Genetically encoded protein photocrosslinker with a transferable mass spectrometry-identifiable label

Yi Yang¹, Haiping Song¹, Dan He¹, Shuai Zhang¹, Shizhong Dai¹, Shixian Lin¹, Rong Meng¹, Chu Wang¹,² & Peng R. Chen¹,²

Coupling photocrosslinking reagents with mass spectrometry has become a powerful tool for studying protein–protein interactions in living systems, but it still suffers from high rates of false-positive identifications as well as the lack of information on interaction interface due to the challenges in deciphering crosslinking peptides. Here we develop a genetically encoded photo-affinity unnatural amino acid that introduces a mass spectrometry-identifiable label (MS-label) to the captured prey proteins after photocrosslinking and prey–bait separation. This strategy, termed IMAPP (In-situ cleavage and MS-label transfer After Protein Photocrosslinking), enables direct identification of photo-captured substrate peptides that are difficult to uncover by conventional genetically encoded photocrosslinkers. Taking advantage of the MS-label, the IMAPP strategy significantly enhances the confidence for identifying protein–protein interactions and enables simultaneous mapping of the binding interface under living conditions.
Techniques for discovering and characterizing protein–protein interactions under physiological conditions are under constant development, among which chemical and photo-affinity crosslinking strategies have drawn increasing attention in recent years. The ability of converting non-covalent interactions between biomolecules into covalent linkages allows capture of weak and transient protein–protein interactions frequently found in nature. Genetically encoded photocrosslinkers, typically in the form of photo-affinity ‘unnatural amino acids’ (UAAs), can be incorporated into proteins at desired positions via the genetic code expansion strategy. This method uses an orthogonal aminoacyl-transfer RNA (tRNA) synthetase (aaRS)-tRNA pair to incorporate the desired UAA such as a photocrosslinker in response to an in-frame amber codon in the target gene, allowing site-specific photocrosslinking for capturing protein–protein interactions under living conditions. However, the downstream procedures for target separation and identification still follow regular affinity pull-down protocols that suffer from the problem of false identification due to non-specific protein binding and/or indirect protein interactions. More importantly, the crosslinking peptides and modification sites are difficult to uncover by current photocrosslinkers, which would otherwise provide valuable information regarding the interaction interface.

We envision that these limitations can be overcome by integrating a stable transferable mass spectrometry-identifiable label (MS-label) into the photo-affinity probe, which can be subsequently transferred to the crosslinked prey proteins through a cleavage linker after protein photocrosslinking. By searching the MS-label modified peptides, the crosslinked interacting proteins can be readily distinguished from the background, thus improving the specificity, confidence as well as robustness of the target identification process. Meanwhile, the MS-label modified peptides can provide structural information of interaction interface. Herein, we report the design and development of such a genetically encoded photo-affinity UAA that contains a cleavable linker for prey–bait separation and an in situ generated MS-label that can be transferred to the prey proteins upon cleavage. Embarked on this unique photocrosslinker, we create a novel chemical proteomic strategy, termed ‘IMAPP’ (In-situ cleavage and MS-label transfer After Protein Photocrosslinking), that enables simultaneous identification of the captured peptides and the exact crosslinking sites, which is highly valuable for uncovering target proteins as well as mapping protein–protein interaction interfaces under living conditions.

**Results**

**Design and characterization of DiZHSeC.** We have recently developed a genetically encoded photocrosslinker-DiZPK containing the diazirine group that can be used for highly efficient photo-affinity capture of protein–protein interactions in living systems. By replacing the γ-carbon with a selenium (Se) atom, DiZPK can be further converted to a cleavable photocrosslinker-DiZSeK in which the selenium–carbon bond can undergo oxidative cleavage upon H₂O₂ treatment, facilitating the prey–bait separation as well as the downstream target identification. Both these UAAs are pyrrolysine (Pyl) analogues that can be specifically recognized by a mutant pyrrolysyl-tRNA synthetase (PylRS) with its cognitive tRNA_pyl from archaeal species such as Methanosarcina barkeri (Mb). However, the selenenic acid moiety generated after oxidative cleavage on the prey proteins is too labile for MS identification and, therefore, cannot serve as a stable MS-label. To overcome this limitation, we herein designed an alternative UAA structure by replacing the δ-carbon on DiZPK by a Se atom and changing the ε-amine to a methylene group. The resulting photocrosslinker (Se-(N-(3-(3-methyl-3H-diazirin-3-yl)propyl)propanamide)-3-yl-homoselenocysteine), named as DiZHSeC, can undergo the oxidation-mediated C–Se bond cleavage and produce an N-(4,4-bis-substituted-pentyl)acrylamide (NPAA) moiety that is stable and readily identifiable by mass spectrometry (Fig. 1a,b). We synthesized DiZHSeC according to the procedure shown in Supplementary Fig. 1 and monitored its oxidative cleavage by ultra-performance liquid chromatography mass spectrometry. Upon treatment with 8 mM H₂O₂ for 60 min, we observed full cleavage of the C–Se bond in DiZHSeC and the generation of a new peak corresponding to the cleaved product P1 containing the NPAA moiety (Supplementary Fig. 2). As expected, the resulting acrylamide group from H₂O₂-mediated cleavage survived through the oxidative condition and remained intact, making it a stable tag suitable for chemical derivatization and MS identification.

**Figure 1 | Development of genetically encoded photocrosslinker with a transferable MS-label.** (a) In situ generation of MS-label on prey proteins by using a genetically encoded cleavable photocrosslinker. (b) Chemical design of the photocrosslinker (DiZHSeC) with transferable MS-label. The in situ generated NPAA MS-label can be verified by either fluorogenic labelling or can be directly identified by mass spectrometry.
Site-specific encoding of DiZHSeC in *Escherichia coli*. After verifying the chemical properties of DiZHSeC, we started by testing the site-specific incorporation of DiZHSeC into model proteins. The DiZPK-recognizing PyrRS mutant (L274A, C313S and Y349F)-tRNACUA<sub>Pyl</sub> pair showed a similar amber suppression efficiency in the presence of DiZHSeC or DiZPK when an in-frame amber mutation site was introduced at residue N149 in green fluorescent protein (Fig. 2a). We expressed and purified the resulting protein GFP-N149DiZHSeC in *E. coli* and verified its molecular weight by electrospray ionization (ESI)-MS, which confirmed the incorporation specificity and fidelity of DiZHSeC (Fig. 2b). Furthermore, we incorporated DiZHSeC at residue V58 in an *E. coli* acid stress chaperone HdeA and found that the resulting protein HdeA-V58DiZHSeC exhibited similar efficiency in photo-affinity capturing of client proteins in *E. coli* periplasm as that of HdeA-V58DiZPK bearing the DiZPK photocrosslinker (Fig. 2c).

Identification of the MS-label by mass spectrometry. To test if the NPAＡ moietys could be utilized as a MS-label for target identification by mass spectrometry, we used the HdeA-F35DiZHSeC monomer to photocrosslink with the WT-HdeA monomer at pH 7 (Supplementary Fig. 4). The crosslinked dimer was then subjected to in-gel oxidative cleavage, trypsin digestion and liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis on a Thermo Velos-Elite Orbitrap mass spectrometer. Searching the MS/MS data with Mascot identified an adduct peptide ([KPVNSWTCD-EFLAVDESFQPTAVGFAEALNNK]2) with the MS-label (C<sub>H</sub>13<sub>N</sub>O) modification (Fig. 3b). The monoisotopic mass of this adducted peptide was observed as 3722.8006 Da (calculated 3722.7927 Da), equal to sum of the mass of NPAＡ (139.0997 Da) plus the mass of the unmodified peptide (3583.6929 Da). The modification site was unambiguously assigned to E37 on HdeA based on the MS/MS spectra (Fig. 3b).
Development of IMAPP. Next, we utilized DiZHSeC to develop a MS-label assisted ‘IMAPP’ strategy to directly capture and identify protein – protein interactions within living cells (Fig. 4a). In brief, the DiZHSeC-incorporated bait protein was expressed in cells and its native interactions with ‘prey’ proteins were photo-captured under living conditions. The photocrosslinked prey–bait complexes were enriched by affinity purification using an epitope tag on the bait protein, separated by SDS-PAGE gel and then subjected to in-gel H$_2$O$_2$-mediated oxidative cleavage and trypsin digestion. The digested peptides were analysed by LC–MS/MS, and the identity of the prey proteins as well as the crosslinking sites were assigned by searching for the specific MS-label (C$_8$H$_{13}$NO) modification on the peptides (Fig. 4a). Because DiZHSeC was installed within the specific interaction region on the bait protein, only the specific interaction partners can be photo-captured with an in situ generated MS-label upon cleavage. Therefore, by only focusing on peptides with a uniform MS-label modification after the IMAPP strategy, the crosslinking peptides can be readily distinguished from the remaining non-specific contaminants even after affinity purification$^{17}$, which significantly increase the confidence for identifying the genuine pool of prey proteins.

Identification of HdeA client proteins using IMAPP. As an essential acid stress chaperone in preserving the periplasmic proteostasis of enteric pathogens under acid stress (for example, pH 1–3 in human stomach), HdeA is able to bind a broad range of acid-susceptible client proteins to prevent them from aggregation$^{8,29}$. Our previous efforts in identifying HdeA-binding partners at acidic conditions were hampered by false target identification due to non-specific protein binding and indirect protein interactions$^{14,19}$. Therefore, we decide to apply the IMAPP strategy to re-evaluate this important protein – protein interaction network during bacterial acid resistance. The HdeA variants carrying DiZHSeC at different positions were confirmed to exhibit similar chaperone activity as WT-HdeA (Supplementary Fig. 5). E. coli cells expressing HdeA-V53DiZHSeC were treated at pH 2.3 for 30 min, irradiated with 365 nm ultraviolet light for 15 min and then subjected to the IMAPP strategy (Supplementary Fig. 6). During the traditional analysis without accounting for MS-label, a hit was assigned when the Mascot search identified at least two unique peptides from a given protein. A total of 967 hits were identified with 767 proteins appearing in at least two of three replicates, among which 402 (52% of the total) are cytosolic proteins (Supplementary Fig. 7). Because HdeA is a periplasmic chaperone that is expected to only interact with the envelope proteins, these cytosolic proteins are obviously false-positive hits, most likely originating from the affinity purification process. In contrast, when using IMAPP analysis that take the MS-label as a further criteria, a hit is assigned only when Mascot search identifies at least two unique peptides from a given protein, including one crosslinking peptide with the MS-label modification and another non-crosslinking peptide without MS-label modification. A total of 71 proteins were identified under this condition from E. coli K12 proteome, with 52 proteins identified in at least two of three replicates (Fig. 4b). These 52 proteins were all covered by the aforementioned 767 proteins identified by the traditional analysis. No hits containing MS-label were found in the control groups that were prepared either without ultraviolet irradiation or with WT-HdeA as the bait protein for ultraviolet irradiation (Supplementary Fig. 6). Among the 52 IMAPP hits, 50 (96%) are bacterial ‘envelope’ proteins (located in periplasm as well as outer and inner membrane) while only 2 (4%) are cytosolic proteins (Fig. 4c). Therefore, in comparison with traditional analysis, the false-positive rate can be dramatically decreased with our IMAPP strategy. In addition, by focusing on the MS-label modified peptides, IMAPP strategy avoids the tedious peptide comparison between the experimental and control groups.

It is worth mentioning that the 50 envelope proteins identified by IMAPP strategy are not simply a reflection of protein abundance present in the periplasmic extract (Supplementary Table 1). This data indicates that HdeA indeed has its own client...
specify when protecting clients from acid stress, with the underlying mechanism remaining elusive. The information regarding these 50 envelope clients as well as the photo-captured peptides carrying the MS-label are listed in Supplementary Tables 2 and 3. Twenty-two clients in this list have been reported before, including DegP and SurA, two essential periplasmic protein quality control (PQC) factors that have been previously postulated that our IMAPP strategy could be further utilized to address how crosslinking peptides may be protected by HdeA within E. coli periplasm under acid stress. Other newly discovered HdeA clients by our IMAPP method include inner membrane lipoproteins such as YfHm (α-macroglobulin), transport proteins such as YtfQ (ABC transporter periplasmic-binding protein) as well as other functional proteins such as Ecotin (a serine protease inhibitor). To further validate these MS results, we randomly chose four candidates (Tsp, BglX, DsaA, AspG2) and unambiguously detected their interaction with HdeA at pH 2 by fluorescence anisotropy (Supplementary Fig. 8), which confirmed the reliability of our IMAPP strategy for profiling intracellular protein–protein interactions.

Finally, as a further control, we applied the IMAPP strategy on the DiZHSeC-bearing HdeA at neutral pH, under which condition HdeA only binds to its dimer partner but not the client proteins. Photocrosslinking of the HdeA variant containing DiZHSeC at residue F35 on its dimer interface (HdeA-F35DiZHSeC) at pH 7 only yielded three candidate proteins, with HdeA itself being the dominant hit (Supplementary Fig. 9). Taken together, our IMAPP strategy offers a straightforward approach for highly efficient and confident identification of intracellular protein–protein interaction partners.

**Mapping HdeA dimer interface using IMAPP.** Identification of the crosslinking sites is highly valuable for studying protein interaction interface. However, traditional crosslinking experiments often require intensive software development to deconvolute the complicated MS spectra in order to identify such crosslinking peptides. In contrast, this technical challenge can be addressed by using our photocrosslinker DiZHSeC. We postulated that our IMAPP strategy could be further utilized to map the interface of protein–protein interactions via these assigned crosslinking sites. For proof-of-concept, we first mapped HdeA dimer interface by photocrosslinking between the DiZHSeC-containing HdeA monomer and WT-HdeA monomer at pH 7. The crosslinking sites between WT-HdeA and HdeA-F35DiZHSeC monomer were identified by IMAPP as residues E37, D43, K44, DAVLD (47–51) and W82 (Supplementary Table 4 and Supplementary Fig. 10c). Based on the crystal structure of HdeA dimer (PDB: 1D8J), the distances between F35 and these identified crosslinking residues (measured from C(α) of F35 to the
Mapping HdeA–DegP interaction interface using IMAPP. Furthermore, we directly mapped a previously unknown protein–protein interaction interface using IMAPP. DegP is an essential PQC factor with dual protease and chaperone functions, and it contains one protease domain and two PDZ domains (PDZ1 and PDZ2; refs 14,34). Our previous photocrosslinking experiments using DiZPK did not yield any information regarding the HdeA–DegP interaction interface on DegP side. According to our previous study, HdeA interacts mainly with its client proteins through two hydrophobic regions (Supplementary Fig. 16a; ref. 14). We thus incorporated DiZHSeC at residues T31, L39, V49, V58 (within these two hydrophobic regions) or residue A6 (in its N-terminal hydrophilic region) as a control, and performed photocrosslinking between these HdeA variants and DegP-S210A (the catalytic dead mutant of DegP to avoid self-proteolysis) under acidic conditions. In agreement with the previous study, the crosslinking results showed that HdeA interacts with DegP mainly through its two hydrophobic regions (Supplementary Fig. 16a). The crosslinked complexes were further subjected to IMAPP strategy to analyse the crosslinking sites on DegP-S210A. The results indicated that HdeA directly interacted with DegP’s protease domain and PDZ1 domain, but not the PDZ2 domain (Supplementary Fig. 16b–g, and Fig. 5b). We further applied the IMAPP strategy in living E. coli cells expressing HdeA.

Probing the interaction dynamic change using IMAPP. In addition to static protein–protein interactions, we expected that our IMAPP strategy could also be applied to probe dynamic conformational changes that frequently occur at protein–protein interaction interfaces. Because HdeA is known to display pH-dependent conformational change28,29,33, we performed the IMAPP experiment at multiple pH conditions in order to capture such dynamic changes at the HdeA dimer interface (Supplementary Fig. 14). Upon acidification (pH <3), HdeA-F35DiZHSeC failed to crosslink with W82 in the peptide 78VKG EWDK84 located at the C-terminus of HdeA and this result is consistent with a structural model in which acid triggers the opening of the C terminal region of HdeA33. In addition, the original crosslinking site E37 on the peptide 11KPVNSWTCEDFLAVDESFPQTDVGFAEALNNK12 at pH 7 became spread to more residue positions within the peptide, indicating a potential order-to-disorder transition within this region when the environmental pH drops from 7 to 2 (Supplementary Fig. 15).

Figure 5 | Mapping protein–protein interaction interface using IMAPP. (a) Mapping of HdeA dimer interface using the IMAPP strategy. DiZHSeC was incorporated at different sites (F28, T31 and L39), respectively, on HdeA to photocrosslink with WT-HdeA at pH 7. The incorporation sites of DiZHSeC (coloured in magentas) and the crosslinking sites (coloured in yellow) are displayed on the crystal structure of HdeA (PDB: 1D8J; ref. 28). The bait HdeA is coloured in cyan and the crosslinked HdeA is coloured in green. Close-up view of the crosslinking interface is shown in the left. Crosslinked HdeA monomer is shown as a surface representation with crosslinking residues coloured in yellow and the other residues coloured in green. (The representative result from three replicates is shown). (b) Interaction interface of HdeA-DegP mapping through the photocrosslinking sites identified by IMAPP strategy. The crystal structure of DegP (PDB: 3MH4; ref. 34) contains a protease domain (coloured in blue) and two PDZ domains (colour in yellow and green, respectively). Based on the integrated crosslinking sites identified by IMAPP (coloured in magenta), the HdeA–DegP interface can be mapped to the protease domain and the PDZ1 domain (The representative result from two replicates is shown). (c) Multiple HdeA may bind to a single DegP molecule under acidic conditions. HdeA-V58DiZHSeC or WT-HdeA was used to photocrosslink with DegP-S210A and the crosslinked complexes were analysed by immunoblot. The crosslinked complex with a ratio of 1:1(HdeA/DegP) binding is marked with a black arrow. Crosslinked complexes with higher HdeA/DegP binding stoichiometries are marked with red arrows. (The representative result from three replicates is shown). (d) A proposed model illustrating that HdeA may interact with different regions on DegP while multiple HdeA chaperone molecules may also simultaneously bind to the same DegP molecule under acid stress.
V58DiZHSeC and searched for the MS-label modified DegP peptides. Consistent with our *in vitro* data, the *in vivo* experiments also identified several crosslinking sites within the protease domain (Supplementary Table 3), suggesting that binding of HdeA at this region could be essential for protecting DegP from acid-induced damage. Furthermore, crosslinked complexes with the molecular weight higher than the 1:1 ratio complex (HdeA:DegP) were observed (Supplementary Fig. 17, and Fig. 5c). This result, together with the fact that the crosslinking sites and locations of DegP and HdeA are different within the RH1 domain of DegP while multiple HdeA chaperone molecules may also simultaneously bind to the same DegP molecule under acid stress (Fig. 5d). The structural basis for this intriguing domain-recognition specificity and the binding stoichiometry of HdeA towards DegP remains to be further verified.

**Expanding IMAPP to mammalian cells.** Finally, as the Pyl-based genetic code expansion system has been widely adapted for various prokaryotic and eukaryotic systems, we expect that our IMAPP strategy can be applied to identify protein–protein interactions in mammalian cells. As a proof-of-concept, we chose to study the interaction between RhoA and RTKN as well as their binding interface. RhoA is a Rho family small GTPase that has important roles in regulating actin cytoskeleton as well as cell adhesion. As an effector protein for RhoA, RTKN can specifically interact with activated RhoA through its Rho-binding domain 1 (RH1 domain), which can inhibit the GTPase activity of RhoA. To illustrate this specific and important interaction interface, especially in the unbiased global proteomics strategy for identification of interaction partners as well as simultaneous mapping of binding networks under living conditions. To our delight, RTKN-L45DiZHSeC can efficiently capture its RhoA partner through a similar binding interface as that on PKN. This observation is also in agreement with the expected region known to be involved for contacting RhoA (Fig. 18b). Furthermore, the crosslinking sites on RhoA were identified RhoA as the dominant hit (91% of relative abundance as calculated by NASF value) (Supplementary Fig. 18a), and the subsequent IMAPP analysis identified RhoA as the dominant hit (~91% of relative abundance as calculated by NASF value) (Supplementary Fig. 18b). Furthermore, the crosslinking sites on RhoA were identified as residues T66/D67/R68 that are located in the expected region known to be involved for contacting RhoA effector proteins such as PKN (Supplementary Fig. 18c; refs 41, 42). This observation is also in agreement with the previous report showing that RTKN may interact with RhoA through a similar binding interface as that on PKN. Taken together, these data further demonstrated the capability of our IMAPP strategy for high-confidence identification of protein–protein interaction as well as simultaneous mapping of binding interface in various living species.

**Discussion**

Coupling chemical crosslinking or photocrosslinking reagents with mass spectrometry has become a valuable tool for identifying protein–protein interactions, particularly under living conditions. Chemical crosslinking is a well-developed strategy for identifying interaction partners as well as interaction interface, especially in the unbiased global proteomics study without the need for protein engineering. However, such strategies suffer from the following limitations, especially in living systems: (i) the restriction to only a small set of chemically reactive amino acids, which may miss some chemically reactive information when such residues are lacking; (ii) low efficiency under extreme conditions, for example, certain chemical crosslinkers such as disuccinimidyl suberate are not compatible with acidic conditions below pH 4.5; (iii) complicated crosslinking results that often include both intermolecular and intramolecular complexes, which requires intensive software development to decipher.

By bearing a photo-labile moiety that can be site-specifically incorporated into proteins of interest, the genetically encoded photocrosslinkers allow facile capture of transient protein–protein interactions with high spatiotemporal resolution in living cells, which can largely overcome the aforementioned limitations for chemical crosslinking. Nevertheless, the downstream process after photocrosslinking still falls into the regular affinity purification procedure that suffers from false-positive identifications due to high contamination backgrounds. Laborious optimization of purification procedures including the washing condition, the design of tandem purification protocol, as well as tedious comparison with the control groups are usually needed in order to remove these contaminants, and the effectiveness of such efforts are highly variable in different cases. Moreover, certain protein–protein complexes may still be difficult to be effectively removed, especially for those sticky indirect binders. In addition, the information regarding protein interaction interface is particularly lacking, mainly due to the difficulties in deciphering crosslinking peptides and sites with traditional photocrosslinkers.

To overcome these limitations, we have developed a selenium-based, genetically encoded photo-affinity probe DiZHSeC that contains a transferable MS-label, allowing simultaneous identification of protein–protein interactions and mapping the corresponding contacting interfaces under living conditions. Taking the *in situ* transferred MS-label as an internal criterion, IMAPP allows high-confidence target identification from complicated proteome backgrounds without tedious optimization of purification procedure or time-consuming comparison with the control groups. Although the utilization of MS-label as the searching criterion may miss certain hits due to the relatively low abundance of modified peptides in the context of non-modified peptides during MS analysis, we expect that such a limitation can be addressed by introducing an affinity tag to the current photocrosslinking system that enables the enrichment of the MS-label-modified peptides.

Notably, the exact crosslinking sites can be readily identified from the MS data through IMAPP analysis, permitting simultaneous illustration of protein–protein interaction interface. Albeit being a relatively low-resolution method in mapping such interfaces, this approach is highly valuable in providing dynamic and real-time structural information regarding protein interaction networks under living conditions. To our knowledge, there are currently no genetically encoded photocrosslinking probes that are able to achieve these multiple challenging goals simultaneously. In particular, as exemplified by our study on the acid chaperone HdeA with its client DegP here, this strategy is especially well-suited for studying protein–protein interactions that involve conditionally or intrinsically disordered proteins which remain almost inaccessible by other conventional structural methods such as X-ray crystallography.

**Methods**

**Plasmids construction.** The plasmid pBAD-AR-MbpYIRBS encoding the mutant MBpYIRBS and its cognate rRNA<sub>CUA</sub><sup>Pyl</sup> (MbpYIRBS recognizes and transfers the UAA to tRNA<sub>CUA</sub> Pyl, which inserts the UAA into the in-frame amber code site on target genes) for DiZHSeC in *E. coli* cells were described in the previous literature. The plasmids encoding the GFP, HdeA, DegP, S210A or their mutant variants (carrying a C-terminal His-tag) and the plasmid encoding the WT-HdeA carrying no tag on its C-terminus were described in previous literature. The plasmid pBAD-HdeA-A67TAG-His<sub>6</sub>, and pBAD-HdeA-F28TAG-His<sub>6</sub>, pBAD-HdeA-T31TAG-His<sub>6</sub>, pBAD-HdeA-L39TAG-His<sub>6</sub>, pBAD-HdeA-V49TAG-His<sub>6</sub> encoding the mutant HdeA...
(carrying a C-terminal His tag) were generated using site-directed mutation on the plasmid pBAD-HdeA-His6. The Tsp sequence was amplified from the genome of E. coli (DH10B) by PCR with primers 5'-CAGTCGAGCTGTTCTTTTGTATTCG (5'-primer) and 3'-CCGGAATTCGATCTTTTGTATTCG (3'-primer). Then, the insert was cloned into pNeo/EcoR1 sites on a pBAD-myc-His/A vector to produce the final plasmid pBAD-AspGl-His6. The AspG2 sequence was amplified from the genome of E. coli (DH10B) by PCR with primers 5'-CAGTCGAGCTGTTCTTTTGTATTCG (5'-primer) and 3'-CCGGAATTCGATCTTTTGTATTCG (3'-primer). Then, the insert was cloned into pNeo/EcoR1 sites on a pBAD-myc-His/A vector to produce the final plasmid pBAD-AspG2-His6.

Reagents and equipments. Compounds used in the synthesis of DiZHSec were purchased from J&K Scientific, Aladdin or Alamine. All chemicals used in this study were analytical grade or above. Primary antibody: anti-His antibody (ZSBG-bio, 1:2,000 diluted). Horseradish peroxidase-linked secondary antibodies were purchased from Coolsinglialing Technology (1:5,000 diluted).

1H NMR and 13C NMR spectra were recorded on a Bruker-500 MHz NMR instrument. Full image of the spectra in the main paper are available in Supplementary Fig. 25. Circular dichroism spectroscopy measurement was performed on a J-815 CD spectrometer (JASCO). Fluorescence anisotropy experiment and light scattering experiment were performed on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies).

The plasmid pCMV-MbPyrRS encoding the constitutive active RhoA14V was generated using the site directed mutation on the plasmid pCMV-RhoA-myc. pBAD-myc-His/A vector.

The synthesis route for DiZHSeC is described in Supplementary Fig. 1. The method for synthesis of the target compounds in this article; see Supplementary Figs 19–24. The synthesis route for DiZHSeC was carried out with a BHE300 C4 Acquity column (1.7 mm, 2.1 × 100 mm), and the mobile phase was chosen for ES-MS analysis. Total mass of the protein was calculated by Masslynx V4.1 software (Waters). Theoretical mass of the protein was calculated using PROTEIN CALCULATOR v3.3 (http://www.scripps.edu/~cddutnam/protcalc.html) and theoretical mass of the modified protein was adjusted manually.

In vivo crosslinking of HdeA dimmers. E. coli (DH10B) cells expressing HdeA-F35DiZHSeC or HdeA-F35DiZPK (carrying a C-terminal His-tag) were harvested at 4,000 rpm for 10 min. The supernatant was discarded, and the bacterial pellet was suspended in LB buffer at pH 7. Then the sample was irradiated by ultraviolet for 15 min with UVP CL-1000 ultraviolet Cross-Linker installed with 365 nm ultraviolet lamps (Hitachi FT85/Black Light, 8-watt) at a distance of 5 cm (~2 mw cm⁻²) on ice. The crosslinked dimer was purified according to the method mentioned above.

Fluorogenic labelling. The in vivo crosslinked HdeA (carrying a C-terminal His-tag) dimmers (pH 7) using DiZHSec or DiZPK as the photoscissor (or the non-crosslinked bivalent protein HdeA-F35DiZHSeC, carrying a C-terminal His-tag) (1 mg ml⁻¹) were denatured by 1% SDS and incubated with 8 mM H2O2 in PBS (pH 8.0) at 37 °C for 3 h. After removal of the remaining H2O2 by desalting with Micro Bio-spin 6 columns (BIO-RAD), the solution was incubated with 200 nM Tet probe upon ultraviolet irradiation for 5 min. Then the proteins were separated by SDS-PAGE gel, and followed by coomassie blue staining or in-gel fluororescence analysis.

**Characterization of newly synthesized compounds.** The synthesis route for DiZHSeC is described in Supplementary Fig. 1. The method for synthesis is described in Supplementary Methods. For NMR analysis and HRMS analysis of the compounds in this article; see Supplementary Figs 19–24.

**Compound 2.** 1H NMR (500 MHz, CDCl₃); δ 3.64 (t, J = 6.6 Hz, 2H), 2.32–3.25 (m, 2H), 2.78 (t, J = 6.6 Hz, 2H), 1.39–1.45 (m, 4H), 1.02 (s, 3H); 13C NMR (125 MHz, CDCl₃); δ 71.05 (m, 2H), 39.08, 31.67, 27.71, 25.46, 24.12, 19.75. HRMS (m/z): [M + H]⁺ calculated for C₆H₁₂N₂BrNO₂, 248.03985; found 248.03903.

**Compound DiZHSeC.** 1H NMR (500 MHz, CDCl₃); δ 4.14 (t, J = 6.4 Hz, 1H), 3.18 (t, J = 6.6 Hz, 2H), 2.85 (t, J = 7.1 Hz, 2H), 2.75–2.79 (m, 2H), 2.25–2.33 (m, 2H), 1.40–1.42 (m, 4H), 1.09 (s, 3H); 13C NMR (125 MHz, CDCl₃); δ 173.54, 170.66, 52.43, 45.43, 38.79, 36.33, 31.30, 31.13, 25.05, 23.40, 18.52, 17.97, 17.77. HRMS (m/z): [M + H]⁺ calculated for C₆H₁₂N₂O₂Se, 351.09354; found 351.09232.

**Expression of UAA-incorporated proteins in E. coli.** Expression of UAA-incorporated proteins was carried out in DH10B cells co-transformed with plasmids carrying both the MbPyrRS mutants/mtRNA⁺CUA 5’/3’ pair and the target protein gene with an in-frame amber codon on the incorporation site. The overnight cultured bacterial cultures harboring the two plasmid-bearing DH10B cells were diluted with fresh LB medium containing 50 mg l⁻¹ ampicillin and 34 mg l⁻¹ chloramphenicol. The bacteria were then grown at 37 °C to an OD₆₀₀ ~ 0.5 before
The crosslinked complex was separated by the SDS–PAGE gel followed by coomassie blue staining.

In vivo photo-crosslinking of HdeA with its client proteins. E.coli (DH10B) cells expressing Hda-A-V5/DiZHSeC were harvested at 4,000 r.p.m. for 10 min. The supernatant was discarded, and the bacterial pellet was suspended in LB buffer. The pH was adjusted to 2.3 using 5 M HCl, and the cells were incubated at 37 °C for 30 min, followed by ultraviolet irradiation for 15 min. The pH of the solution was then recovered to 7 using 5 M NaOH and the crosslinked prey–bait complexes were purified according to the method mentioned above.

In-gel cleavage and digestion in IMAPP strategy. The crosslinked prey–bait complexes were separated by the SDS–PAGE gel, and the corresponding protein bands were excised and cut into pieces. The gel pieces were dehydrated in acetonitrile, then incubated in buffer I (10 mM dithiothreitol, 50 mM ammonium bicarbonate) at 36 °C for 30 min, and were further incubated in buffer II (55 mM iodoacetamide, 50 mM ammonium bicarbonate) at ambient temperature for 1 h in the dark before being dehydrated. The samples were then incubated in 8 M H2O2 in PBS (pH 8.0) at 37 °C for 3 h, followed by washing with H2O for three times, and dehydrated again. Then the samples were in-gel digested with sequencing grade trypsin (5 ng · μL–1 trypsin, 50 mM ammonium bicarbonate, pH 8.0) overnight at 37 °C. The resulting peptides were extracted twice with 5% formic acid/50% acetonitrile in water, and then vacuum-centrifuged to dryness.

LC-MS/MS analysis in IMAPP strategy. To analyse the in vivo crosslinked protein complexes, the corresponding protein bands were excised and divided into six samples and, respectively, subjected to the in-gel cleavage and digestion procedure mentioned above. The extracted peptides were reconstituted in 0.2% formic acid, loaded onto a 100 μm × 2 cm pre-column and separated on a 75 μm × 20 cm capillary column both of which were packed with 4 μm C18 bulk materials (InnosepBio, China). A EASY nLC 1000 system (Thermo Scientific, USA) was used to generate the following HPLC gradient: 7–35% B in 40 min, 35–75% B in 4 min, then held at 75% B for 20 min (A = 0.1% formic acid in water, B = 0.1% formic acid in acetonitrile). The eluted peptides were sprayed into an LTQ-Orbi-trap-Elite (Thermo Scientific, USA) equipped with a nano-EIS source. The mass spectrometer was operated in data-dependent mode with one MS scan in FT mode at a resolution of 30,000 followed by 10 CID (Collision Induced Dissociation) MS/MS scans in the ion trap for each cycle.

Data analysis in IMAPP strategy. Raw data files produced in the Xcalibur software (Thermo Scientific) were transferred to mgf files through MScanX and then searched with Mascot V.2.3.02 (Matrix Science) against Swiss-Prot 57.15 (515,203 sequences; 181,334,896 residues) E. coli database (22,466 sequences). Searches were performed with a precursor mass tolerance set to 7 p.p.m., fragment mass tolerance set to 0.6 Da and a maximum number of missed cleavages set to 3. In addition to the regular cysteine carbamidomethylation (C4-HN=O, 57.0215 Da), methionine oxidation (O, 15.9949 Da) and cysteine carboxamidomethylation (C4-HN=O, 163.1614 Da), the following extra variable modifications were defined in the search in order to account for all possible scenarios of MS/MS label modifications: 1. MS-label modification on each of all 20 amino acids (C8H13NO2, 155.0946 Da). 2. MS-label modification on methionine or cysteine (C8H13NO2, 196.1212 Da). 3. MS-label modification on cysteine (C2H3NO2, 73.0164 Da), the following extra variable modifications were added: 1. MS-label modification on carbamidomethylated cysteine (C2H3NO2, 98.0644 Da). 2. MS-label modification on carbamidomethylated methionine (C4-HN=O, 103.0556 Da). 3. MS-label modification on carbamidomethylated methionine (C4-HN=O, 103.0556 Da). 4. MS-label modification on carbamidomethylated methionine (C4-HN=O, 103.0556 Da). The resulting peptides were extracted twice with 5% formic acid/50% acetonitrile in water, and then vacuum-centrifuged to dryness.

The search was performed using Mascot V.2.3.02 (Matrix Science) against Swiss-Prot 57.15 (515,203 sequences; 181,334,896 residues) Homo sapiens (human) database (20,266 sequences). The peptide ion score threshold was set according to the score distribution (P value < 0.05, E value < 0.05). A protein was assigned as a ‘hit’ when the Mascot search identifies at least two unique peptides, including one crosslinking peptide with MS-label modification and another non-crosslinking peptide without MS-label.

Data availability. The authors declare that the data supporting the findings of this study are available within the article and its Supplementary Information files. Any further relevant data concerning the techniques used in the paper are available from P.R.C or C.W. on request.

References

1. Berggard, T., Linse, S. & James, P. Methods for the detection and analysis of protein–protein interactions. Proteomics 7, 2833–2842 (2007).
2. Gingeras, A.-C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. Nat. Rev. Mol. Cell Biol. 8, 645–654 (2007).
3. Sinz, A. Investigation of protein–protein interactions in living cells by chemical crosslinking and mass spectrometry. Anal. Bioanal. Chem. 397, 3433–3440 (2010).
4. Pham, N. D., Parker, R. B. & Kohler, J. J. Photoscrosslinking approaches to interactome mapping. Curr. Opini. Chem. Biol. 17, 90–101 (2013).
5. Preston, G. W. & Wilson, A. J. Photo-induced covalent cross-linking for the analysis of biomolecular interactions. Chem. Soc. Rev. 42, 3289–3301 (2013).
6. Winnacker, M., Breeger, S., Strasser, R. & Carrell, T. Novel diazirine-containing DNA photofragment probes for the investigation of DNA-protein interactions. Chembiochem. 10, 109–118 (2009).
7. Yang, L. & Zhang, H. C. Bifunctional fatty acid chemical reporter for analyzing protein–protein protein interactions with a genetically encoded photo-crosslinking amino Acid. Chembiochem. 12, 1854–1857 (2011).
8. Hino, N. et al. Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photocleavable amino acid. Nat. Methods 2, 201–206 (2005).
9. Mori, H. & Ito, K. Different modes of SecY–SecA interactions revealed by site-directed in vivo photo-cross-linking. Proc. Natl Acad. Sci. U.S.A. 103, 16159–16164 (2006).
10. Freinken, E., Chng, S.-S. & Kahne, D. The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. Proc. Natl Acad. Sci. U.S.A. 108, 2486–2491 (2011).
11. Con, I. et al. Genetically encoded chemical probes in cells reveal the binding mode of urocortin 1 to CRF-B receptor. Cell 153, 1258–1269 (2013).
12. Zhang, M. et al. A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. Nat. Chem. Biol. 7, 671–677 (2011).
13. Hino, N. et al. Genetic incorporation of a photo-crosslinkable amino acid reveals novel protein complexes with GRB2 in mammalian cells. J. Mol. Biol. 406, 343–353 (2011).
14. Wu, H. et al. A peptide aldehyde microarray for high-throughput profiling of cellular events. J. Am. Chem. Soc. 133, 1946–1954 (2011).
15. Dunham, W. H., Mullin, M. & Gingras, A.-C. Affinity-purification coupled to mass spectrometry: basic principles and strategies. Proteomics 12, 1576–1590 (2012).
16. Cravatt, B. F., Wright, A. T. & Kozaich, J. W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. Annu. Rev. Biochem. 77, 383–414 (2008).
17. Lin, S. et al. Genetically encoded cleavable protein–photo-cross-linker. J. Am. Chem. Soc. 136, 11860–11863 (2014).
18. Yang, L. et al. In vivo application of photocleavable protein interaction reporter technology. J. Proteome. Res. 11, 1027–1041 (2012).
19. Palaniappan, K. K. et al. Isotopic signature transfer and mass pattern prediction (isostamp): an enabling technique for chemically-directed proteomics. ACS Chem. Biol. 6, 829–836 (2011).
20. Pfaffenberger, T. et al. Tet oxidizes thyamine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat. Chem. Biol. 10, 574–581 (2014).
21. Hayashi, T., Yasueda, Y., Tamura, T., Takaoka, Y. & Hamachi, I. Analysis of cell-surface receptor dynamics through covalent labeling by catalyst-tethered antibody. J. Am. Chem. Soc. 137, 5372–5380 (2015).
22. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. Annu. Rev. Biochem. 79, 413–444 (2010).
23. Takimoto, J. K., Dallas, N. J., Bouthenon, J. P. & Wang, L. Stereobchemical basis for engineered pyrrolyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. ACS Chem. Biol. 6, 733–743 (2011).
26. Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* 83, 379–408 (2014).

27. Yu, Z. & Lin, Q. Design of spiro[2.3]hex-1-ene, a genetically encodable double-strained alkenes for superfast photoclick chemistry. *J. Am. Chem. Soc.* 136, 4153–4156 (2014).

28. Gajiwala, K. S. & Burley, S. K. HDEA, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. *J. Mol. Biol.* 295, 605–612 (2000).

29. Hong, W. et al. Periplasmic protein HdeA exhibits chaperone-like activity exclusively within stomach ph range by transforming into disordered conformation. *J. Biol. Chem.* 280, 27029–27034 (2005).

30. Weski, J. & Ehrmann, M. Genetic analysis of 15 protein folding factors and proteases of the *Escherichia coli* cell envelope. *J. Bacteriol.* 194, 3225–3233 (2012).

31. Rinner, O. et al. Identification of cross-linked peptides from large sequence databases. *Nat. Methods* 5, 315–318 (2008).

32. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nat. Methods* 9, 904–906 (2012).

33. Ahlstrom, L. S., Law, S. M., Dickson, A. & Brooks lii, C. L. Multiscale Modeling of a Conditionally Disordered pH-Sensing Chaperone. *J. Mol. Biol.* 427, 1670–1680 (2015).

34. Krojer, T., Sawa, J., Huber, R. & Clausen, T. HtrA proteases have a conserved activation mechanism that can be triggered by distinct molecular cues. *Nat. Struct. Mol. Biol.* 17, 844–852 (2010).

35. Lin, S. et al. Site-specific engineering of chemical functionalities on the surface of live hepatitis D virus. *Angew. Chem. Int. Ed.* 52, 13970–13974 (2013).

36. Kaibuchi, K., Kuroda, S. & Amano, M. Regulation of the cytoskeleton and cell adhesion by the Rho family GTPases in mammalian cells. *Annu. Rev. Biochem.* 68, 459–486 (1999).

37. Reid, T. et al. Rhotekin, a new putative target for Rho bearing homology to a serine/threonine kinase, PKN, and rhophilin in the Rho-binding domain. *J. Biol. Chem.* 271, 13556–13560 (1996).

38. Fu, Q. et al. Molecular cloning, expression characterization, and mapping of a novel putative inhibitor of Rho GTase activity, RTKN, to D2S145–D2S286. *Genomics* 66, 328–332 (2000).

39. Ihara, K. et al. Crystal structure of human RhoA in a dominantly active form complexed with a GTP analogue. *J. Biol. Chem.* 273, 9656–9666 (1998).

40. Wong, K.-W., Mohammadi, S. & Isberg, R. R. Disruption of RhoGDI and RhoA regulation by a Rac1 specificity switch mutant. *J. Biol. Chem.* 281, 40379–40388 (2006).

41. Maesaki, R. et al. The structural basis of Rho effector recognition revealed by the crystal structure of human RhoA complexed with the effector domain of PKN/PKR1. *Mol. Cell* 4, 793–803 (1999).

42. Dworsky, R., Blumenstein, L., Vetter, I. R. & Ahmadian, M. R. Structural Insights into the interaction of ROCK1 with the switch regions of RhoA. *J. Biol. Chem.* 279, 7098–7104 (2004).

43. Strambio-de-Castilla, C., Tettenbaum-Novatt, J., Imai, B. S., Chait, B. T. & Rout, M. P. A method for the rapid and efficient elution of native affinity-purified protein A tagged complexes. *J. Proteome. Res.* 4, 2250–2256 (2005).

44. Rigaut, G. et al. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032 (1999).

45. Trinkle-Mulcahy, L. Resolving protein interactions and complexes by affinity purification followed by label-based quantitative mass spectrometry. *Proteomics* 12, 1623–1638 (2012).

46. Selbach, M. & Mann, M. Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK). *Nat. Methods* 3, 981–983 (2006).