

Electronic Supplementary Information

Genetic incorporation of 1,2-aminothiol functionality for site-specific protein modification via thiazolidine formation†

Xiaobao Bi, Kalyan Kumar Pasunooti, Ahmad Hussen Tareq, John Takyi-Williams and Chuan-Fa Liu*

Division of Structural Biology and Biochemistry, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551

*Corresponding author: cfliu@ntu.edu.sg

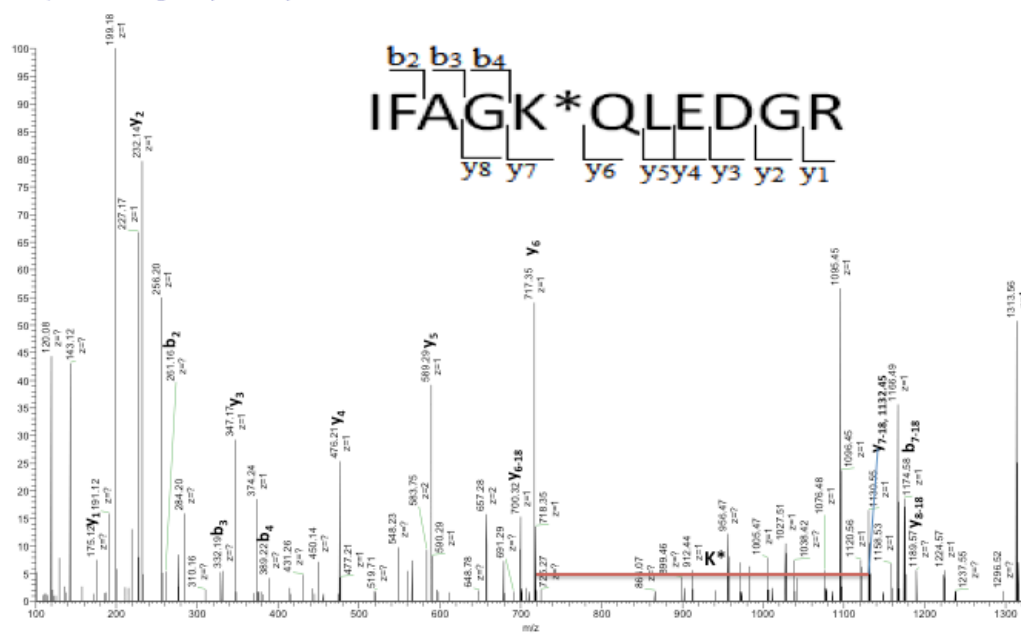
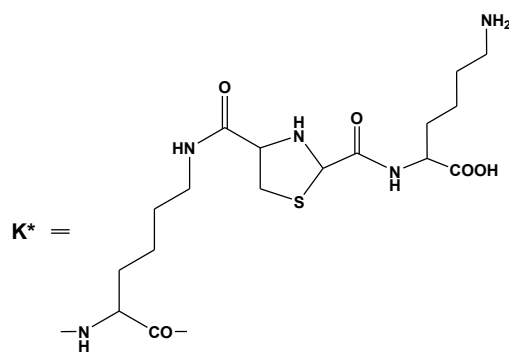


Fig S1. Tandem MS spectrum of IFAGK*QLEDGR from Ubiquitin 5.



Ubiquitin **5** was treated with trypsin and the digested peptide mixture was analyzed with nanoLC-MS/MS on the Q Exactive mass spectrometer. The tryptic peptide **IFAGK*QLEDGR** which contains the modification was identified by its fragment ion spectrum. Note that trypsin digestion also cleaved the K-K(biotin) peptide bond in the labeling reagent.

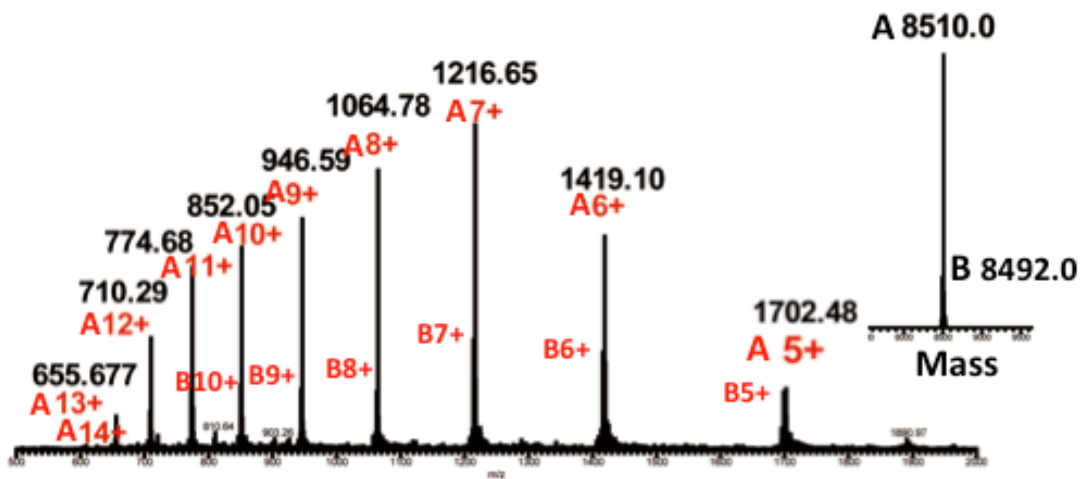


Fig S2. ESI-MS analysis of ubiquitin 6. Expected mass of non-hydrated form of ubiquitinR74-aldehyde is 8491.87, expected mass of hydrated form is 8509.87; Observed mass of peak A (ab. 70%, hydrated form) is 8510, observed mass of peak B (ab. 30%, non-hydrated form) is 8492.0.

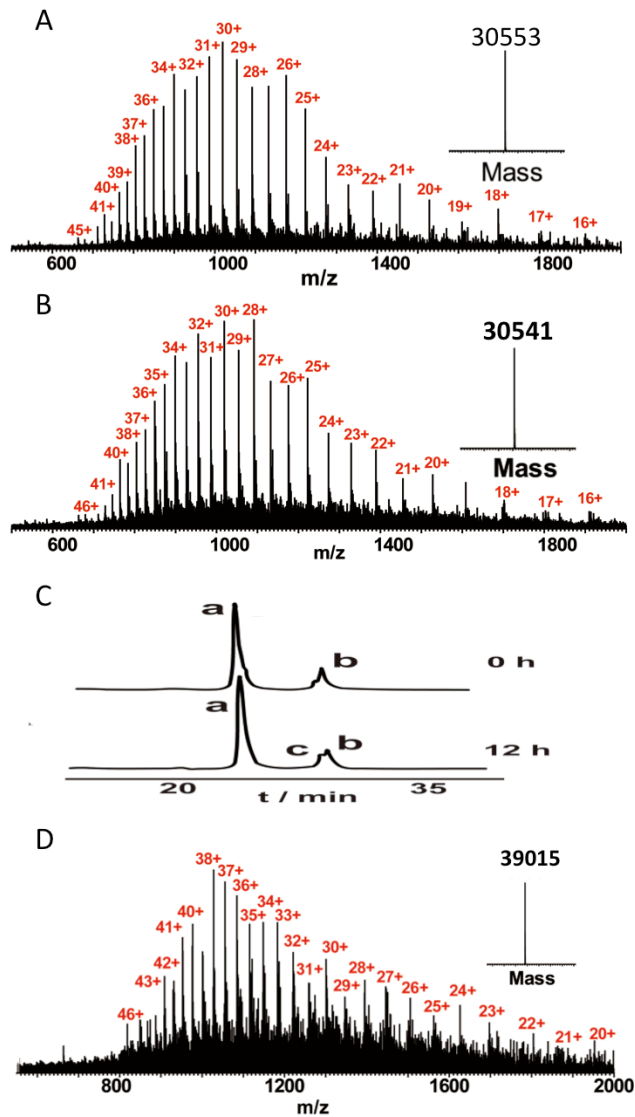


Fig S3. (A) ESI-MS analysis of yeast Rpn10 incorporating ThzK. The initiating Met was cleaved during protein expression. Expected mass of Rpn10-ThzK: 30554.27; Observed mass: 30553; (B) ESI-MS analysis of yeast Rpn10-ThzK after thiazolidine ring deprotection (Rpn10-aminothiols). Expected mass: 30542.26; Observed mass: 30541. (C) C4 analytic HPLC analysis of ubiquitination of Rpn10 at K99 via thiazolidine ligation. Peak a: ubiquitinR74-aldehyde; Peak b: Rpn10-aminothiols; Peak C: desired product, ubRpn10. (D) ESI-MS analysis of ubRpn10. Expected mass: 39016.13; Observed mass: 39015.

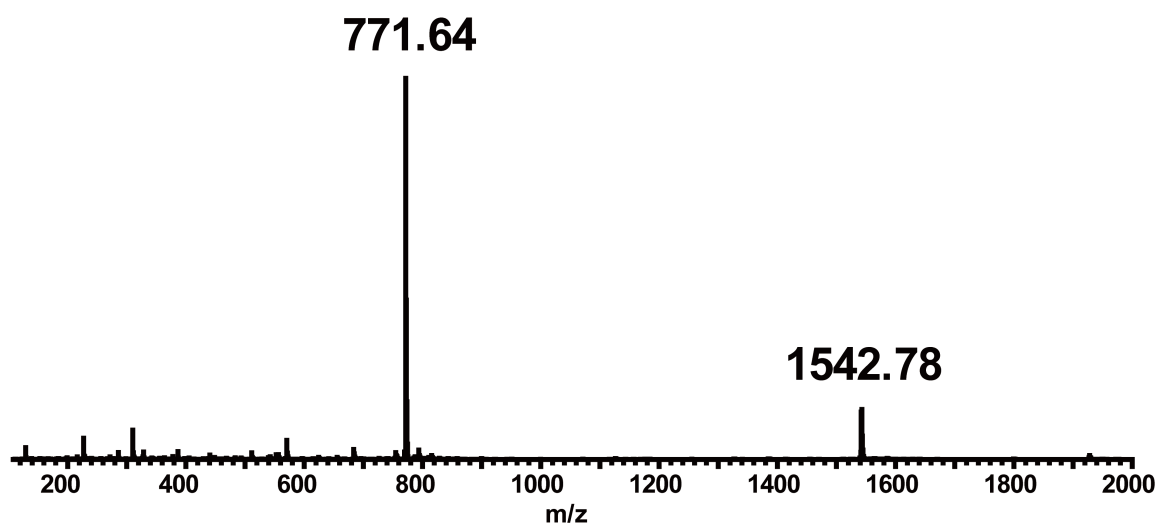


Fig S3. ESI-MS analysis of H-SKK(biotin)LA-NH₂ peptide. Expected mass: 770.45; Observed mass: 771.64.

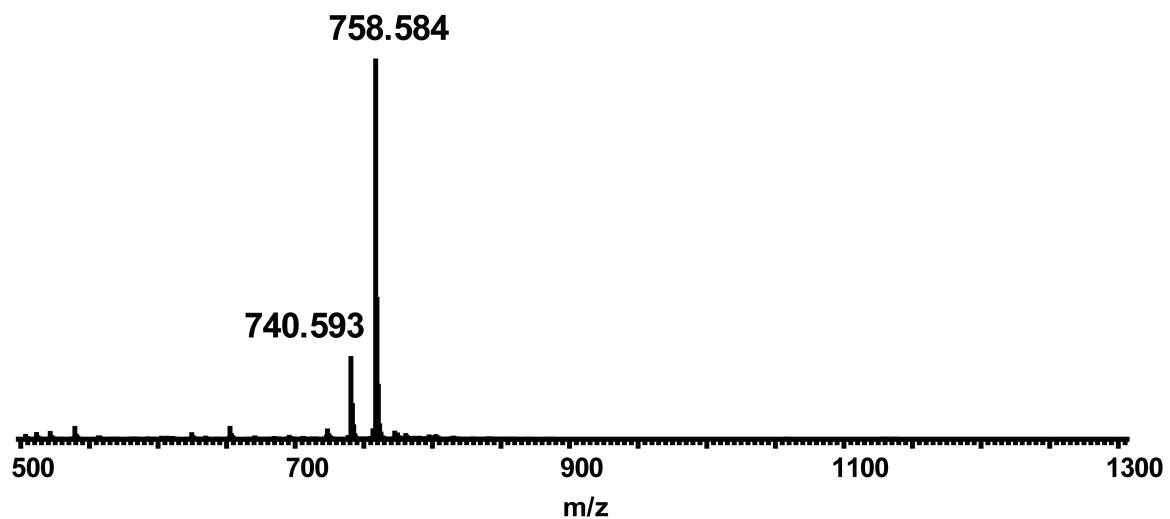


Fig S5. ESI-MS analysis of aldehyde functionalized peptide: $\text{O}=\text{CHCO}\text{-KK}(\text{biotin})\text{LA-NH}_2$ peptide. Expected mass: 739.41; Observed mass: 740.593 and 758.584 (hydrated form).

Materials and Reagents

Amino acids, coupling reagents and resins were obtained from chemimpex, Novabiochem and GL biochem. All other chemical reagents were of analytical grade, obtained from Sigma Aldrich, alfa aesar and Acros Organics. All solvents and chemicals were used as received without purification unless otherwise indicated. T4 DNA ligase and restriction enzymes were obtained from New England BioLabs. Primer synthesis and gene sequencing service were provided by Integrated DNA Technologies (IDT) and Axil Scientific Pte Ltd in Singapore, respectively. The anti-biotin antibody was purchased from Abcam. Chitin Resin was purchased from NEB. Ni-NTA Agarose was purchased from Qiagen.

High Performance Liquid Chromatography (HPLC)

Analytical HPLC analyses were performed using a Shimadzu HPLC system equipped with a Jupiter C18 (5 µm, 4.6 x 250 mm) or a Jupiter C4 (5 µm, 4.6 x 250 mm) reverse-phase column with a flow rate of 1.0 mL/min. Semi-preparative HPLC were performed using a semi-preparative HPLC column (Jupiter C18, 5 µm, 10 x 250 mm) on a Shimadzu system with a flow rate of 2.5 mL/min. Preparative HPLC analyses were performed on Waters 600 systems equipped with a Vydac C4 column (10 µm, 22 x 250 mm) with a flow rate of 10 mL/min. Detection was done with a UV-VIS-detector at 220 nm. The buffer system for all the analyses was buffer A H₂O (containing 0.045% TFA) and buffer B 90% acetonitrile (ACN) in H₂O (containing 0.045% TFA).

Mass spectrometry

Small peptides ESI mass spectra data were obtained on a Thermo Finnigan LCQ DECA XP MAX with ESI ion source. Protein ESI mass spectra were measured on Q Exactive hybrid Quadrupole-Orbitrap Mass Spectrometer. The deconvoluted data were obtained using the software of MagTran 1.03 .

Cloning of plasmids used in our study

Construction of pEVOL-PylT-ThzKRS

The codon optimized sequence of *Methanosarcina barkeri* MS PylS gene (MbPylRS) [gb:AAQ19545.1] for *E. coli* use was synthesized by Genscript based on the reported work by Chin and coworkers.¹ The synthetic MbPylRS gene was inserted into plasmid pUC57 to yield pUC57-MbPylRS. ThzKRS gene containing four mutations (A276S, C313V, M315F, D344G) was generated via Overlap-PCR using the following primers: ThzKRS-NdeI-F1 (5'-GGG AAT TCC ATA TGG ATA AAA AAC CGC TGG ATG-3'), ThzKRS-A267S-R1(5'-GTT ATA CAG GGT CGG gga CAG CAT CGG ACG CAG GCA C-3'), ThzKRS-A267S-F2(5'-tcc CCG ACC CTG TAT AAC TAT C-3'), ThzKRS-C313V-M315F-R2(5'-GGT GCA GCC GCT GCC aaa TTG gac AAA GTT AAC CAT GGT GAA TTC-3'), ThzKRS-C313V-M315F-F3(5'-ttt GGC AGC GGC TGC ACC CGTG -3'), ThzKRS-D344G- R3 (5'-ATA CAC

CAT GCA GCT ccc GCC CAC AAT TTC GAA GTC G-3'), ThzKRS-D344G-F4 (5'-ggg AGC TGC ATG GTG TAT GGC G-3'), ThzKRS-PstI-R4(5'- AAA ACT GCA GTT ACA GGT TCG TGC TAA TGC C-3'). The DNA product was purified using PCR purification Kit and digested by NdeI and PstI for overnight. The digested ThzKRS was ligated with the precut pUC57 to obtain pUC57-ThzKRS.

The pEVOL-PyIT plasmid was derived from a pEVOL-pBpF, which was a gift from Peter Schultz (Addgene plasmid # 31190).² According to previous method,³ the pyIT gene with proK promoter and proK terminator was synthesized by overlap PCR with eight primers: pEVOL-PyIT-ApaI-F-1 (5'-gat atg atc agt gca cgg cta act aag cgg cct gct gac ttt ctc g-3'), pEVOL-PyIT-R-2 (5'-caa tcc ctt aat agc aaa atg cct ttt gat cgg cga gaa agt cag cag-3'), pEVOL-pyIT-F-3 (5'-gct att aag gga ttg acg agg gcg tat ctg cgc agt aag atg cgc ccc-3'), pEVOL-pyIT-R-4 (5'-agt cca ttc gat cta cat gat cag gtt tcc aat gcg ggg cgc atc tta c-3'), pEVOL-pyIT-F-5 (5'-gta gat cga atg gac tct aaa tcc gtt cag ccg ggt tag att ccc ggg g-3'), pEVOL-pyIT-R-6 (5'-ggc ttt tcg aat ttg gcg gaa acc ccggaatctaac-3'), pEVOL-pyIT-F-7(5'-caaattcgaaaagcctgctcaacgagcaggccttttt g-3'), pEVOL-pyIT-Xho1-R-8 (5'-ctgagctgctcgagcatgcaaaaaagcctgctc-3'). The PCR product was digested by ApaI and XhoI and purified using DNA gel extraction kit (Axygen). The purified DNA was ligated with precut pEVOL-pBpF to produce pEVOL-PyIT.

To introduce two copies of ThzKRS into pEVOL-PyIT, the first copy of ThzKRS was amplified from the pUC57-ThzKRS plasmid by flanking primers (pEVOL-mPyIRS-BglII-F:5'-GGAAGATCTATGGATAAAAAACCGCTGAACAC CCTGATC-3' pEVOL-mPyIRS-SalI-R:5'-ACGCGTCTGACTTACAGGTTTCGTTG AGATGCC-3') to introduce two restriction enzymes BglII and SalI at 5' and 3' end, respectively. The digested DNA by BglII and SalI was ligated to a precut pEVOL-PyIT plasmid with T4 ligase (NEB). The resulted plasmid was digested by NdeI and PstI enzymes and used to insert the second copy of ThzKRS gene that was cut from pUC57-ThzKRS by NdeI and PstI restriction enzymes. The resulted plasmid is pEVOL-PyIT-ThzKRS.

Construction of pTXB1-UbiquitinR74

The human ubiquitin gene (1-74) was amplified by PCR using the following primers pTXB1-Ubi-NdeI-F:5'-GGGAATTCCATATGCAGATCTTCGTC AAGAC GT-3'(forward), Ub74R-SapI-R:5'-GCGCAGGAAAGCTCTTCCGCATCTTAGT CTTAAGACAAGATG-3'(reverse). The PCR product was digested with NdeI and SapI, purified and ligated with the identically digested pTXB1 vector (New England Biolabs). Then the correct insert was confirmed by sequencing.

Construction of pETDuet-Rpn10K99TAG

pETDuet-Rpn10K99TAG was derived from ZM552, which was obtained from DNA Resource Core at Harvard Medical School. One amber codon mutation was introduced into Rpn10 at K99 position via Overlap-PCR using the following primers: Rpn10-F1-NcoI:5'-CAT GCC ATG GTA TTG GAA GCT ACAGTG-3', Rpn10-K99TAG-R1:5'-GACCTTATTCTGGCGATGCTaCAAAGT CAGCTGAGCGAT-3', Rpn10-K99TAG-F2:5'-TAGCATCGCCAGAATAAGGTC -3', Rpn10-R2-HindIII:5'-cccAAGCTTtattaATGATGATGATGATGATGTTTGT CTTGGTGTGTTTC-3'. The PCR product was double digested with NcoI and HindIII and purified using DNA gel extraction kit. The purified Rpn10K99TAG gene was ligated with the pre-cut pETDuet to produce pETDuet-Rpn10K99TAG

Construction of pETDuet-UbiquitinK48TAG and pETDuet-Ubiquitin63TAG

Amber codon mutation was introduced into ubiquitin at K63 position via Overlap-PCR using the following primers: Ub-NdeI-F1: 5'-5'-GGGAATTCATATGCAGATCTTCGTCAAG-3', UbK63TAG-R1: 5'-GTA AGG TCG ACT CCT ACT GAA TGT TGT AAT CAG AC-3', UbK63TAG-F2: 5'-GAG TCG ACC TTA CAT CTT-3', UbH6-KpnI-R: 5'-CGG GGT ACC tta tta ATG ATG ATG ATG ATG ACC ACC TCT TAG TCT TAA G-3'. Digested ubiquitin-K63TAG by NdeI and KpnI was purified by DNA gel extraction kit. The purified DNA product was ligated with pre-cut pETDuet to get pETDuet-Ubiquitin63TAG. pETDuet-UbiquitinK48TAG was constructed in a similar way.

Expression and Purification of protein

Expression and Purification of UbiquitinR74-MES:

The expression vector for Ub(1-74)-MES was transformed into *E. coli* BL21 (DE3) cells. The overnight culture of cells was grown in 2 L of LB media containing ampicillin (100 mg/L) at 37 °C. When OD₆₀₀ reached 0.6, the cells were induced with IPTG (0.05 mM) for 16 h at 16 °C. The cells were then harvested through centrifugation at 6,500 g. The cells were then resuspended in 50 mL lysis buffer (20 mM HEPES, 50 mM sodium acetate, 75 mM NaCl, pH 6.5). The cells were lysed by sonication. The cell lysate was centrifuged at 23,000 g for 30 min. For the binding of the ubiquitin-intein fusion protein to the chitin beads, the supernatant were mixed with 10 ml of chitin beads which had been pre-equilibrated with lysis buffer. The mixture was stirred at 4 °C for 2 h. The beads was then collected by filtration through a silica pad and washed extensively with lysis buffer. The thiolysis of the ubiquitin-intein fusion protein was performed by mixing the beads with 20 mL of lysis buffer containing 0.2 M MESNa (pH 7.3). After overnight cleavage at 37 °C, the cleavage solution was subjected to C4-preparative HPLC purification. The fractions were analyzed by ESI-MS. The desired product was lyophilized. The yield of ub(1-74)-MES was about 5 mg/L.

Expression and Purification of Ubiquitin and Rpn10 incorporating ThzK:

For expression of proteins with ThzK, *E. coli* BL21(DE3) cells were cotransformed with the plasmid encoding the target gene containing the amber stop codon and the vector encoding the orthogonal ThzKRS/pyIT pair. Cells were grown in LB media supplemented with 50 µg/ml chloramphenicol and 100 µg/ml ampicillin at 37 °C. When an OD₆₀₀ of 0.6 was reached, unnatural amino acid **2** was added to a final concentration of 1 mM and protein expression was induced with 1 mM IPTG. Cultures were then incubated for another 8 h at 37 °C. All cells were harvested by centrifugation, resuspended in the respective buffer, and lysed by sonification. His-tagged Ubiquitin and Rpn10 were purified on Ni-NTA agarose (Qiagen) equilibrated in Ni-NTA buffer A (50 mM Tris, 300 mM NaCl, pH 8.0) supplemented with 10 mM imidazole. After washing with Ni-NTA buffer A supplemented with 10 mM imidazole, protein was eluted out from the agarose by using 250 mM imidazole. Purified proteins were dialyzed against dialysis buffer and stored at – 80°C for further use.

Synthesis of ubiquitin 6

2 mg of UbiquitinR74-MES was first dissolved in 100 µl DMSO then 100 µl of aminoacetaldehyde diethyl acetal 98% (Sigma) was added and incubated at room temperature. Usually the near complete conversion from thioester to acetal was observed after 1h. Then excess cold ether was added to quench the reaction. The aminoacetaldehyde diethyl acetal dissolved in ether can be removed by centrifuge, leaving the ubiquitin acetal as pellet. The protein pellet was dissolved in HPLC buffer A again and injected into C4 semi-HPLC for purification. The desired product was confirmed by ES-MS analysis and lyophilized for drying. The ubiquitin acetal was treated with 200 µL of TFA/TIS/H₂O (95/2.5/2.5) for 20 min. After ether precipitation, the crude deprotection product was purified by reverse-phase HPLC to give pure ubiquitin aldehyde, which is confirmed by ESI-MS analysis. The desired product was lyophilized for drying and stored at -20 °C for further use.

Preparation of biotin-labeled ubiquitin 5 at K48 position

Aldehyde-functionalized peptide (5, 10 or 20 equivalent) was added to the reaction buffer (0.2 M phosphate, 10 mM aniline, pH 7.0) at a concentration of about 25 µM ubiquitin **4**. The reaction was incubated for 8 h at room temperature. The reaction was monitored by C4 analytic HPLC and ESI-MS analysis. The purified biotin-labeled ubiquitin **5** was collected for western blot analysis using anti-biotin antibody.

Preparation of ubiquitin conjugated protein via thiazolidine ligation

For the ubiquitin dimer formation via thiazolidine ligation, a typical reaction was set as the followings: ubiquitin **4** and **6** was mixed in equimolar amounts of 100 µM each in the reaction buffer (0.2 M sodium acetate buffer, 6M guanidine

chloride, 1 mM TCEP, pH 5.4) for 12 h. The total reaction volume ranged from 1.5 to 2 ml. The desired ubiquitin dimer analog can be purified using C4 semipreparative HPLC. For the preparation of ubiquitinated Rpn10 via thiazolidine ring formation, Rpn10-aminothiol (50 μ M) was mixed together with ubiquitinR74-aldehyde (500 μ M) in a 1:10 ratio in the above reaction buffer, 6M guanidine chloride, pH 5.4) for 12 h. The total reaction volume ranged from 0.2 to 0.4 ml. The reaction was monitored by C4 analytical HPLC.

Synthesis of biotin labeled peptide

The peptides H-SKK(biotin)LA-NH₂ were synthesized using standard Fmoc chemistry. The synthesis of H-SKK(biotin)LA-NH₂ were done on Rink amide MBHA resin. Before use, resin was pre-swelled using DCM for 20 min. Before the first coupling, an Fmoc deprotection was performed using 20% piperidine in dimethylformamide (DMF) for 2 min and a subsequent 18 min. After deprotection, the resin was washed with DMF, DCM and DMF alternatively. For the coupling reactions, 3 eq. of Fmoc-AA-OH, 3 eq. of PyBOP were first dissolved in DMF/DCM mixture. After the mixture was added to the resin, 6 eq. of DIEA was added. Coupling reactions were carried out for 60 to 90 min. Coupling efficiency was examined by Kaiser test. The Fmoc amino acids used for the synthesis of H-SKK(biotin)LA-NH₂ were Fmoc-Ser(tBu)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Ala-OH. For the coupling of D-biotin with the first lysine in the peptide, the Boc group was deprotected using 1M HCl in DCM for 30-40 min. Then 3 equivalent of D-Biotin and 3 equivalent of PyBOP was added into the above resin containing NMP solvent. The coupling was performed for overnight. After successful coupling of D-Biotin to the peptide, the next coupling was done as prescribed above. After the whole peptide was assembled, peptide was cleaved from the resin with a cocktail containing 95% TFA, 2.5% water and 2.5% TIS for one hour. After precipitated with cold ether, the peptide was purified using HPLC and characterized by mass spectrometry. Desired peptides were obtained in the white powder form after lyophilization. The preparation of aldehyde functionalized peptide was performed as the following typical reaction: Peptide (10mg) was dissolved in 1.5 ml of 0.1 M sodium phosphate buffer (pH 7.0) supplemented with 20 mg Methionine and treated with 193.5 μ l of 100 mM NaIO₄ on ice for 2 min. The reaction was stopped by injecting the sample directly into HPLC. The product was separated on a semipreparative HPLC column using a gradient from 0 to 50% B within 50min and gave a major peak from the analytical HPLC. ESI-MS found 740.593 and 758.584 (hydrated form), which were consistent with the calculated mass.

Western blot:

About 2 μ g of non-labeled ubiquitin and 1.5 μ g of biotin-labeled ubiquitin were analyzed by 5-15% SDS-PAGE. The proteins on the gel were then electrotransferred to nitrocellulose membrane. The membrane was blocked by 10 mL of 5% w/v non-fat milk powder in TBS buffer (100 mM Tris-HCl, 150 mM

NaCl, pH 7.5) containing 0.1% Tween-20 for 1 h at room temperature. A 1:5000 dilution of Anti-Biotin antibody (HRP) (ab53468) was added to the blocking mixture and incubated for 1 h at room temperature. The biotin-labeled ubiquitin were visualized by chemilluminescence (SuperSignal West Dura Trial Kit, Pierce, USA).

Reference:

- [1] H. Neumann, S. Y. Peak-Chew and J. W. Chin, *Nat. Chem. Biol.*, 2008, **4**, 232-234.
- [2] J. W. Chin, A. B. Martin, D. S. King, L. Wang and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 11020-11024.
- [3] Y. S. Wang, X. Fang, A. L. Wallace, B. Wu and W. R. Liu, *J. Am. Chem. Soc.*, 2012, **134**, 2950-2953.