Uncover New Reactivity of Genetically Encoded Alkyl Bromide Non-Canonical Amino Acids

Xin Shu[1,2†], Sana Asghar[1,2†], Fan Yang[3], Shang-Tong Li[4]*, Haifan Wu[5]* and Bing Yang[1,2]*

[1]Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, China, [2]Cancer Center, Zhejiang University, Hangzhou, China, [3]Department of Biophysics, Kidney Disease Center of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, [4]Glbizzia Biosciences Co., Ltd, Beijing, China, [5]Department of Chemistry and Biochemistry, Wichita State University, Wichita, KS, United States

Genetically encoded non-canonical amino acids (ncAAs) with electrophilic moieties are excellent tools to investigate protein-protein interactions (PPIs) both *in vitro* and *in vivo*. These ncAAs, including a series of alkyl bromide-based ncAAs, mainly target cysteine residues to form protein-protein cross-links. Although some reactivities towards lysine and tyrosine residues have been reported, a comprehensive understanding of their reactivity towards a broad range of nucleophilic amino acids is lacking. Here we used a recently developed OpenUaa search engine to perform an in-depth analysis of mass spec data generated for Thioredoxin and its direct binding proteins cross-linked with an alkyl bromide-based ncAA, BprY. The analysis showed that, besides cysteine residues, BprY also targeted a broad range of nucleophilic amino acids. We validated this broad reactivity of BprY with Affibody/Z protein complex. We then successfully applied BprY to map a binding interface between SUMO2 and SUMO-interacting motifs (SIMs). BprY was further applied to probe SUMO2 interaction partners. We identified 264 SUMO2 binders, including several validated SUMO2 binders and many new binders. Our data demonstrated that BprY can be effectively used to probe protein-protein interaction interfaces even without cysteine residues, which will greatly expand the power of BprY in studying PPIs.

Keywords: protein-protein interactions, genetic code expansion, non-canonical amino acid, chemical cross-linking, and SUMO interactome

INTRODUCTION

Protein-protein interactions (PPIs) are essential for virtually all cellular processes in all living organisms. Thus, there is a significant effort in mapping protein-protein interaction networks to understand relevant biological processes in detail, and many techniques have been developed for this purpose (Low et al., 2021). Affinity purification mass spectrometry (AP-MS) have been successfully applied to map protein-protein interactomes in many organisms (Ho et al., 2002; Butland et al., 2005; Krogan et al., 2006; Gordon et al., 2020; Richards et al., 2021), although these methods cannot distinguish between direct binders and indirect binders of a protein of interest (POI) (Morris et al., 2014). Moreover, weak and transient interactions are typically not comprehensively detected under the conditions of AP-MS (Richards et al., 2021). Proximity labeling (PL) by introducing covalent labels to proteins proximal to a POI allows large scale analysis of protein-protein interactions with potential spatial and temporal resolutions in cells (Han et al., 2018; Roux et al., 2018). However, limited information is available to map the interaction interfaces to gain further structural understanding of these interactions.

FIGURE 1 | Broad reactivity of BprY to nucleophilic amino acids identified by data mining. **(A)** Scheme of *in-situ* BprY cross-linking and in-depth data analysis by OpenUaa. **(B)** MS/MS fragmentation patterns of cross-linked peptides with different nucleophilic AAs targeted by BprY. **(C)** Number of cross-linked peptides identified for different targeted AAs.
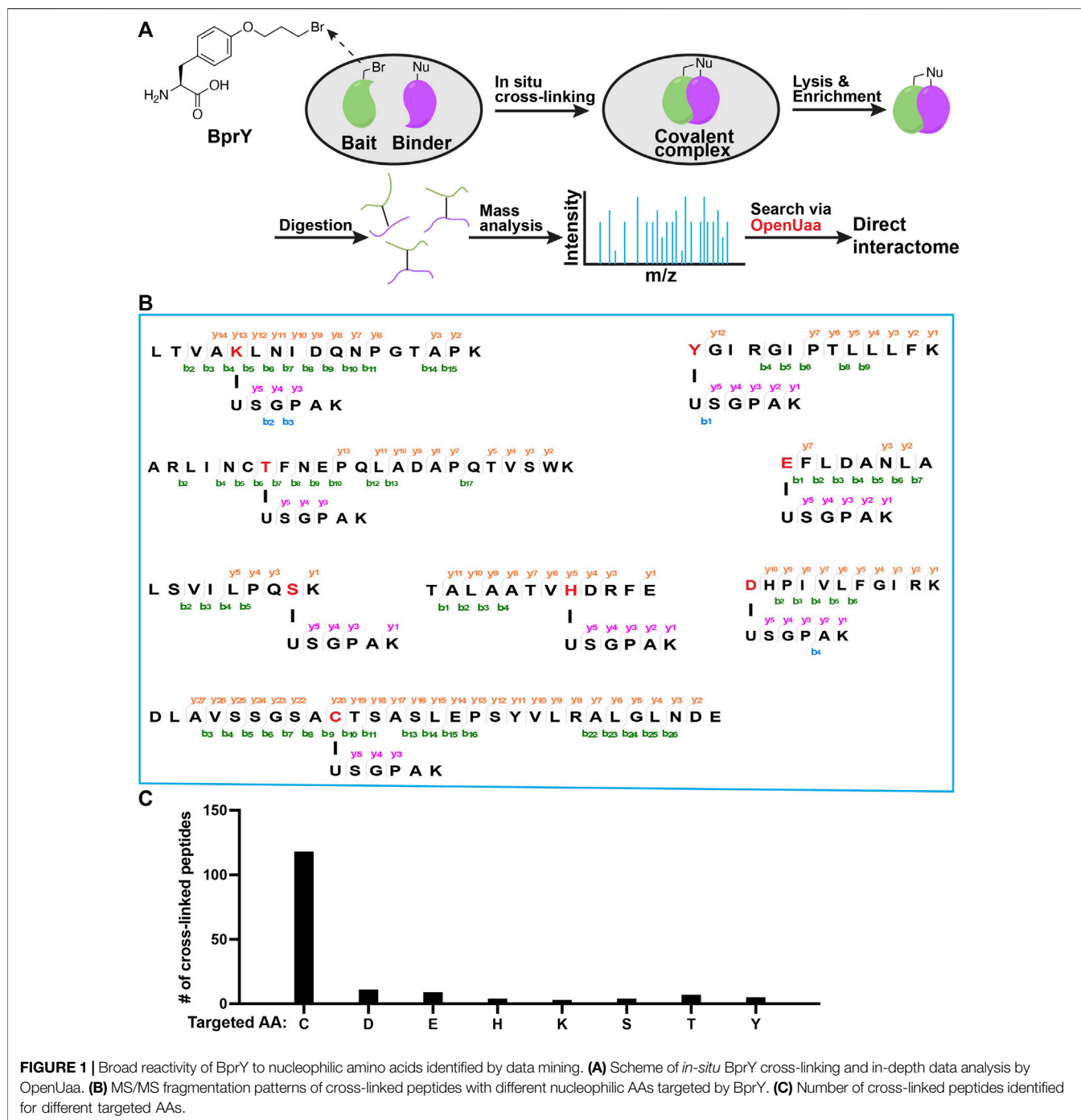
A complementary method to AP-MS and PL for analyzing PPIs involves covalent cross-linking (Liu et al., 2015; Wang, 2017; Nguyen et al., 2018; Yu and Huang, 2018). Genetic code expansion by amber codon suppression has enabled site-specific incorporation of non-canonical amino acids (ncAAs) into proteins (Wang et al., 2001; Wang and Schultz, 2004; Wang et al., 2006; de la Torre and Chin, 2021; Shandell et al., 2021). Many ncAAs, including photo-activated ncAAs (Chin et al., 2002a; Chin et al., 2002b; Zhang et al., 2011; Lin et al., 2014;

Yang et al., 2016) and those with fine-tuned bio-reactivity to capture protein binders (Xiang et al., 2013; Chen X.-H. et al., 2014; Furman et al., 2014; Xiang et al., 2014; Xuan et al., 2016; Cigler et al., 2017; Wang, 2017; Nguyen et al., 2018; Shang et al., 2018; Wang et al., 2018; Yang et al., 2019; Liu et al., 2020; Liu J. et al., 2021), has been successfully incorporated into proteins. One example is a chemical cross-linking ncAA BprY with an electrophilic alkyl bromide group (**Figure 1A**), which is typically unreactive towards biomolecules in cells after incorporation into

a POI unless there is a proximal nucleophilic cysteine residue from the binder of this POI through the proximity-enabled reactivity. Given its ability to cross-link proximal cysteine residues, BprY has been successfully applied to capture proteome-wide protein-protein interactions in live cells (Yang et al., 2017), allowing the development of GECX-MS (Genetically Encoded Chemical Cross-linking of proteins coupled with Mass Spectrometry) as a powerful tool to identify direct binding partners of target proteins (**Figure 1A**). In GECX-MS, a cross-linkable ncAA, such as BprY, is genetically incorporated into a bait protein to covalently capture binder proteins including those weak and transient PPIs *in situ*. Analysis of cross-linked peptides by mass spectroscopy identifies not only binder proteins but also their corresponding cross-linking sites, which could be used to map binding interfaces or binding motifs.

One limitation of BprY and related alkyl halide-based ncAAs is that the cross-linking reaction requires proximal cysteine residues in binder proteins. Because cysteine residues are in relatively low abundance and they often form disulfide bonds, potentially binder proteins without proximal cysteine residues may not be captured. To maximize the potential of BprY to probe protein-protein interactions, its amino acid reactivity beyond cysteine is critical. Although there are some reports of extended reactivity of alkyl bromide based ncAAs towards glutamate, lysine and histidine (Chen X.-H. et al., 2014; Xiang et al., 2014; Cigler et al., 2017), a comprehensive analysis of BprY reactivity towards different nucleophilic amino acids under physiological conditions is not available. Here we show that BprY can target a broad range of nucleophilic amino acid residues, including Cys, Asp, Glu, Ser, Thr, His, and Tyr, among which the reactivity to Cys is still the highest. BprY can also be used to probe PPIs without cysteine at the binding interface. Finally, we successfully applied BprY to identify SUMO2-interacting proteins at a whole proteome level.

MATERIALS AND METHODS

Plasmid Construction

pBad-Affibody and mutants. The gene encoding Affibody was PCR amplified with Affibody-NdeI-F and Affibody-HindIII-R primers, and the PCR product was cloned into a commercial pBad vector pre-treated with NdeI and HindIII enzymes. Affibody-D2N mutant was generated by PCR amplification of pBad-Affibody with Affibody-D2N-F and Affibody-HindIII-R primers. Affibody mutants pBad-Afb*(K7X) were generated by PCR amplification of pBad-Affibody with Affibody-Mutant-F and Affibody-Mutant-R primers.

Affibody-NdeI-F:
GGAGATATACATATGGTAGACAACGCCTTCAAC
Affibody-HindIII-R:
AAAACAGCCAAGCTTTTAGTGATGGTGATGGTG
ATGA
Affibody-D2N-F:
GGAGATATACATATGGTAAACAACGCCTTCAACAAG
Affibody-Mutant-F:
AACAACGCCTTCAACCxxxCAACTATCAGTCGCC

Affibody-Mutant-R:
GGCGACTGATAGTTGxxxGTTGAAGGCGTTGTT

pBad-MBP-Z-24TAG. The gene encoding MBP-Z fusion protein was PCR amplified with the following primers, and the PCR product was cloned into the pBad vector pre-digested with NdeI and HindIII enzymes. pBad-MBP-Z-24TAG was generated by PCR amplification of pBad-MBP-Z with MBP-Z-24TAG-E25Q-Mutant primers.

MBP-Z-NdeI-F:
GGAGATATACATATGATGAAAATCGAAGAAGGT
AAACTG
MBP-Z-HindIII-R:
CAAAACAGCCAAGCTTTTAATGATGATGATGATGAT
GCTTAGG
MBP-Z-24TAG-E25Q-Mutant-F:
TTACCTAACCTGAATTAGCAGCAGCGTAATGCCTTC
MBP-Z-24TAG-E25Q -Mutant-R:
GAAGGCATTACGCTGCTGCTAATTCAGGTTAGGTAA

pBad-SUMO2 and mutants. The gene encoding *Homo sapiens* SUMO-2 (NCBI Reference Sequence: NM_006937.3) was PCR amplified with the following primers, and the PCR product was cloned into the pBad vector pre-treated with NdeI and HindIII enzymes. Sites of hSUMO2-E49 and R50 were mutated to a TAG codon respectively.

hSUMO2-NdeI-F:
GGAGATATACATATGATGGCCGACGAAAAGC
hSUMO2-HindIII-R:
AACAGCCAAGCTTTCAGTGATGGTGATGGTGATGGT
AGACACCTCCCGTCT

pET28a-MBP-RNF111$_{293-391}$. The gene encoding *Homo sapiens* RNF111$_{293-391}$ (NCBI Reference Sequence: NM_001270530.1) was PCR amplified with the following primers, and the PCR product was cloned into a pET28a-MBP vector.

pET28a-MBP-RNF111-F:
AAGTTCTGTTCCAGGGGCCCCATATGATGTCAGGAA
GTATTGATGAAGATGTTG
pET28a-MBP-RNF111-R:
CAGTGGTGGTGGTGGTGGTGCTCGAGTTCATCTTCA
TCAACGGTAAGGTC

Protein Expression

Affibody, SUMO2 and MBP-RNF111$_{293-391}$. The corresponding plasmids were individually transformed into DH10B cells, which were plated on LB agar plates supplemented with 100 μg/mL ampicillin. Colonies were picked from the plate of each plasmid and individually inoculated to 100 mL LB (5 g/L NaCl, 10 g/L Tryptone, 10 g/L Yeast extract). Cells were grown at 37°C and 200 rpm to an OD of 0.6 with good aeration and the relevant antibiotic selection. For induction of MBP-RNF111$_{293-391}$, 400 μM IPTG was added. 0.2% L-arabinose was added to induce the expression of Affibody and SUMO2. The expression was carried out at 30°C, 200 rpm for 5h. Cells were harvested by centrifugation at 6,000 *g*, 4°C for 10 min. The cell pellet was washed with cold PBS buffer and centrifuged again at 6,000 g, 4°C for 10 min. Cell pellets were then frozen in liquid nitrogen and stored at −80°C.

ncAA constructs: MBP-Z(E24BprY), SUMO2(E49BprY), and SUMO2*(R50BprY).* The corresponding plasmids were individually transformed into DH10B cells together with pEvol-BprY plasmid. Cells were plated on LB agar plates supplemented with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. Colonies were picked from these plates and individually inoculated to 25 mL LB (5 g/L NaCl, 10 g/L Tryptone, 10 g/L Yeast extract). Cells were grown at 37°C and 200 rpm to an OD of 0.4 with good aeration and the relevant antibiotic selection. Then the medium was added 1 mM ncAA BprY and 0.2% L-arabinose. The expression was carried out at 18°C and 200 rpm shaking for 18 h. Cells were harvested by centrifugation at 3,260 g and 4°C for 20 min. Cell pellets were washed with cold PBS buffer, centrifuged again at 3,260 g and 4°C for 20 min. Cell pellets were then frozen in liquid nitrogen and stored at −80°C.

His-Tag Protein Purification

Frozen cells were rapidly thawed and resuspended in lysis buffer (50 mM Tris, pH8.0, 500 mM NaCl, 0.1% Tween-20). EDTA S3 free protease inhibitor cocktail was added followed by vortexing for 2 min. Cells were lysed by sonication after which the cell lysate was clarified by centrifugation at 13,000 rpm and 4°C for 30 min. The supernatant was collected and incubated with 200 µL Ni-NTA Affinity resin at 4°C for 1 h. The resin was washed with an equal volume of wash buffer (50 mM Tris pH8.0, 500 mM NaCl, 20 mM imidazole) for 2 times at 4°C. Elution was done with 200 µl elution buffer (50 mM Tris pH8.0, 100 mM NaCl, 250 mM imidazole) for five times. The fractions containing the target protein were determined by SDS-PAGE analysis with 10% Tricine gel.

Proteins Cross-Linking *in vitro*

A 20 µL reaction mixture containing MBP-Z(E24BprY) (0.6 mg/mL) and Affibody (1.2 mg/mL) in HEPES buffer (pH 7.5) was incubated at 37°C for 8 h. A 20 µL reaction mixture containing SUMO2 (10 µM) and RNF111 (293-391) (10 µM) in HEPES buffer (pH 7.5) was incubated at 37°C for 8 h.

Protein Digestion

Protein samples were precipitated by the addition of six volumes of cold acetone (−20°C) and incubated at −20°C for 30 min. Precipitated proteins were air dried and resuspended in 8 M urea, 100 mM Tris, pH 8.5. After reduction with 5 mM TCEP for 20 min and alkylation with 10 mM iodoacetamide for 15 min in the dark, samples were diluted to 2 M urea with 100 mM Tris, pH 8.5, and digested with trypsin (at 50:1 protein: enzyme ratio) at 37°C for 16 h. Digestion was terminated by adding formic acid to 5% final concentration, and digested peptides were desalted with StageTips.

Cell Lysate Cross-Linking and Two-step His-Tag Purification to Enrich Cross-Linked Peptides

Cell pellets expressing His-tagged SUMO2*(E49BprY) or SUMO2*(R50BprY) were resuspended in 4 ml lysis buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 0.1% Tween-20), separately. Cells were lysed by sonication after which the cell lysate was clarified by centrifugation at 13,000 rpm and 4°C for 30 min. The supernatants were collected and incubated with 200 µl Ni-NTA Affinity resin at 4°C for 2 h. The resin was washed with 4 ml of wash buffer (50 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole) for 3 times at 4°C, followed by a second rinsed with an equal volume of wash buffer 2 (50 mM Tris pH8.0, 500 mM NaCl, 40 mM imidazole) for 3 times at 4°C. Then the resin was equilibrated with lysis buffer and incubated with 293T cell lysates from one 10 cm plate at RT for overnight. The next day, the resin was washed with 4 mL wash buffer (50 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole) for 3 times at 4°C, and then rinsed with an equal volume of wash buffer 2 (50 mM Tris pH8.0, 500 mM NaCl, 40 mM imidazole) for 3 times at 4°C. Elution was done with 200 µl elution buffer (50 mM Tris pH 8.0, 100 mM NaCl, 250 mM imidazole) for five times. The eluates were concentrated, and buffer exchanged into 100 µl of protein storage buffer (50 mM HEPES, pH 7.5, and 100 mM NaCl) using 10 k Amicon Ultra columns. Purified proteins were digested with Lys-C at 37°C for overnight, and digested peptides were incubated with pre-equilibrated Ni-NTA Agarose resin (50 µL) at 4°C for 2 h to further enrich cross-linked peptides (all contain C-terminal His tag after Lys-C digestion). Resin was rinsed with wash buffer 2 for three times and 50 mM NH_4HCO_3 twice. Bound peptides were digested on-bead with Trypsin at 37°C for 8 h, and digested peptides were desalted with StageTips before MS analysis.

Tandem MS Analysis

MS experiments were performed on a Q Exactive HF-X instrument (ThermoFisher) coupled with an Easy-nLC 1200 system. Mobile phase A and B were water and 80% acetonitrile, respectively, with 0.1% formic acid. Digested peptides were loaded directly onto analytical column (75 µm × 20 cm, 1.9 µm C18, 5 µm tip) at a flow rate of 300 nL/min. All peptide samples were separated using a linear gradient of 6–22% B for 38 min, 22–35% B for 17 min, 35–90% B in 2 min, 90% B for 1 min, 100% B for 2 min. Survey scans of peptide precursors were performed from 350 to 1500 m/z at 60,000 FWHM resolution with a 1×10^6 ion count target and a maximum injection time of 20 ms. The instrument was set to run in top-speed mode with 1-s cycles for the survey and the MS/MS scans. After a survey scan, tandem MS was then performed on the most abundant precursors exhibiting a charge state from 3 to 7 of greater than 1×10^5 intensity by isolating them in the quadrupole at 1.6 m/z. Higher energy collisional dissociation (HCD) fragmentation was applied with 27% collision energy and resulting fragments detected in the Orbitrap detector at a resolution of 15,000. The maximum injection time limited was 30 ms, and dynamic exclusion was set to 30 s with a 10 ppm mass tolerance around the precursor.

MS Data Analysis

MS/MS spectra were extracted by parsing from RAW file. Datasets of model proteins were searched against the corresponding proteins by OpenUaa. OpenUaa was also used

to search data of two-step purified Trx sample and SUMO interaction protein sample against *E. coli* proteome and human proteome downloaded from the UniProt database and the reversed decoy proteins, separately. OpenUAA search parameters: 5% false discovery rate (FDR) at the peptide-spectrum match (PSM) level, 10 ppm precursor mass tolerance, 20 ppm fragment mass tolerance, variable modification Cys 57.02146, and three maximum number of missed cleavage sites.

Molecular Docking Study

For molecular docking, we first relaxed the structures of SUMO2 and RNF4 (PDB ID: 6JXX and 4PPE, respectively) in the Rosetta software suite (Leaver-Fay et al., 2011). 1000 models were generated for each protein and the model with the lowest total score was chosen for docking. In the protein-protein docking using Rosetta suite, RNF4 was first randomly oriented relative to SUMO2, and then 10,000 docking models were generated. The final docking model was chosen as the model with the largest binding energy (Rosetta energy term: I_sc). The docking model was rendered in the UCSF Chimera (Pettersen et al., 2004).

Bioinformatic Analysis

For the analysis of sequence motif around the cross-linked sites, MEME software was used to scan a ±15 amino acids sequence window around the cross-linked site to generate a consensus motif. Gene ontology (GO) term and KEGG pathway enrichment for functional analysis were performed using the clusterProfiler package under the R software (Yu et al., 2012). The human proteome was used as the background for enrichment analysis. The significance of the enrichment analysis was defined using a hypergeometric test, and the resulted p values were corrected for multiple hypothesis testing using the BH method (Hendriks and Vertegaal, 2016). The final reported enriched terms and pathways were filtered according to the adjusted p values <0.05. STRING network analysis was performed using the STRING database, which integrate multiple information sources for protein interaction speculation (Szklarczyk et al., 2019). Using all identified interacting proteins of SUMO2 as input, protein interactions between SUMO2 interacting proteins were filtered at a STRING interaction confidence of >0.4. Statistical analysis of the interaction network was performed based on the interactions in the networks compared to randomly expected frequency of interactions. Network visualization was performed using Cytoscape software version 3.7.2. The MCODE plugin of Cytoscape was used to extract the 8 interconnected modules under default settings for other parameters (Bader and Hogue, 2003).

RESULTS

Deep mining of interactome data identified a broad reactivity of BprY. We chose to investigate those nucleophilic amino acids known to carry out alkylation reactions (deGruyter et al., 2017). Nucleophilic substitution reactions with Asn/Gln/Arg sidechains are rare. Thus, we think they are unlikely to react with BprY. For
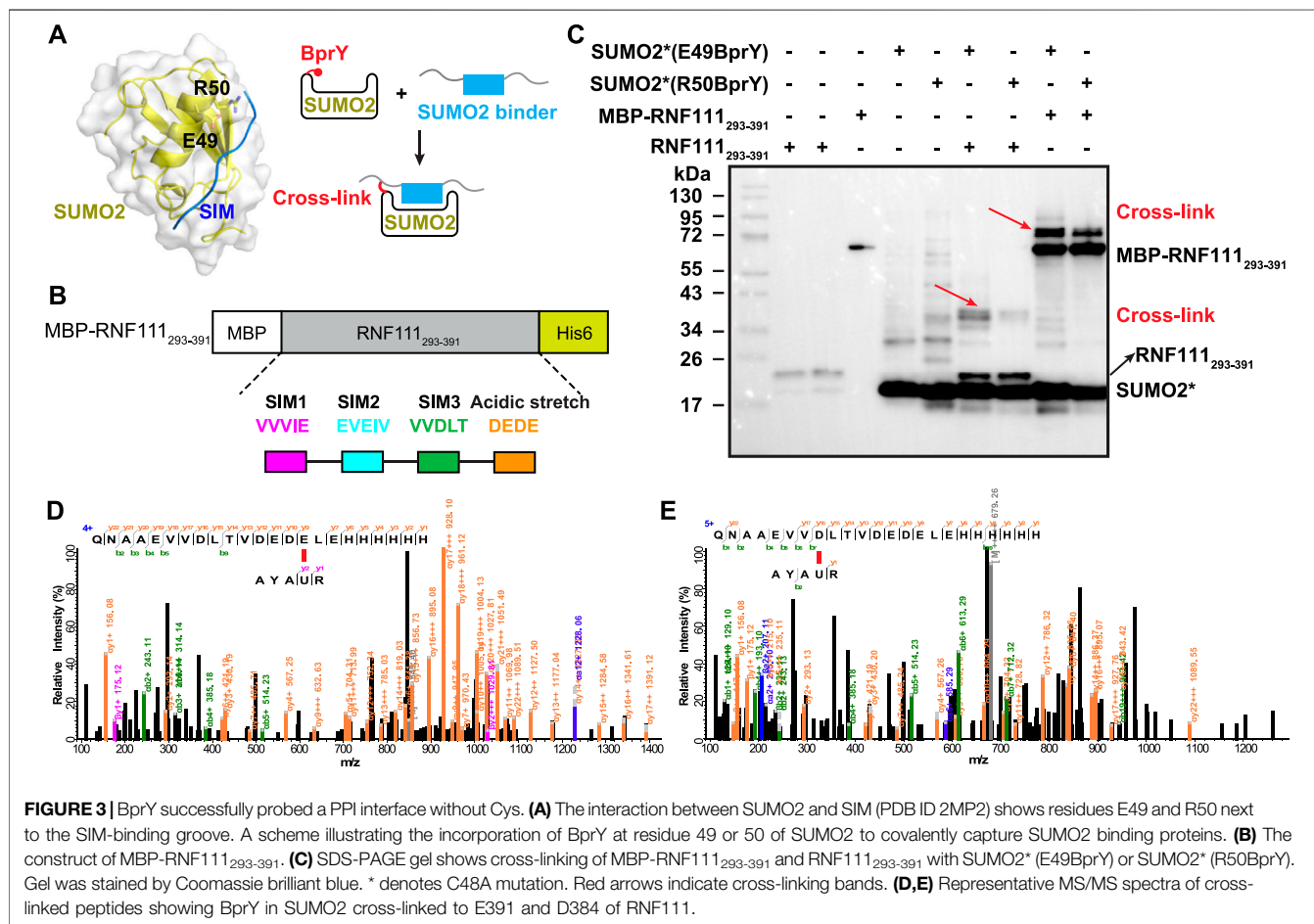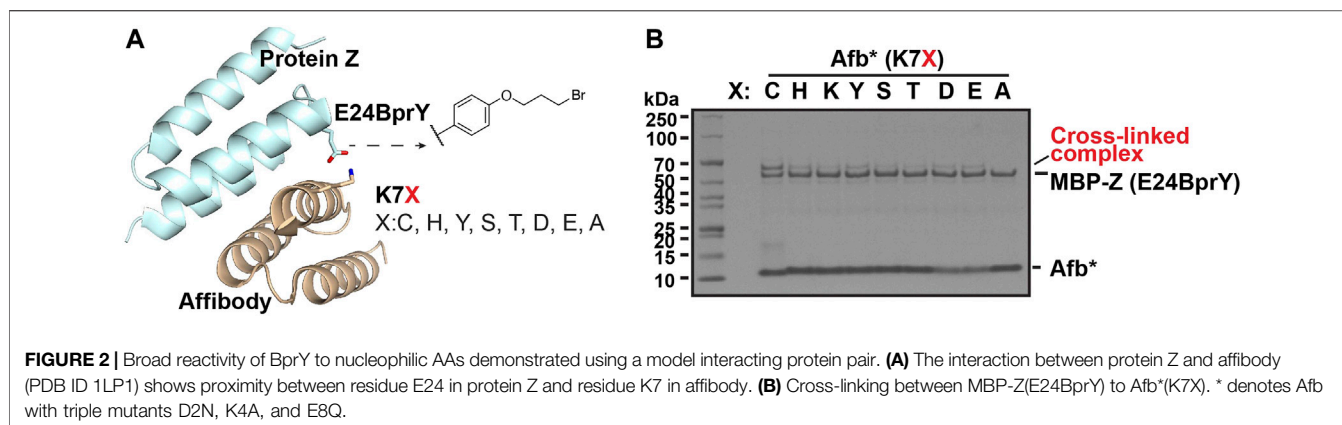
Met, alkylation with benzyl bromides has been reported using model peptides under acidic conditions (Kramer and Deming, 2013). It is also known that iodoacetamide, a common alkylation reagent used in proteomics, can alkylate Met (Kruger et al., 2005). Therefore, we think it is possible that Met might also react with BprY. One concern is that the alkylation product sulfonium can undergo a dealkylation reaction (Kramer and Deming, 2013) and can dissociate upon collision to cause neutral mass loss (Kruger et al., 2005), which can further complicate the MS analysis. Due to these reasons, we decided to focus on Cys, Asp, Glu, His, Lys, Ser, Thr and Tyr in the current study. To systematically evaluate the reactivity of BprY to different types of amino acid residues in proteins, we re-analyzed previously published direct interactome data (Yang et al., 2017) of thioredoxin (Trx) probed by BprY (**Figure 1A**). Using a recently developed searching algorism OpenUaa (Liu C. et al., 2021), which allowed deeper mining of the interactome data, we observed a broad reactivity of BprY to multiple nucleophilic amino acids—Cys, Asp, Glu, His, Lys, Ser, Thr and Tyr, supported by high-quality MS/MS spectra (Figures 1B; **Supplementary Figure S1**). As expected, the reactivity is dominated by Cys with 118 Cys-targeted peptides identified. However, we also observed 43 non-Cys cross-linked peptides which still made a significant contribution to determining the direct interactome (**Figure 1C**).

Validate the reactivity using a model interacting protein pair. To validate this broad reactivity of BprY, we employed a model interacting protein pair—affibody (Afb) and protein Z fused to maltose-binding protein (MBP-Z) (**Figure 2A**). Efficient cross-linking was observed when a Cys residue and BprY were introduced to replace residue K7 of Afb and residue E24 of MBP-Z, respectively (**Supplementary Figure S2A**), consistent with a previous study (Yang et al., 2017). Several other AAs were also introduced at residue 7 of Afb, and the extent of cross-linking between Afb(K7X, X: variable AA) and MBP-Z(E24BprY) was evaluated. To our surprise, even some degree of cross-linking was observed between the "unreactive" control Afb(K7A) and MBP-Z(E24BprY) (**Supplementary Figure S2A**), suggesting nucleophilic amino acid residues adjacent to residue 7 in Afb can also facilitate cross-linking. We found that a triple mutation D2N, K4A, and E8Q of Afb completely eliminated cross-linking of the "unreactive" control (**Figure 2B**). Therefore, this triple mutation of Afb, denoted as Afb* thereafter, was further investigated. As shown in **Figure 2B**, Afb*(K7C) was efficiently cross-linked to MBP-Z(E24BprY), suggesting that the triple mutation didn't affect the binding of Afb to MBP-Z. Further cross-linking of MBP-Z with Afb*(K7X) showed that, compared to Afb*(K7C), several nucleophilic amino acids displayed reactivity to BprY, although the cross-linking efficiency as evaluated from the intensity of the cross-linking band is lower than that of Cys.

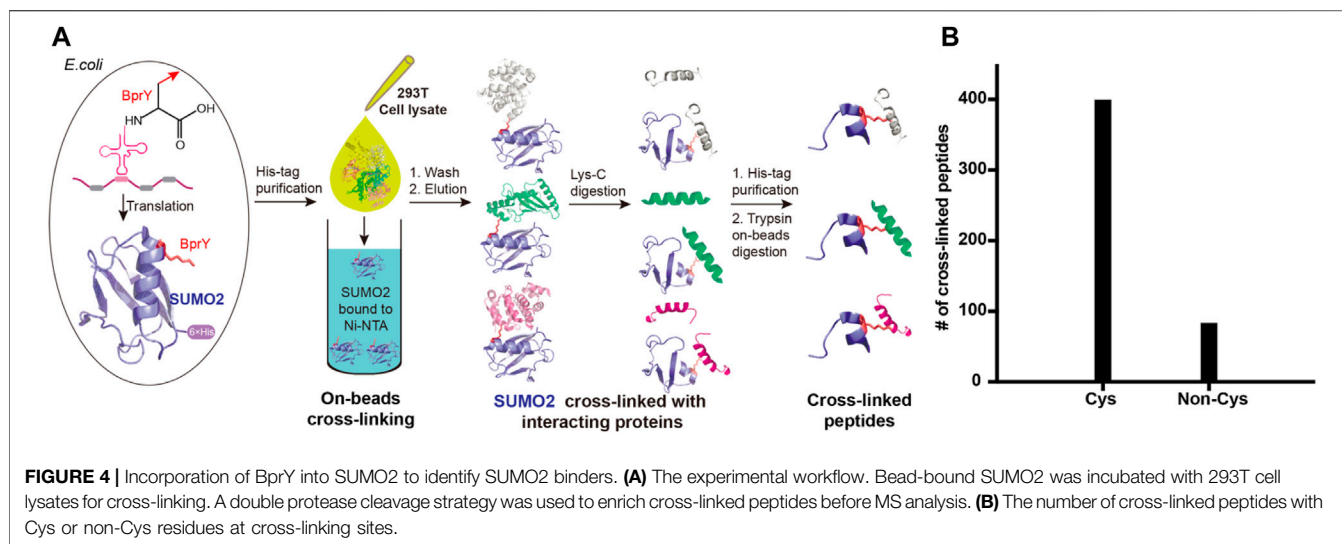*Apply BprY to study PPI without Cys at the interface in vitro.* The study with model interacting protein pair further supported that BprY can target multiple nucleophilic amino acid residues beyond Cys, suggesting that BprY can be used to probe protein-protein interactions even without Cys residues at the interaction interface. To test this idea, we used BprY to capture the interaction between small ubiquitin-like modifiers (SUMOs)

FIGURE 2 | Broad reactivity of BprY to nucleophilic AAs demonstrated using a model interacting protein pair. **(A)** The interaction between protein Z and affibody (PDB ID 1LP1) shows proximity between residue E24 in protein Z and residue K7 in affibody. **(B)** Cross-linking between MBP-Z(E24BprY) to Afb*(K7X). * denotes Afb with triple mutants D2N, K4A, and E8Q.

FIGURE 3 | BprY successfully probed a PPI interface without Cys. **(A)** The interaction between SUMO2 and SIM (PDB ID 2MP2) shows residues E49 and R50 next to the SIM-binding groove. A scheme illustrating the incorporation of BprY at residue 49 or 50 of SUMO2 to covalently capture SUMO2 binding proteins. **(B)** The construct of MBP-RNF111$_{293-391}$. **(C)** SDS-PAGE gel shows cross-linking of MBP-RNF111$_{293-391}$ and RNF111$_{293-391}$ with SUMO2* (E49BprY) or SUMO2* (R50BprY). Gel was stained by Coomassie brilliant blue. * denotes C48A mutation. Red arrows indicate cross-linking bands. **(D,E)** Representative MS/MS spectra of cross-linked peptides showing BprY in SUMO2 cross-linked to E391 and D384 of RNF111.

and one of their binding partners—RNF111. SUMOs can be reversibly conjugated to lysine side-chains of target proteins by an enzymatic cascade involving E1-E2-E3 enzymes, and this modification plays a key role in genome stability and transcription. The interaction between SUMOs and their binding partners is mediated primarily by SUMO-interacting motifs (SIMs) containing 3-4 aliphatic amino acid residues (**Figure 3A**) (Flotho and Melchior, 2013). RNF111 is a

SUMO-targeted ubiquitin ligase with three SIMs and an acidic stretch adjacent to SIM3 (Poulsen et al., 2013; Sriramachandran et al., 2019).

To probe the binding between SUMO2 and RNF111, we generated two SUMO2 constructs with BprY individually incorporated into SUMO2 at residue E49 or R50, adjacent to the SIM-binding groove (**Figure 3A**). In both constructs, residue C48 was mutated to Ala to prevent intra-protein cross-linking.
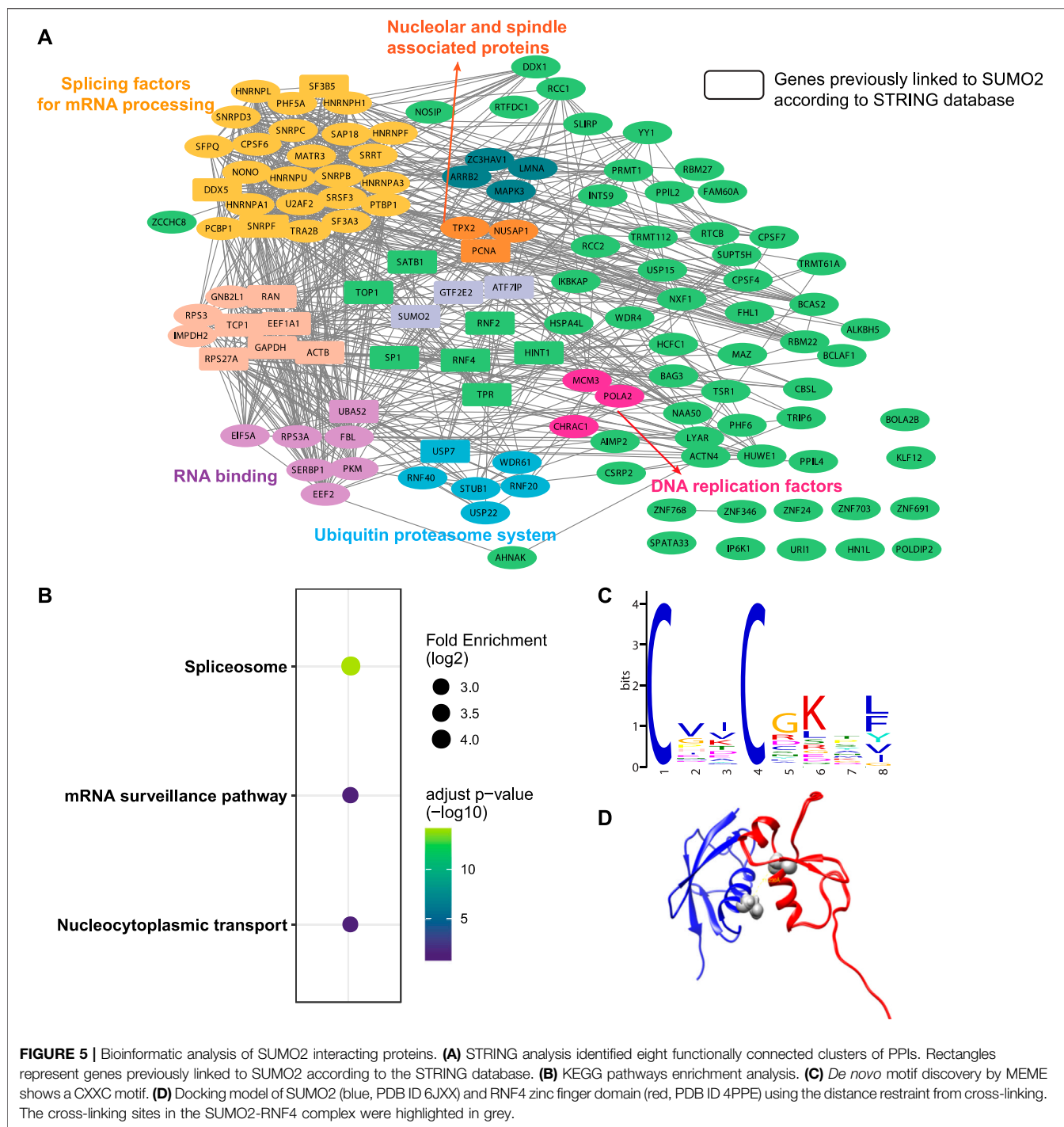
FIGURE 4 | Incorporation of BprY into SUMO2 to identify SUMO2 binders. **(A)** The experimental workflow. Bead-bound SUMO2 was incubated with 293T cell lysates for cross-linking. A double protease cleavage strategy was used to enrich cross-linked peptides before MS analysis. **(B)** The number of cross-linked peptides with Cys or non-Cys residues at cross-linking sites.

Mutation of these residues has been known to have a small effect on SUMO2 binding (Bouchenna et al., 2019; Bruninghoff et al., 2020). These two constructs SUMO2*(E49BprY) and SUMO2*(R50BprY) were then individually incubated with MBP-RNF111$_{293-391}$ (**Figure 3B**). As shown in **Figure 3C**, both SUMO2*(E49BprY) and SUMO2*(R50BprY) formed inter-protein cross-links with MBP-RNF111$_{293-391}$. The same result was also observed with RNF111$_{293-391}$, suggesting the cross-links are specific to RNF111. Incorporation of BprY at residue 49 of SUMO2* appeared to give more efficient cross-linking, suggesting that the side-chain of residue 49 might be better positioned for the cross-linking reaction. Another possibility is that mutation of R50 caused weaker binding to SIMs. Further mass analysis identified cross-links between BprY in SUMO2* to residues D384 in SIM3 and E391 in the acidic stretch of RNF111 (**Figures 3D,E**), supporting that this acidic stretch contributes to the binding of SIM3 in RNF111 to SUMO2.

Proteome-wide identification of SUMO2 interacting proteins by BprY. We next applied BprY to covalently capture SUMO2 interaction partners at the whole proteome level using 293T cell lysates (**Figure 4A**). After the expression of C-terminal His-tagged SUMO2*(E49BprY) and SUMO2*(R50BprY) in *E. coli* DH10B cells, both constructs were purified by Ni-NTA resin, and the resin-bound SUMO2 was incubated with 293T cell lysates to capture SUMO2 binders by forming cross-links. The resin was then stringently washed followed by elution to give SUMO2 and SUMO2 cross-linked to its interacting proteins. Because there is no lysine residue after K44 in SUMO2, Lys-C digestion followed by a second-step Ni-NTA purification can further enrich cross-linked peptides (**Figure 4A**). Final on-bead trypsin digestion was done before mass analysis.

A total of 482 cross-linked peptides were detected (399 Cys and 83 non-Cys, **Figure 4B**), corresponding to 264 SUMO2 binders among which 139 are proteins with nucleus localization. Cysteine residues only account for about 2% of the total residues in proteins and likely have low frequency to

appear at binding interfaces compared with all the other amino acids investigated here. However, our SUMO interactome data showed that the number of identified cysteine-targeted cross-links is four times higher than that of non-cysteine cross-links. This is mainly due to higher intrinsic reactivity of cysteine sidechain to BprY. We believe that, for non-cysteine cross-links to occur efficiently, the interaction needs to be relatively strong, and the non-cysteine sidechain needs to be in an optimal geometry to react with BprY. Because SUMO2 is mainly localized in the nucleus, we first attempted to analyze these nuclear protein binders. Among them, we observed a previously reported SUMO2 binder RNF4 as well as many known SUMOylation substrates, including PCNA, SATB1, *etc* (**Figure 5A**). An analysis using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database showed that most of the SUMO2 binders were situated in a large network in which proteins are functionally/physically connected. At medium STRING confidence, 126 of the 139 proteins are in a single interaction network, with a PPI enrichment *p*-value < 1e-16, suggesting the identified SUMO2 binders are functional connected. Furthermore, we performed MCODE analysis on the SUMO2 binder network and identified 8 highly interconnected clusters within the core network (**Figure 5A**), including the spliceosome, ubiquitin proteasome system, DNA replication factors, and RNA processing factors. Interestingly, members (MCM3/CHRAC1/POLA2) of a cluster are all newly identified SUMO2 binders (Li et al., 2018; Dang and Morales, 2020). Since SUMOylation plays critical roles in DNA damage response (DDR) and numerous SUMO conjugates have been identified, the interaction between SUMO2 and MCM3/CHRAC1/POLA2 may establish a new link between SUMOylation mediated double strand breaks repair and maintenance of genome stability.

The KEGG pathway enrichment analysis highlighted three major pathways associated with these binders—spliceosome, mRNA surveillance pathway, and nucleocytoplasmic transport, consistent with some of the major functions of SUMOylation (**Figure 5B**). Interestingly, *de novo* motif discovery using MEME

FIGURE 5 | Bioinformatic analysis of SUMO2 interacting proteins. (A) STRING analysis identified eight functionally connected clusters of PPIs. Rectangles represent genes previously linked to SUMO2 according to the STRING database. (B) KEGG pathways enrichment analysis. (C) *De novo* motif discovery by MEME shows a CXXC motif. (D) Docking model of SUMO2 (blue, PDB ID 6JXX) and RNF4 zinc finger domain (red, PDB ID 4PPE) using the distance restraint from cross-linking. The cross-linking sites in the SUMO2-RNF4 complex were highlighted in grey.

based on sequences flanking the cross-linking site of SUMO2 binders revealed a CXXC motif (**Figure 5C**), which is commonly found in zinc finger proteins. Indeed, many identified proteins have zinc finger domains (ZNF24/346/691/703/768 and ZC3HAV1). This finding is consistent with previous reports suggesting that other than SIMs, zinc fingers can also bind SUMOs (Danielsen et al., 2012; Guzzo et al., 2014; Diehl et al., 2016). Besides, the enrichment of lysine in the motif suggests a close distance between cross-linked sites and SUMOylation sites.

One advantage of GECX-MS is the ability to identify cross-linking sites, which can be used to generate distance restraints for structural modeling. We focused on RNF4, a known binder of SUMO2, and attempted to reveal the potential conformation of the SUMO2-RNF4 complex with molecular docking. RNF4 has four SIMs and a zinc finger domain. Previous studies have identified SIM2 and SIM3 as the major contributor to SUMO binding (Kung et al., 2014; Xu et al., 2014). Interestingly, in this study, we found a cross-link between SUMO2 to the zinc finger domain of RNF4,

suggesting that this domain may also be involved in SUMO binding. By applying the distance restraint from cross-linking, we performed molecular docking of SUMO2 and RNF4 zinc finger domain using the Rosetta software suite (Leaver-Fay et al., 2011). The docking model (**Figure 5D**) revealed a binding interface on SUMO2 close to the SIM-binding groove. One thing to be noted is that because the BprY mediated cross-linking in this study was done in cell lysates with mixed cellular compartments, we also identified many non-nuclear proteins as SUMO2 binders. Gene ontology (GO) analysis suggested a potential link between SUMO2 to protein translation and localization (**Supplementary Figure S3**). Although there have been studies showing this link (Xu et al., 2010a; Xu et al., 2010b; Chen L.-z. et al., 2014), more detailed investigation will be needed in future studies.

CONCLUSION

In conclusion, we have demonstrated that the alkyl bromide containing ncAA BprY can react with not only cysteine residues but also a broad range of nucleophilic amino acids. Therefore, the application of BprY will not be limited to PPIs containing cysteine residues at the binding interface. Indeed, this aspect has been successfully demonstrated by our *in vitro* study of SUMO2/RNF111 interaction in which there are no cysteine residues at the binding site. With this broad reactivity, we applied BprY to covalently capture and identify 264 SUMO2 interacting proteins at a whole proteome level. This study further demonstrated that BprY and the relevant alkyl halide ncAAs are excellent tools to study protein-protein interactions.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD031159 and 10.6019/PXD031159.

AUTHOR CONTRIBUTIONS

XS conducted experiments. SA and S-TL performed bioinformatic analysis. FY performed protein modeling. BY and HW conceived and directed the project. S-TL, BY, and HW wrote the article.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.815991/full#supplementary-material

REFERENCES

Bader, G. D., and Hogue, C. W. (2003). An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks. *BMC Bioinformatics* 4, 2. doi:10.1186/1471-2105-4-2

Bouchenna, J., Sénéchal, M., Drobecq, H., Stankovic-Valentin, N., Vicogne, J., and Melnyk, O. (2019). The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis. *Bioconjug. Chem.* 30 (10), 2684–2696. doi:10.1021/acs.bioconjchem.9b00598

Brüninghoff, K., Aust, A., Taupitz, K. F., Wulff, S., Dörner, W., and Mootz, H. D. (2020). Identification of SUMO Binding Proteins Enriched after Covalent Photo-Cross-Linking. *ACS Chem. Biol.* 15 (9), 2406–2414. doi:10.1021/acschembio.0c00609

Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., et al. (2005). Interaction Network Containing Conserved and Essential Protein Complexes in Escherichia coli. *Nature* 433 (7025), 531–537. doi:10.1038/nature03239

Chen, L.-z., Li, X.-y., Huang, H., Xing, W., Guo, W., He, J., et al. (2014a). SUMO-2 Promotes mRNA Translation by Enhancing Interaction between eIF4E and eIF4G. *PLoS One* 9 (6), e100457. doi:10.1371/journal.pone.0100457

Chen, X.-H., Xiang, Z., Hu, Y. S., Lacey, V. K., Cang, H., and Wang, L. (2014b). Genetically Encoding an Electrophilic Amino Acid for Protein Stapling and Covalent Binding to Native Receptors. *ACS Chem. Biol.* 9 (9), 1956–1961. doi:10.1021/cb500453a

Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002a). Addition of a Photocrosslinking Amino Acid to the Genetic Code of Escherichia coli. *Proc. Natl. Acad. Sci.* 99 (17), 11020–11024. doi:10.1073/pnas.172226299

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002b). Addition of P-Azido-L-Phenylalanine to the Genetic Code of Escherichia coli. *J. Am. Chem. Soc.* 124 (31), 9026–9027. doi:10.1021/ja027007w

Cigler, M., Müller, T. G., Horn-Ghetko, D., von Wrisberg, M.-K., Fottner, M., Goody, R. S., et al. (2017). Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes *In Vitro* and *In Vivo*. *Angew. Chem. Int. Ed.* 56 (49), 15737–15741. doi:10.1002/anie.201706927

Dang, T. T., and Morales, J. C. (2020). Involvement of POLA2 in Double Strand Break Repair and Genotoxic Stress. *Ijms* 21 (12), 4245. doi:10.3390/ijms21124245

Danielsen, J. R., Povlsen, L. K., Villumsen, B. H., Streicher, W., Nilsson, J., Wikström, M., et al. (2012). DNA Damage-Inducible SUMOylation of HERC2 Promotes RNF8 Binding via a Novel SUMO-Binding Zinc finger. *J. Cel Biol* 197 (2), 179–187. doi:10.1083/jcb.201106152

de la Torre, D., and Chin, J. W. (2021). Reprogramming the Genetic Code. *Nat. Rev. Genet.* 22 (3), 169–184. doi:10.1038/s41576-020-00307-7

deGruyter, J. N., Malins, L. R., and Baran, P. S. (2017). Residue-Specific Peptide Modification: A Chemist's Guide. *Biochemistry* 56 (30), 3863–3873. doi:10.1021/acs.biochem.7b00536

Diehl, C., Akke, M., Bekker-Jensen, S., Mailand, N., Streicher, W., and Wikström, M. (2016). Structural Analysis of a Complex between Small Ubiquitin-like Modifier 1 (SUMO1) and the ZZ Domain of CREB-Binding Protein (CBP/p300) Reveals a New Interaction Surface on SUMO. *J. Biol. Chem.* 291 (24), 12658–12672. doi:10.1074/jbc.M115.711325

Flotho, A., and Melchior, F. (2013). Sumoylation: a Regulatory Protein Modification in Health and Disease. *Annu. Rev. Biochem.* 82, 357–385. doi:10.1146/annurev-biochem-061909-093311

Furman, J. L., Kang, M., Choi, S., Cao, Y., Wold, E. D., Sun, S. B., et al. (2014). A Genetically Encoded Aza-Michael Acceptor for Covalent Cross-Linking of Protein-Receptor Complexes. *J. Am. Chem. Soc.* 136 (23), 8411–8417. doi:10.1021/ja502851h

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* 583 (7816), 459–468. doi:10.1038/s41586-020-2286-9

Guzzo, C. M., Ringel, A., Cox, E., Uzoma, I., Zhu, H., Blackshaw, S., et al. (2014). Characterization of the SUMO-Binding Activity of the Myeloproliferative and Mental Retardation (MYM)-type Zinc Fingers in ZNF261 and ZNF198. *PLoS One* 9 (8), e105271. doi:10.1371/journal.pone.0105271

Han, S., Li, J., and Ting, A. Y. (2018). Proximity Labeling: Spatially Resolved Proteomic Mapping for Neurobiology. *Curr. Opin. Neurobiol.* 50, 17–23. doi:10.1016/j.conb.2017.10.015

Hendriks, I. A., and Vertegaal, A. C. O. (2016). A High-Yield Double-Purification Proteomics Strategy for the Identification of SUMO Sites. *Nat. Protoc.* 11 (9), 1630–1649. doi:10.1038/nprot.2016.082

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S.-L., et al. (2002). Systematic Identification of Protein Complexes in Saccharomyces cerevisiae by Mass Spectrometry. *Nature* 415 (6868), 180–183. doi:10.1038/415180a

Kramer, J. R., and Deming, T. J. (2013). Reversible Chemoselective Tagging and Functionalization of Methionine Containing Peptides. *Chem. Commun.* 49 (45), 5144–5146. doi:10.1039/c3cc42214c

Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., et al. (2006). Global Landscape of Protein Complexes in the Yeast Saccharomyces cerevisiae. *Nature* 440 (7084), 637–643. doi:10.1038/nature04670

Krüger, R., Hung, C.-W., Edelson-Averbukh, M., and Lehmann, W. D. (2005). Iodoacetamide-alkylated Methionine Can Mimic Neutral Loss of Phosphoric Acid from Phosphopeptides as Exemplified by Nano-Electrospray Ionization Quadrupole Time-Of-Flight Parent Ion Scanning. *Rapid Commun. Mass. Spectrom.* 19 (12), 1709–1716. doi:10.1002/rcm.1976

Kung, C. C.-H., Naik, M. T., Wang, S.-H., Shih, H.-M., Chang, C.-C., Lin, L.-Y., et al. (2014). Structural Analysis of Poly-SUMO Chain Recognition by the RNF4-SIMs Domain. *Biochem. J.* 462 (1), 53–65. doi:10.1042/BJ20140521

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., et al. (2011). Rosetta3. *Methods Enzymol.* 487, 545–574. doi:10.1016/B978-0-12-381270-4.00019-6

Li, M., Xu, X., Chang, C.-W., Zheng, L., Shen, B., and Liu, Y. (2018). SUMO2 Conjugation of PCNA Facilitates Chromatin Remodeling to Resolve Transcription-Replication Conflicts. *Nat. Commun.* 9 (1), 2706. doi:10.1038/s41467-018-05236-y

Lin, S., He, D., Long, T., Zhang, S., Meng, R., and Chen, P. R. (2014). Genetically Encoded Cleavable Protein Photo-Cross-Linker. *J. Am. Chem. Soc.* 136 (34), 11860–11863. doi:10.1021/ja504371w

Liu, C., Wu, T., Shu, X., Li, S. T., Wang, D. R., Wang, N., et al. (2021a). Identification of Protein Direct Interactome with Genetic Code Expansion and Search Engine OpenUaa. *Adv. Biol.* 5 (3), 2000308. doi:10.1002/adbi.202000308

Liu, F., Rijkers, D. T. S., Post, H., and Heck, A. J. R. (2015). Proteome-wide Profiling of Protein Assemblies by Cross-Linking Mass Spectrometry. *Nat. Methods* 12 (12), 1179–1184. doi:10.1038/nmeth.3603

Liu, J., Cao, L., Klauser, P. C., Cheng, R., Berdan, V. Y., Sun, W., et al. (2021b). A Genetically Encoded Fluorosulfonyloxybenzoyl-L-Lysine for Expansive

Covalent Bonding of Proteins via SuFEx Chemistry. *J. Am. Chem. Soc.* 143 (27), 10341–10351. doi:10.1021/jacs.1c04259

Liu, J., Cheng, R., Van Eps, N., Wang, N., Morizumi, T., Ou, W.-L., et al. (2020). Genetically Encoded Quinone Methides Enabling Rapid, Site-specific, and Photocontrolled Protein Modification with Amine Reagents. *J. Am. Chem. Soc.* 142 (40), 17057–17068. doi:10.1021/jacs.0c06820

Low, T. Y., Syafruddin, S. E., Mohtar, M. A., Vellaichamy, A., A Rahman, N. S., Pung, Y.-F., et al. (2021). Recent Progress in Mass Spectrometry-Based Strategies for Elucidating Protein-Protein Interactions. *Cell. Mol. Life Sci.* 78 (13), 5325–5339. doi:10.1007/s00018-021-03856-0

Morris, J. H., Knudsen, G. M., Verschueren, E., Johnson, J. R., Cimermancic, P., Greninger, A. L., et al. (2014). Affinity Purification-Mass Spectrometry and Network Analysis to Understand Protein-Protein Interactions. *Nat. Protoc.* 9 (11), 2539–2554. doi:10.1038/nprot.2014.164

Nguyen, T.-A., Cigler, M., and Lang, K. (2018). Expanding the Genetic Code to Study Protein-Protein Interactions. *Angew. Chem. Int. Ed.* 57 (44), 14350–14361. doi:10.1002/anie.201805869

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., et al. (2022). The PRIDE Database Resources in 2022: a Hub for Mass Spectrometry-Based Proteomics Evidences. *Nucleic Acids Res.* 50 (D1), D543–D552. doi:10.1093/nar/gkab1038

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera?A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 25 (13), 1605–1612. doi:10.1002/jcc.20084

Poulsen, S. L., Hansen, R. K., Wagner, S. A., van Cuijk, L., van Belle, G. J., Streicher, W., et al. (2013). RNF111/Arkadia Is a SUMO-Targeted Ubiquitin Ligase that Facilitates the DNA Damage Response. *J. Cel Biol* 201 (6), 797–807. doi:10.1083/jcb.201212075

Richards, A. L., Eckhardt, M., and Krogan, N. J. (2021). Mass Spectrometry-based Protein-Protein Interaction Networks for the Study of Human Diseases. *Mol. Syst. Biol.* 17 (1), e8792. doi:10.15252/msb.20188792

Roux, K. J., Kim, D. I., Burke, B., and May, D. G. (2018). BioID: A Screen for Protein-Protein Interactions. *Curr. Protoc. Protein Sci.* 91, 19 23 11–19 23 15. doi:10.1002/cpps.51

Shandell, M. A., Tan, Z., and Cornish, V. W. (2021). Genetic Code Expansion: A Brief History and Perspective. *Biochemistry* 60, 3455–3469. doi:10.1021/acs.biochem.1c00286

Shang, X., Chen, Y., Wang, N., Niu, W., and Guo, J. (2018). Oxidation-induced Generation of a Mild Electrophile for Proximity-Enhanced Protein-Protein Crosslinking. *Chem. Commun.* 54 (33), 4172–4175. doi:10.1039/c8cc01639a

Sriramachandran, A. M., Meyer-Teschendorf, K., Pabst, S., Ulrich, H. D., Gehring, N. H., Hofmann, K., et al. (2019). Arkadia/RNF111 Is a SUMO-Targeted Ubiquitin Ligase with Preference for Substrates Marked with SUMO1-Capped SUMO2/3 Chain. *Nat. Commun.* 10 (1), 3678. doi:10.1038/s41467-019-11549-3

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi:10.1093/nar/gky1131

Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001). Expanding the Genetic Code of Escherichia coli. *Science* 292 (5516), 498–500. doi:10.1126/science.1060077

Wang, L. (2017). Genetically Encoding New Bioreactivity. *New Biotechnol.* 38 (Pt A), 16–25. doi:10.1016/j.nbt.2016.10.003

Wang, L., and Schultz, P. G. (2005). Expanding the Genetic Code. *Angew. Chem. Int. Ed.* 44 (1), 34–66. doi:10.1002/anie.200460627

Wang, L., Xie, J., and Schultz, P. G. (2006). Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* 35, 225–249. doi:10.1146/annurev.biophys.35.101105.121507

Wang, N., Yang, B., Fu, C., Zhu, H., Zheng, F., Kobayashi, T., et al. (2018). Genetically Encoding Fluorosulfate-L-Tyrosine to React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins In Vivo. *J. Am. Chem. Soc.* 140 (15), 4995–4999. doi:10.1021/jacs.8b01087

Xiang, Z., Lacey, V. K., Ren, H., Xu, J., Burban, D. J., Jennings, P. A., et al. (2014). Proximity-enabled Protein Crosslinking through Genetically Encoding Haloalkane Unnatural Amino Acids. *Angew. Chem. Int. Ed.* 53 (8), 2190–2193. doi:10.1002/anie.201308794

Xiang, Z., Ren, H., Hu, Y. S., Coin, I., Wei, J., Cang, H., et al. (2013). Adding an Unnatural Covalent Bond to Proteins through Proximity-Enhanced Bioreactivity. *Nat. Methods* 10 (9), 885–888. doi:10.1038/nmeth.2595

Xu, X., Vatsyayan, J., Gao, C., Bakkenist, C. J., and Hu, J. (2010a). HDAC2 Promotes eIF4E Sumoylation and Activates mRNA Translation Gene Specifically. *J. Biol. Chem.* 285 (24), 18139–18143. doi:10.1074/jbc.C110.131599

Xu, X., Vatsyayan, J., Gao, C., Bakkenist, C. J., and Hu, J. (2010b). Sumoylation of eIF4E Activates mRNA Translation. *EMBO Rep.* 11 (4), 299–304. doi:10.1038/embor.2010.18

Xu, Y., Plechanovová, A., Simpson, P., Marchant, J., Leidecker, O., Kraatz, S., et al. (2014). Structural Insight into SUMO Chain Recognition and Manipulation by the Ubiquitin Ligase RNF4. *Nat. Commun.* 5, 4217. doi:10.1038/ncomms5217

Xuan, W., Li, J., Luo, X., and Schultz, P. G. (2016). Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins. *Angew. Chem. Int. Ed.* 55 (34), 10065–10068. doi:10.1002/anie.201604891

Yang, B., Tang, S., Ma, C., Li, S.-T., Shao, G.-C., Dang, B., et al. (2017). Spontaneous and Specific Chemical Cross-Linking in Live Cells to Capture and Identify Protein Interactions. *Nat. Commun.* 8 (1), 2240. doi:10.1038/s41467-017-02409-z

Yang, B., Wang, N., Schnier, P. D., Zheng, F., Zhu, H., Polizzi, N. F., et al. (2019). Genetically Introducing Biochemically Reactive Amino Acids Dehydroalanine and Dehydrobutyrine in Proteins. *J. Am. Chem. Soc.* 141 (19), 7698–7703. doi:10.1021/jacs.9b02611

Yang, Y., Song, H., He, D., Zhang, S., Dai, S., Lin, S., et al. (2016). Genetically Encoded Protein Photocrosslinker with a Transferable Mass Spectrometry-Identifiable Label. *Nat. Commun.* 7, 12299. doi:10.1038/ncomms12299

Yu, C., and Huang, L. (2018). Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* 90 (1), 144–165. doi:10.1021/acs.analchem.7b04431

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, M., Lin, S., Song, X., Liu, J., Fu, Y., Ge, X., et al. (2011). A Genetically Incorporated Crosslinker Reveals Chaperone Cooperation in Acid Resistance. *Nat. Chem. Biol.* 7 (10), 671–677. doi:10.1038/nchembio.644