protein interaction is of great importance in cell biology and drug discovery. X-ray crystallography, nuclear magnetic resonance (NMR), and various spectroscopic techniques are commonly used in the field of structural biology to study the conformational change of proteins *in vitro*. Nevertheless, observations made *in vitro* studies do not necessarily translate to actual changes that occurred *in vivo*. In addition, identical drugs may bind to different targets or to the same target with different affinities across different cell lines.[1] Thus, there is a crucial need to study protein structures and interactions occurring in cells.

Recently, with the development of liquid-phase mass spectrometry (LC/MS), some MS-proteomics methods have been applied to study dynamic change of protein structures in cells, including the Hydrogen-Deuterium Exchange (HDX),[2] Fast Photochemical Oxidation of Proteins (FPOP), [3-5] Covalent Protein Painting, [6-8] Chemical Cross-Linking, [9-12] etc. In general, Chemical Cross-Linking suffers from low reaction stoichiometry that is detectable only on the most abundant proteins in the proteome,

restricting the methodology for comprehensive profiling of protein structure and protein interactions. HDX can be used to resolve solvent accessibility and weak hydrogen bonding, but depends on reversible covalent labeling that could compromise the detection of structural changes.[13] The labeling of proteins by FPOP is non-specific for global proteome study. However, the disadvantages of this method are the increase of hydroxyl free radicals in the sample, which are highly reactive but less specific[14] and cannot enrichment by chemical probe. Covalent Protein Painting (CPP), where solvent-accessible amino acids on proteins are covalently labeled with chemicals, is a promising chemical proteomics strategy for intracellular studies of protein structures and conformational changes.[7] Chemically modified regions or specific amino acid sites constitute a "footprint" that is affected by protein structure as well as interaction with protein, metabolites and xenobiotics such as drugs. Compared to other methods, CPP with protein mass spectrometry does not require isolation and purification of proteins, and can detect multiple proteins and thousands of peptides in parallel.

The amino acids that are often labeled in common chemical proteomics strategies are cysteine and lysine, of which lysine accounts for 6% of all residues in human proteins and is one of the most represented amino acids with an average of 30 lysine sites per protein.[15] Lysine is present in many functional sites such as enzyme active sites,[16] protein-protein interaction interfaces,[17] and post-translational modification sites. Currently, the commonly used transmembrane biomolecular probes for labeling and cross-linking primary amines are *N*-hydroxysuccinimidyl ester (NHS ester) but suffer from poor selectivity (also react with histidine, serine, and tyrosine), poor stability and low labeling efficiency under physiological conditions.[18] In contrast, there are also in vitro studies showing that *ortho*-phthalaldehyde (OPA) is capable of amine reactions with peptides and proteins under physiological conditions and is better than NHS esters for protein labeling.[19] Moreover, OPA is also used to modify proteins and peptides by intramolecular OPA-amine-thiol three-component processes for cyclization[20, 21] and OPA-amine two-component reactions for bioconjugation.[22, 23] In microbiology, OPA is used for sterilization as it binds to membrane receptors and impair their functions.[24]

Here, we screened 11 membrane-permeable chemical probes revealing that OPA exhibited the highest reactivity intracellularly and was used to develop an intracellular covalent labeling method with new data analysis strategy termed RAPID (Reactive Amino acid Profiling with Inverse Detection) to identify intracellular changes in structures and molecular interactions of proteins. We applied RAPID-OPA to investigate conformational changes of protein during thermal denaturation as well as conformational change of protein PKM2 induced by allosteric inhibitors TEPP-46. In addition, RAPID-OPA was also applied to identify drug binding sites of geldanamycin in cells.

Results and discussion

Profiling of RAPID strategy

Lysine-reactive probes could react with lysine on proteins at low stoichiometry that often necessitate enrichment steps to increase the identification of peptides containing probe-modified lysine while fragmentation during MS analysis could also lead to cleavage of probe generating daughter ions with unpredictable mass-to-charge ratio (m/z).[25] Here, we hypothesized that these challenges can be circumvented by the identifying lysine probes, preferably membrane-permeable, that

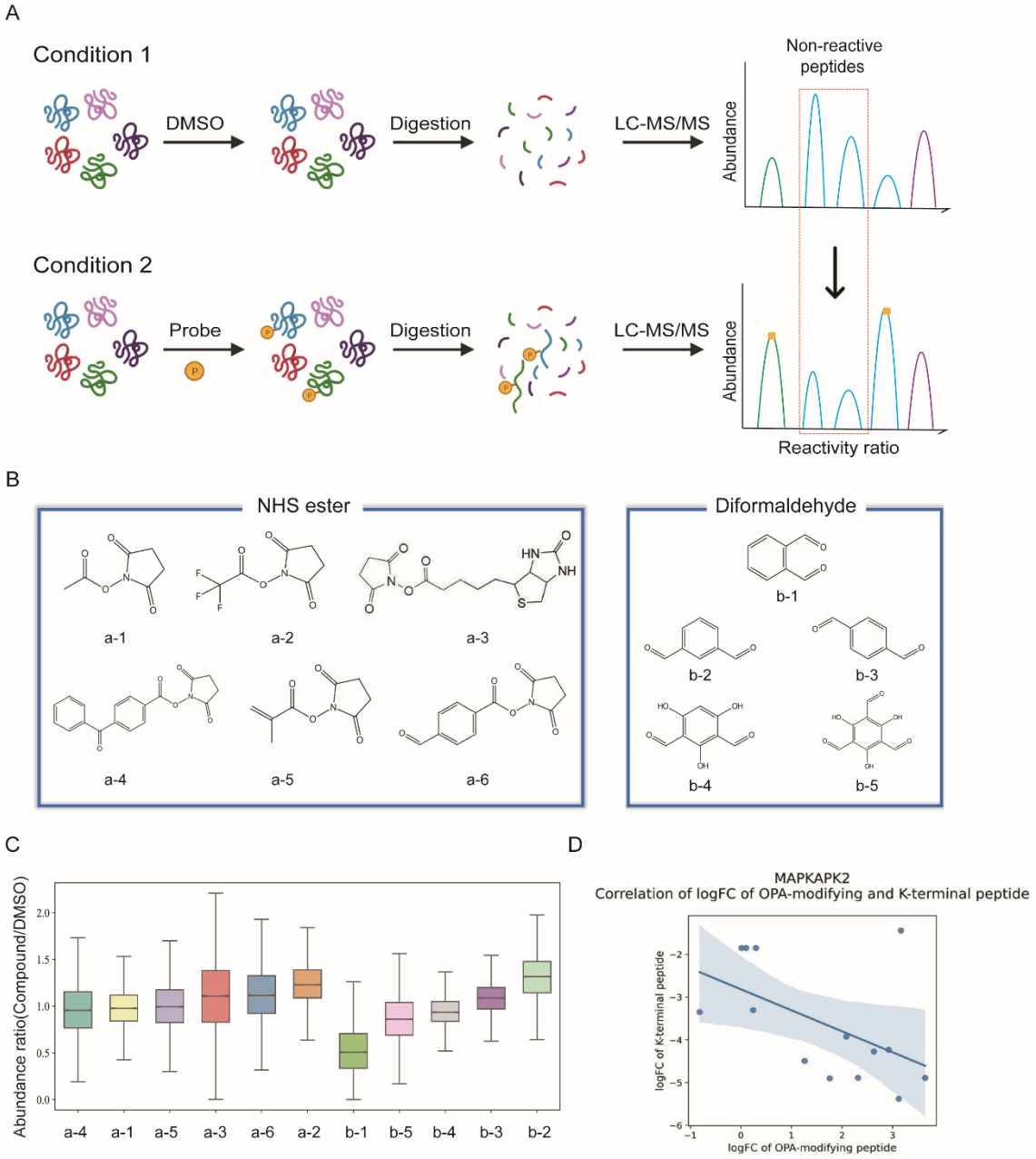

Fig 1. Proposal of the entire workflow for lysine reactivity profiling in complicated proteomes. (A) Principle of RAPID strategy. (B) Chemical structure of lysine reactivity probes grouped into NHS esters and diformaldehydes. (C) Boxplot visualization of abundance changes for peptides in HEK293T cell induced by each lysine reactivity probe. (D) Scatter plot shows the correlation in abundance changes of OPA-labeling peptides and their native peptides in MAPKAPK2 protein.

interfere with cleavage of proteins at C-terminal of arginine or lysine by trypsin. The protease is typically used to cleave proteins into multiple peptides for bottom-up MS analysis, but lysine modified by such probe will not be cleaved by trypsin leading to the decreased abundance of peptides that could have been generated otherwise. In addition, probe-modified lysine may not be amenable to further labelling for multiplex MS analysis such as TMT (Tandem Mass Tag) reagents. Thus, by comparing unmodified peptides from control samples (without probe) to probe-treated samples with an isobaric labelling MS strategy

(e.g. TMT), low abundant probe-modified peptides could be inferred with increased sensitivity while omitting tedious enrichment steps that can increase experimental variability (Fig. 1A). We note that cell intrinsic factors, such as changes in post-translation modification, structural conformation, and interactions of proteins with metabolites, drugs and other proteins, could lead to differentiated changes in lysine reactivity that can be explored with RAPID strategy. Based on above assumption, we first screened 11 chemical probes, consisting of NHS esters and diformaldehydes, incubating them with intact cells followed by cell lysis, tryptic digestion and TMT-labelling for MS analysis. We observed sample treated membrane-permeable OPA (b-1) had the largest decrease in abundance of peptide identified while other compounds have either no effect or an up-regulated effect (Fig. 1B-C). We note that OPA contains two aldehyde groups that can be condensed with an amine by Mannich reaction and/or reductive amination (Fig. S1A).

The observed decrease in native peptide could potentially arise from direct interference of trypsin enzymatic activity by OPA, hence steps are incorporated in our workflow to remove excess OPA prior enzymatic digestion of proteins. [26] Nevertheless, to further validate that the decrease in abundance of native peptides arises from corresponding increase in OPA-labelled peptides, purified MAP kinase-activated protein kinase 2 (MAPKAPK2), which contains 30 lysine residues, was incubated with OPA followed by bottom-up MS analysis where OPA is included as variable lysine modification during MS spectral search. Using purified protein at high concentration increases OPA reaction stoichiometry and the abundance of OPA-labelled peptides for MS detection while reduce interference from other interacting biomolecules such as proteins and metabolites. In total, 82 peptides were quantified of which 19 peptides (23.17%) were OPA-labelled (Fig. S1B). We compared the relative abundance of these OPA-labeled peptides (orange) and their respective unmodified native peptides (green) (Fig. S1C). As anticipated, the native peptides of identified OPA-labeled peptides shown a corresponding decrease in abundance (Fig. 1D and Fig. S1C). These findings strongly suggested that intensity of OPA labelling can be inferred from changes in abundance of native peptides.

Cellular protein labelling of the whole proteome profiling by OPA

We hypothesized that the OPA-labelling reaction follows a second-order kinetic approach,[27] with the modification rate increasing in a dose and time dependent manner. To systematically optimize the labelling conditions, we first incubated HEK293T cells with different OPA concentration from 50 μM to 800 μM for 15 min reaction (Fig. 2A). Overall, we observed peptides with decreased abundance from 100 μM onward which further decrease with higher OPA concentration thereafter (Fig. 2B). Next, we examined how the duration of incubation with 800 μM OPA affects labeling at different time points, namely 5, 10, and 15 min. As anticipated, the percentage of peptide with decreased abundance increases with time where the greatest change was observed with 15 min of OPA incubation (Fig. 2C). The concentration- and time-dependent effect of OPA suggest the decreased peptide abundance observed is largely due to reaction of lysine with OPA and demonstrated that OPA could label proteins progressively under the physiological conditions.

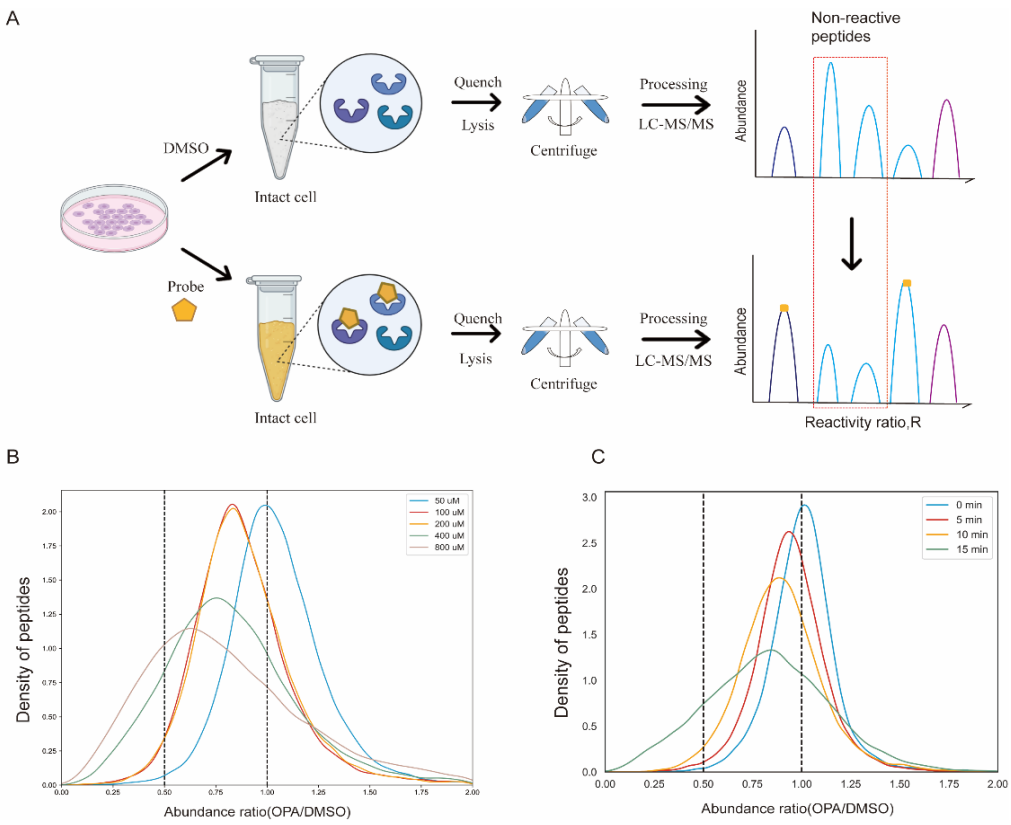

Fig. 2 Experimental process and method optimization. (A) General workflow of RAPID-OPA. (B) Overall density plot of peptide abundance change with different concentrations of OPA. (C) Overall density plot of peptide abundance change with different labeling times of OPA.

RAPID-OPA for monitoring proteome-wide structural changes of proteins in cells

Overall, with 15 min of 800 μM OPA incubation in HEK293T cells, 8487 out of 63,005 native peptides (13.47%) across 7,188 proteins exhibited statistically significant (p-value < 0.05) decrease in abundance. In comparison, spectral search incorporating variable OPA modification could only identified 673 OPA-labeled peptides (1.07%) (Fig. 3A). After identifying OPA and developed RAPID-OPA workflow, we investigate the applicability of the workflow to study proteome-wide intracellular conformational changes of protein structure. First, using monomer structures of human proteins predicted by AlphaFold2, we observed native peptides with largest decrease in abundance after OPA labelling are over-represented in highly surface accessible lysine (Fig 3B). Nevertheless, not all peptides with highly surface accessible lysine exhibit marked decrease in abundance. As monomer structures of proteins were used in calculating solvent accessible surface area (SASA), some of these peptides might be parts of protein involved in interactions with membrane, nucleic acids, and other proteins, which could also influence protein structural conformation not predicted by AlphaFold.[28] We also observed peptides with partially surface accessible lysine could exhibit marked decrease in abundance, possibly due to microenvironment of lysine with proximal amino acids that could influence its reactivity[17]. Thus, we assumed that changes in structural conformation of protein and lysine microenvironment could be reflected as differentiated changes to OPA labelling of lysine. To investigate this, MS data from intact cells treated with OPA at three temperature

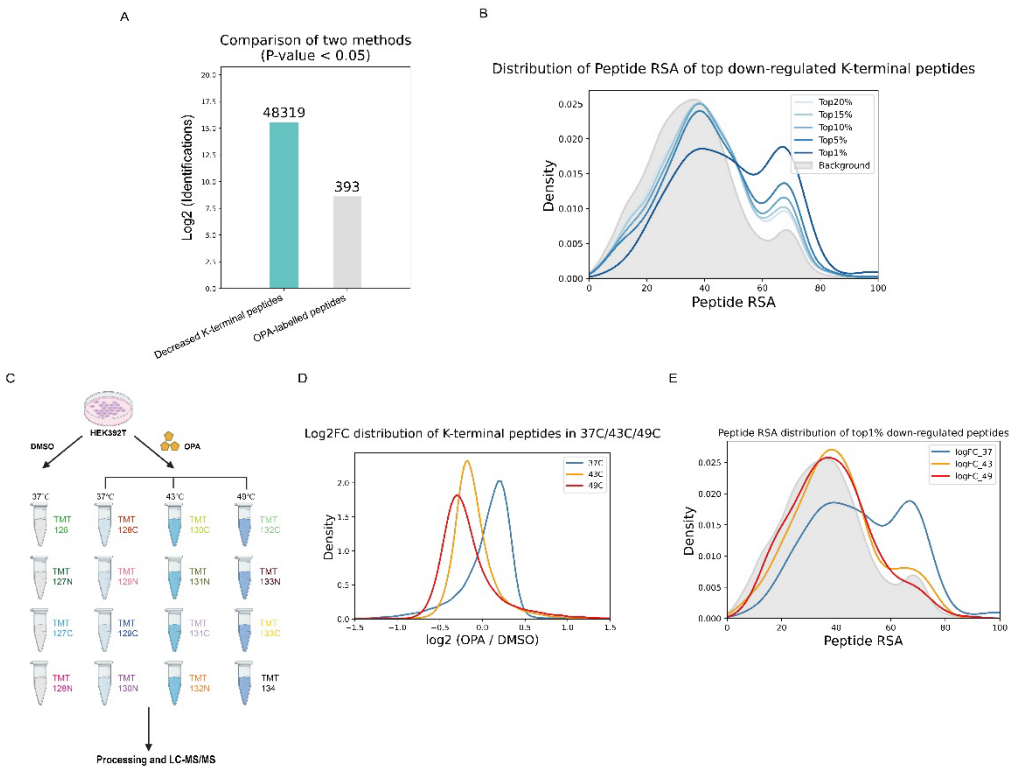

Fig. 3 Monitoring the structural changes in proteomes. (A) Number of peptides with differentiated abundance identified in HEK293T cells. (B) RSA distribution of down-regulated peptides at 37°C. (C) Schema of samples collected. HEK293T cells are treated with DMSO/OPA and heated at 3 temperatures for CPP profiling. (D) Distribution of changes in native peptide abundance across different temperatures. (E) RSA distribution of the top 1% down-regulated peptides.

points (namely 37°C, 43°C and 49°C) were collected and analyzed (Fig. 3C). At these elevated temperatures, proteins are expected to at least partially unfold, exposing inner lysine for OPA labelling which will result in less native peptides for MS detection (Fig. 3D). Indeed, we observed further decrease in native peptide abundance when cells are incubated with OPA at denaturing temperatures (43°C and 49°C) where partially surface accessible lysine is enriched in peptides with largest decrease (Fig. 3D-E and Fig. S2A-B). Overall, more than 23,116 (36.69%) and 30,408 (48.26%) native peptides exhibited a decrease in abundance while only 316 and 171 OPA-labelled peptides were detected at 43°C and 49°C respectively. We reasoned proteins had partially unfolded at 43°C and 49°C with opening pockets that increase accessibility of lysine to OPA. All in all, these results suggest that RAPID-OPA can be used to detect structural changes of proteins in cell.

Identifying protein structure conformation changes with RAPID-OPA

TEPP-46 binds and activate PKM2 (pyruvate kinase M2) by promoting the tetramerization of the protein[29]. To further validate the use of RAPID-OPA to study conformational changes in protein structures, we treated HEK293T cells with TEPP-46 for 20 min, followed by RAPID-OPA where cells are incubated with OPA before lysis and MS analysis using PRM mode (Fig. 4A). Four biological replicates were performed and processed for MS analysis with a single TMT set. Overall, we detected 46 native peptides of PKM2 protein, of which 3 peptides exhibit a statistical significance

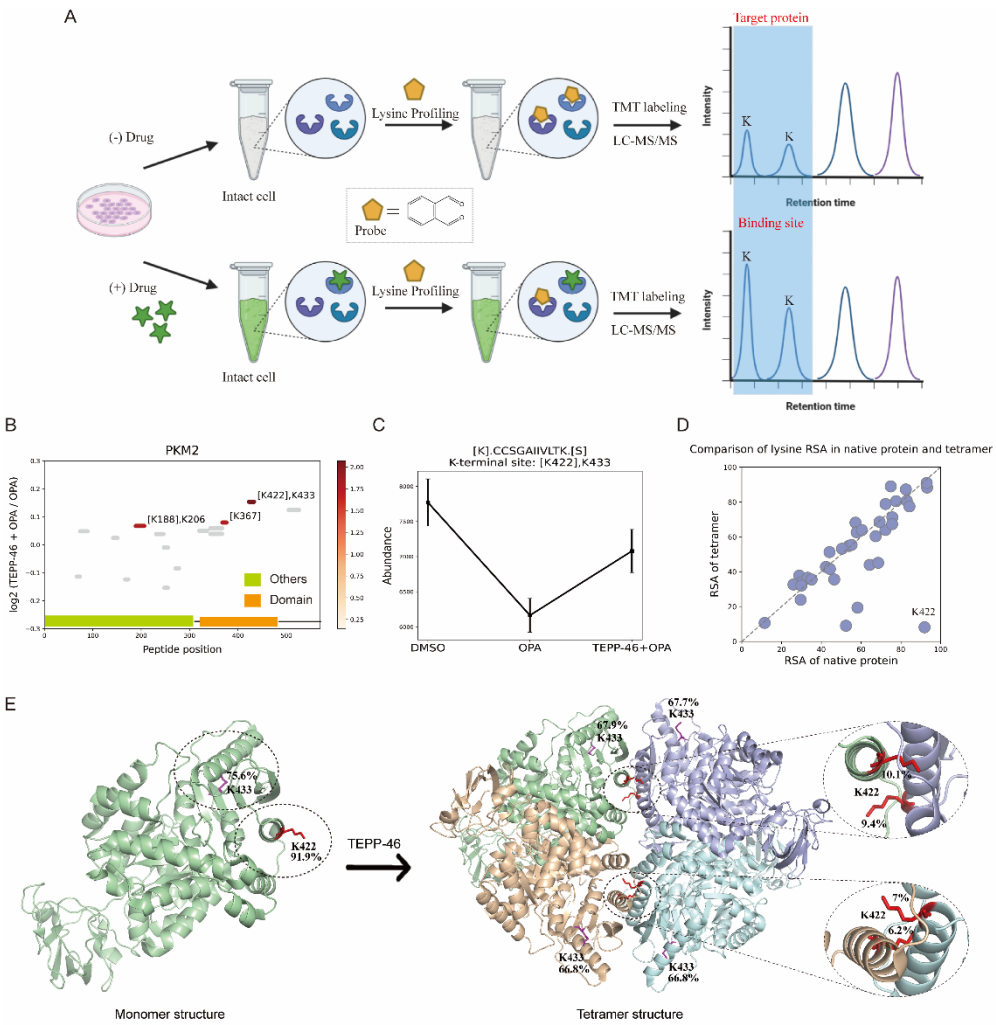

Fig. 4 Interrogation of the allosteric proteins PKM2 induced by TEPP-46. (A) Schema of samples collected. HEK293T cells are treated with drug before probe. (B) Diagram shows the locations and labeling reactivity shifts of identified peptides in PKM2. (C) Abundance of peptides [K^{422}].CCSGAIVLTK^{433}.[S]. (D) Scatter plot shows the comparison of lysine RSA in monomer PKM2 and tetrameric PKM2. (E) RSA of K422 in tetrameric PKM2 and monomer PKM2. The four PKM2 monomers and K422 are represented in cartoon mode with different colors. The red square is the binding region of the PKM2 protein and the drug TEPP-46.

decrease in abundance in samples incubated with OPA. However, in cells treated prior with TEPP-46 before OPA incubation, the decrease in abundance is significantly muted for native peptide [K^{422}].CCSGAIVLTK^{433} of PKM2 protein suggesting either the preceding Lys ^{422} or the Lys ^{433} within the peptide had been OPA-labelled (Fig.4B-C). Both Lys ^{422} and Lys ^{433} are exposed in monomer PKM2 (Fig.4E) but Lys ^{422} are lysine found on binding interface of tetramer PKM2 (Fig.4D). Specifically, the RSA of Lys^{422} and Lys^{433} in tetrameric PKM2 is less than 15% while is over 90% in monomer PKM2 (Fig. 4E). As the structure of target protein PKM2 will change from monomer to tetramer when treated with TEPP-46, this result validates the use of RAPID-OPA for studying changes in structural conformation and molecular interaction of proteins.

Interrogation of the target proteins and binding sites of geldanamycin

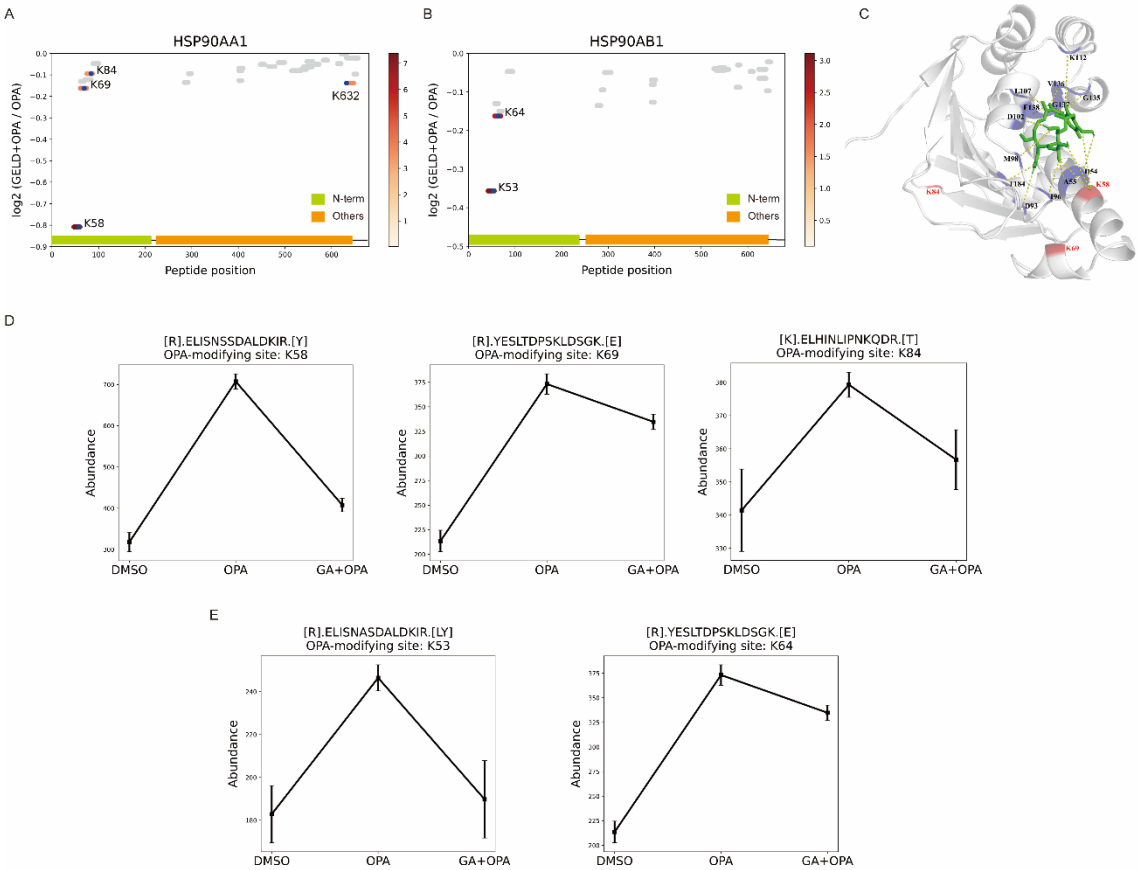

Fig. 5 Interrogation of the target proteins and binding sites of geldanamycin. (A-B) Diagram shows the locations and labeling reactivity shifts of identified peptides in HSP90AA1 and HSP90AB1 (C) PDB cocrystallization result of HSP90AA1. (D) Abundance of peptides [R].ELISNSSDALDK$^{58}_{OPA}$IR.[Y], [R].YESLTDPSK$^{69}_{OPA}$LDSGK^{74}.[E] and [K].ELHINLIPNK$^{84}_{OPA}$QDR.[T] in HSP90AA1. (E) Abundance of peptides[R].ELISNASDALDK$^{53}_{OPA}$IR.[Y] and [R].YESLTDPSK$^{64}_{OPA}$LDSGK^{69}.[E] in HSP90AB1.

Next, we used geldanamycin to assess applicability of RAPID-OPA to study drug-protein binding sites. Geldanamycin (GA) is a natural heat shock protein 90 (HSP90) inhibitor that specifically binds to ATP binding sites in the N-terminal domain of Hsp90 (residue 1-220).[30] Since drug-induced changes in the reactivity of lysine residues can be small in some cases and quantification at the peptide level are much more variable than at the protein level, reproducible experimental workflows with accurate quantification of peptide is crucial. Accordingly, four technical replicates are performed, and Pearson's correlation analysis showed a good reproducibility of peptide quantification among the four replicates (Fig. S3A-C). Totally, about 59,530 were identified in whole proteomics while 99 peptides and 92 peptides of known target proteins HSP90AA1 and HSP90AB1 were identified, respectively. Perhaps due to the high abundance of target proteins HSP90AA1 and HSP90AB1, their OPA-labelled peptides were identified with significant changes. Specifically, the peptides with significant lysine residues shifts in reactivity of these target proteins were [R].ELISNSSDALDK$^{58}_{OPA}$IR.[Y], [R].YESLTDPSK$^{69}_{OPA}$LDSGK^{74}.[E] and [K].ELHINLIPNK$^{84}_{OPA}$QDR.[T] for HSP90AA1, and [R].ELISNASDALDK$^{53}_{OPA}$IR.[Y] and [R].YESLTDPSK$^{64}_{OPA}$LDSGK^{69}.[E] for HSP90AB1 (Fig. 5A-B), all which were

located in the known binding regions in the target proteins.[26] Specifically, Lys^{58} was located directly in the interaction region, while Lys^{69} and Lys^{84} are located near the interaction region (Fig. 5C). The decrease in abundance is significantly muted for OPA-labeling peptide in drug treated group suggesting either the preceding Lys^{58}, Lys^{69}, Lys^{84}, Lys^{53} or Lys^{64} within the binding site had been occupied by geldanamycin before OPA treatment (Fig. 5D-E).

Conclusion

In brief, we developed an intracellular chemical covalent labeling method based on lysine reactive shift to analyze the conformational characteristics of proteins on a large scale from proteomics. RAPID is able to detect the extent of structural changes in proteins by denaturing them through heating. In addition, two model drugs were used to benchmark RAPID, and we successfully found the lysine site Lys ^{422}, which is masked by TEPP-46 to turn PKM2 into a tetramer, as well as the target protein and drug binding site regions of geldanamyci. Overall, RAPID can be used to study intracellular protein conformational changes and the identification of drug and target protein binding sites. We also envision RAPID can be further applied to detect protein complex binding interfaces and protein interactions in the future, enabling proteins to become effective biomarkers or for the study of drug target proteins.

Material methods

Reagents and Cell Culture. TEPP-46 and geldanamycin were purchased from MedChem Express. All the compounds were purchased from bidepharm (China). All the compounds were made up as solutions in DMSO (100x).

HEK293T cells were cultured in DMEM medium (Gibco) supplemented with 10% FBS (PAN) and 1% penicillin-streptomycin (Gibco) at 37℃, and 5% CO_2 in a humidified environment.

In vivo covalent small molecules for lysine reactivity. HEK293T cells were treated with 800 μM individual compounds or corresponding concentration of vehicle, dimethyl sulfoxide (DMSO), incubated at room temperature for 15 min followed by addition of glycine solution to a final concentration of 2 mM with 10 min incubation to quench the reaction. Then, each mixture subsequently subjected to (2x) lysis buffer containing a final concentration of 100 mM HEPES (pH 7.5), 20 mM $MgCl_2$, 10 mM β-Glycerophosphate (sodium salt hydrate), 2 mM Tris(2-carboxyethyl) phosphine hydrochloride (TCEP), 0.2 mM Sodium orthovanadate,0.2%(w/v) n-dodecyl β-D-maltoside (DDM), and EDTA-free protease inhibitor (Sigma-Aldrich, USA). Cell suspension was subjected to five times flash-freezing in liquid nitrogen and rapid thawing in water to facilitate cell lysis. After centrifugation at 21,000 g for 20 min at 4 °C, the supernatant was transferred to a new tube and the protein concentration was measured by the BCA assay kit (Thermo Fisher Scientific, USA).

RAPID-OPA analysis in purified protein. The MAP kinase-activated protein kinase 2 (MAPKAPK2) samples were prepared at 2 mg/mL with PBS (pH 7.4). Then, the MAPKAPK2 samples were subjected to labeling and quench step as above. After centrifugation at 21,000 g for 20 min at 4 °C, the supernatant was transferred to a new tube and the protein concentration was measured by the BCA assay kit.

Optimize RAPID-OPA labeling-time. HEK293T cells were treated either with 800 μM OPA or corresponding concentration of

DMSO, incubated at room temperature for different duration (15 min, 10 min, 5 min, 0 min) followed by addition of glycine solution to a final concentration of 2 mM with 10 min incubation to quench the reaction. The following steps were same as above.

Optimize RAPID-OPA reaction concentration. HEK293T cells were treated either with different concentration (50 μM, 100 μM, 200 μM, 400 μM, 800 μM) of OPA or corresponding concentration of DMSO, incubated at room temperature for 15 min. The following steps were same as above.

Heat stimulation treatment. HEK293T cells were treated either with 800 μM OPA or corresponding concentration of DMSO and heated in parallel in a PCR (VWR, Doppio Gradient) block for 15 min to the three temperatures (37°C, 43°C, 49°C). The following steps were same as above.

Profiling of lysine reactivity. HEK293T cells were treated with individual drugs (20 μM for TEPP-46 and 100 μM for geldanamycin) and incubated at room temperature for 20 min. Next, the cells were incubated with either 800 μM OPA or corresponding concentration of DMSO at room temperature for 15 min. The following steps were same as above.

In-solution digestion and TMT label. Each sample had 100 μg proteins in 60 μl 1x lysis buffer followed by reduced with 10 mM DTT in 95 °C for 10 min, and alkylated with 30 mM IAA for 30 min at room temperature. Then, protein was purified by methanol-chloroform precipitation (methanol/chloroform/water = 4:1:2). After lyophilized to dryness, protein pellet was resuspended in 8 M urea buffer and diluted to 2 M by adding 50 mM ammonium bicarbonate (ABC). Proteins were digested overnight at 37°C with trypsin (Promega) and Lys-C (Wako,125-05061) (trypsin/Lys-C/protein = 2:1:100 = w/w/w). The peptide concentration was measured by the Nanodrop (Thermo Fisher Scientific, USA). The samples of different compounds, MAPKAPK2, RAPID-OPA labeling-time, and different concentration which no need to label TMT were directly desalted with C18 disk (3 M Empore, U.S.A.) followed by dried vacuum. The other tryptic digests samples were desalted with C18 disk and labeled with TMTpro reagent (Thermo Fisher Scientific, USA) followed by dried vacuum. The TMT-mixed sample is then fractionated into six fractions by a stepwise increasing gradient of ACN. Finally, the eluted peptide samples were lyophilized to dryness and redissolved in 0.1% (v/v) formic acid in water for nano-LC-MS/MS analysis.

LFQ-RAPID analysis. Samples that were not labeled with TMT were subjected to analysis using a Q-Exactive instrument coupled with an Easy-nLC 1200 system (Thermo Fisher Scientific). The analytical column, consisting of an integrated spray tip (100 μm i.d. x 20 cm), was packed with 1.9 μm/120 Å ReproSil-Pur C18 beads (Dr. Maisch GmbH) for peptide separation at a flow rate of 250 nL/min. The samples were directly loaded onto the column with buffer A [0.1% (v/v) FA in water]. An 80 min gradient separation was performed as follows: 4%-8% buffer B [90% (v/v) ACN in buffer A] over 2 min, 8%-25% buffer B for 50 min, 25%-40% buffer B for 10 min, 40%-97% buffer B over 2 min, followed by 16 min wash with 97% buffer B. Peptides were detected on the Q-Exactive instrument, utilizing full MS scans on the Orbitrap mass analyzer with a resolution of 70,000, and top 10 MS/MS scans at a resolution of 35,000. HCD fragmentation was performed with a normalized collision energy (NCE) of 32, and a dynamic exclusion time of 30 seconds was applied.

TMT-DDA analysis. TMT-labeled samples were diluted with 0.1% FA prior to separation on 20 cm x 100 µm EASY-Spray C18 LC column with a 135 min gradient on an UltiMate 3000 HPLC system (Thermo Fisher Scientific). The mobile phase was solvent A (0.5% acetic acid in water) and solvent B (80% ACN, 0.5 % acetic acid in water). Data were acquired by Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific): MS1 scan resolution was 60,000 and MS/MS scan resolution was 30,000 using turbo TMT. Raw files were searched using Proteome Discoverer (PD) software (Version 2.4, Thermo Fisher Scientific) against the human proteome fasta database (Uniprot,20,376 entries, downloaded on May 03, 2022). The maximum missed cleavage for trypsin digestion was set to 2. The mass tolerance for peptide precursors was 10 ppm and the mass tolerance for fragment ions was 0.02 Da. TMT pro modification of lysine residues, OPA modification, oxidation (M) and deamidation (NQ) were selected as variable modifications. FDR control for protein and peptide is 1% at strict level and 5% at relaxed level.

TMT-PRM analysis. For PRM method, the LC-MS/MS instruments and chromatographic setups were identical to TMT-DDA method. The PRM scan mode: MS scan mode was set at the maximum injection time of 50 ms, AGC target set as custom at the resolution of 60,000. MS/MS scan mode was set at the maximum injection time of 100 ms, AGC target set as custom and an isolation window of 0.7 m/z at the resolution of 30,000. Then 38% normalized collision energy was set for HCD fragmentation. Targeted PKM2 precursor transitions and their corresponding m/z were extracted from the previous DDA discovery run and put into the PRM method as an inclusion list (Supplementary Table 1).

RSA prediction and data filtering. Solvent accessibility prediction was calculated using GetArea as described in Fraczkiewicz et al. (http://curie.utmb.edu/getarea.html). All human protein structures were downloaded from AlphaFold2 and imported into GetArea to get the accessibility profiles for each residue. The RSA value is the ratio of real side-chain SASA to average SASA in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations. In this study, we calculated two type of RSA value: peptide RSA and lysine RSA. The former considered the solvent accessibility of entire peptides, which is the average RSA value of all amino acids in the peptide. The latter focused on local solvent accessibility of the lysine site(s), indicated by the (average) RSA value of lysine residue(s).

For analysis of lysine-terminal peptides detected in the proteomics dataset, all unique lysine residues were extracted and matched to GetArea-predicted RSA values by UniProt accession ID and sequence position. Any contaminant proteins and proteins with no predicted RSA values were removed, and lysine-end peptides were then filtered by the following criteria: peptides with no missed cleavage were remained, and low-confidence predictions (Alphafold pLDDT $\geqslant 70$) were removed. Additionally, only unique peptides are retained after deduplicating by sequence.

Author contributions

C.T. conceived and led the project. Z.Z. and Y.Z. analyzed data. Y.Z. and B.L. wrote the algorithms. Z.Z. and K.L. performed the biological experiments. P.L. provide the compounds. Z.Z., Y.Z. and C.T. wrote the paper.

Conflicts of interest

The authors declare no competing interests.

Reference

1. H. Ji, X. Lu, S. Zhao, Q. Wang, B. Liao, L. G. Bauer, K. V. M. Huber, R. Luo, R. Tian and C. S. H. Tan, Target deconvolution with matrix-augmented pooling strategy reveals cell-specific drug-protein interactions, *Cell Chemical Biology*, 2023, **30**, 1-10.
2. M. Kaldmae, A. Leppert, G. Chen, M. Sarr, C. Sahin, K. Nordling, N. Kronqvist, M. Gonzalvo-Ulla, N. Fritz, A. Abelein, S. Laiotan, H. Biverstal, H. Jornvall, D. P. Lane, A. Rising, J. Johansson and M. Landreh, High intracellular stability of the spidroin N-terminal domain in spite of abundant amyloidogenic segments revealed by in-cell hydrogen/deuterium exchange mass spectrometry, *FEBS J*, 2020, **287**, 2823-2833.
3. J. A. Espino, V. S. Mali and L. M. Jones, In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells, *Anal. Chem.*, 2015, **87**, 7971-7978.
4. A. Rinas, V. S. Mali, J. A. Espino and L. M. Jones, Development of a Microflow System for In-Cell Footprinting Coupled with Mass Spectrometry, *Anal. Chem.*, 2016, **88**, 10052-10058.
5. U. Kaur, D. T. Johnson and L. M. Jones, Validation of the Applicability of In-Cell Fast Photochemical Oxidation of Proteins across Multiple Eukaryotic Cell Lines, *J. Am. Soc. Mass Spectrom.*, 2020, **31**, 1372-1379.
6. Q. Zhao, X. Ouyang, X. Wan, K. S. Gajiwala, J. C. Kath, L. H. Jones, A. L. Burlingame and J. Taunton, Broad-Spectrum Kinase Profiling in Live Cells with Lysine-Targeted Sulfonyl Fluoride Probes, *J. Am. Chem. Soc.*, 2017, **139**, 680-685.
7. C. Bamberger, S. Pankow, S. Martinez-Bartolome, M. Ma, J. Diedrich, R. A. Rissman and J. R. Yates, 3rd, Protein Footprinting via Covalent Protein Painting Reveals Structural Changes of the Proteome in Alzheimer's Disease, *J. Proteome Res.*, 2021, **20**, 2762-2771.
8. D. Cox, A. R. Ormsby, G. E. Reid and D. M. Hatters, Protein painting reveals pervasive remodeling of conserved proteostasis machinery in response to pharmacological stimuli, *NPJ Syst Biol Appl*, 2022, **8**, 46.
9. Y. Yang, H. P. Song, D. He, S. Zhang, S. Z. Dai, S. X. Lin, R. Meng, C. Wang and P. R. Chen, Genetically encoded protein photocrosslinker with a transferable mass spectrometry-identifiable label, *Nat Commun*, 2016, **7**.
10. B. Yang, S. B. Tang, C. Ma, S. T. Li, G. C. Shao, B. B. Dang, W. F. DeGrado, M. Q. Dong, P. G. Wang, S. Ding and L. Wang, Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions, *Nat Commun*, 2017, **8**, 2240.
11. W. Hu, Y. Yuan, C. H. Wang, H. T. Tian, A. D. Guo, H. J. Nie, H. Hu, M. J. Tan, Z. Tang and X. H. Chen, Genetically Encoded Residue-Selective Photo-Crosslinker to Capture Protein-Protein Interactions in Living Cells, *Chem-Us*, 2019, **5**, 2955-2968.
12. Z. Li, D. Wang, L. Li, S. Pan, Z. Na, C. Y. Tan and S. Q. Yao, "Minimalist" cyclopropene-containing photo-cross-linkers suitable for live-cell imaging and affinity-based protein labeling, *J. Am. Chem. Soc.*, 2014, **136**, 9990-9998.
13. X. R. Liu, M. M. Zhang and M. L. Gross, Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications, *Chem. Rev.*, 2020, **120**, 4355-4454.

14. G. H. Xu and M. R. Chance, Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chem. Rev.*, 2007, **107**, 3514-3543.
15. S. M. Hacker, K. M. Backus, M. R. Lazear, S. Forli, B. E. Correia and B. F. Cravatt, Global profiling of lysine reactivity and ligandability in the human proteome, *Nat. Chem.*, 2017, **9**, 1181-1190.
16. G. Akcay, M. A. Belmonte, B. Aquila, C. Chuaqui, A. W. Hird, M. L. Lamb, P. B. Rawlins, N. Su, S. Tentarelli, N. P. Grimster and Q. B. Su, Inhibition of Mcl-1 through covalent modification of a noncatalytic lysine side chain, *Nat. Chem. Biol.*, 2016, **12**, 931-936.
17. Z. Liu, W. Zhang, B. Sun, Y. Ma, M. He, Y. Pan and F. Wang, Probing conformational hotspots for the recognition and intervention of protein complexes by lysine reactivity profiling, *Chem. Sci.*, 2020, **12**, 1451-1457.
18. C. L. Tung, C. T. Wong, E. Y. Fung and X. Li, Traceless and Chemoselective Amine Bioconjugation via Phthalimidine Formation in Native Protein Modification, *Org Lett*, 2016, **18**, 2600-2603.
19. Q. Zhang, Y. Zhang, H. Liu, H. Y. Chow, R. Tian, Y. M. Eva Fung and X. Li, OPA-Based Bifunctional Linker for Protein Labeling and Profiling, *Biochemistry*, 2020, **59**, 175-178.
20. C. H. P. Cheung, T. H. Chong, T. Wei, H. Liu and X. Li, Guanidine Additive Enabled Intermolecular ortho-Phthalaldehyde-Amine-Thiol Three-Component Reactions for Modular Constructions, *Angew. Chem. Int. Ed. Engl.*, 2023, **62**, e202217150.
21. Y. Zhang, Q. Zhang, C. T. T. Wong and X. Li, Chemoselective Peptide Cyclization and Bicyclization Directly on Unprotected Peptides, *J. Am. Chem. Soc.*, 2019, **141**, 12274-12279.
22. B. Li, L. Wang, X. Chen, X. Chu, H. Tang, J. Zhang, G. He, L. Li and G. Chen, Extendable stapling of unprotected peptides by crosslinking two amines with o-phthalaldehyde, *Nat Commun*, 2022, **13**, 311.
23. X. Chu, B. Li, H. Y. Liu, X. Sun, X. Yang, G. He, C. Zhou, W. Xuan, S. L. Liu and G. Chen, Bioconjugation via Hetero-Selective Clamping of Two Different Amines with ortho-Phthalaldehyde, *Angew. Chem. Int. Ed. Engl.*, 2023, **62**, e202212199.
24. M. Simoes, L. C. Simoes, S. Cleto, I. Machado, M. O. Pereira and M. J. Vieira, Antimicrobial mechanisms of ortho-phthalaldehyde action, *J. Basic Microbiol.*, 2007, **47**, 230-242.
25. D. A. Polasky, D. J. Geiszler, F. C. Yu, K. Li, G. C. Teo and A. I. Nesvizhskii, MSFragger-Labile: A Flexible Method to Improve Labile PTM Analysis in Proteomics, *Mol. Cell. Proteomics*, 2023, **22**.
26. C. Ruan, J. Zhou, Z. Li, K. Li, Z. Fang, X. Zhang and M. Ye, Proteome-Wide Deconvolution of Drug Targets and Binding Sites by Lysine Reactivity Profiling, *Anal. Chem.*, 2022, **94**, 3352-3359.
27. Y. P. Zhou and R. W. Vachet, Covalent Labeling with Isotopically Encoded Reagents for Faster Structural Analysis of Proteins by Mass Spectrometry, *Anal. Chem.*, 2013, **85**, 9664-9670.
28. N. B. Herrington, D. Stein, Y. C. Li, G. Pandey and A. Schlessinger, Exploring the Druggable Conformational Space of Protein Kinases Using AI-Generated Structures, *bioRxiv*, 2023.
29. D. Anastasiou, Y. Yu, W. J. Israelsen, J.

K. Jiang, M. B. Boxer, B. S. Hong, W. Tempel, S. Dimov, M. Shen, A. Jha, H. Yang, K. R. Mattaini, C. M. Metallo, B. P. Fiske, K. D. Courtney, S. Malstrom, T. M. Khan, C. Kung, A. P. Skoumbourdis, H. Veith, N. Southall, M. J. Walsh, K. R. Brimacombe, W. Leister, S. Y. Lunt, Z. R. Johnson, K. E. Yen, K. Kunii, S. M. Davidson, H. R. Christofk, C. P. Austin, J. Inglese, M. H. Harris, J. M. Asara, G. Stephanopoulos, F. G. Salituro, S. Jin, L. Dang, D. S. Auld, H. W. Park, L. C. Cantley, C. J. Thomas and M. G. Vander Heiden, Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis, *Nat. Chem. Biol.*, 2012, **8**, 839-847.

30. R. R. A. Kitson and C. J. Moody, Learning from Nature: Advances in Geldanamycin- and Radicicol-Based Inhibitors of Hsp90, *J. Org. Chem.*, 2013, **78**, 5117-5141.