Single-site labeling of histidine in proteins, on-demand reversibility, and traceless metal-free protein purification

Pralhad Namdev Joshi, Vishal Rai*

Department of Chemistry, Indian Institute of Science Education and Research Bhopal, Bhauri, Bhopal, Madya Pradesh 462 066 India

vrai@iiserb.ac.in

Contents
1. General.......................................................................................................................... S2
2. Methods........................................................................................................................ S3
   2.1. Procedure for protein labeling.................................................................................. S3
   2.2. Procedure for late-stage modification........................................................................ S3
   2.3 Procedure for protein expression................................................................................ S3
   2.4. Procedure for purification......................................................................................... S4
   2.4. Digestion protocol.................................................................................................... S4
3. Synthesis of reagents .................................................................................................... S5
4. Additional results and discussion.................................................................................. S10
5. Structural analysis ........................................................................................................ S16
   5.1 Stability of 2-cyclohexenone..................................................................................... S16
   5.2. Circular dichroism spectroscopy............................................................................... S17
   5.3. Enzymatic assay .................................................................................................... S18
6. Late-stage modification .................................................................................................. S19
7. MS data of proteins....................................................................................................... S21
8. NMR data ..................................................................................................................... S36
9. Acknowledgements....................................................................................................... S46
10. References .................................................................................................................... S46
1. General
The required chemicals, proteins, and enzymes were procured from Merck and Thermo Fisher Scientific. The reagent grade organic solvents were utilized. The incubator-shaker MaxQ 8000 from Thermo Scientific was used to vortex the reaction mixtures. Organic solvents were evaporated on BUCHI Rotavapor R-210/215. The lyophilization of aqueous samples was carried out on CHRIST ALPHA 2-4 LD plus lyophilizer. The JASCO J-815 CD spectropolarimeter equipped with a Peltier temperature controller was used for recording circular dichroism (CD). The CD spectra were measured in a cuvette of path length 0.2 cm with a scan speed of 50 nm/min at 25 °C and spectral bandwidth of 1 nm. The UV-Vis and CD spectra were plotted with the help of OriginPro 8 software. In the enzymatic assay, absorbance was normalized to its values by dividing its absorbance maxima (OriginPro 8). The reactions were monitored by thin-layer chromatography (TLC) on silica gel coated aluminium TLC plates (Merck, TLC Silica gel 60 F254). The flash column chromatography was carried out on CombiFlash Rf 200 or simple glass columns using 230-400 mesh silica gel from Merck. \( ^1H \) and \( ^{13}C \) NMR spectra were obtained from Bruker Avance III 400 and 500 MHz NMR spectrometer at 298 K. TMS (0 ppm), CDCl\(_3\) (7.26 ppm) and D\(_2\)O (4.79 ppm) were used as references for \( ^1H \) NMR spectra. The \( ^{13}C \) NMR spectra were referenced to CDCl\(_3\) (77.16 ppm). Peak multiplicities are designated by s for singlet, d for doublet, t for triplet, q for quartet, dd for doublet of doublets and m for multiplet. Agilent Technologies 1200 series HPLC paired to Agilent 6130 mass spectrometer (ESI/APCI) was used to monitor the reactions with small molecules and proteins. HPLC experiments were performed on a Poroshell 300SB-C18 column (3.0 × 50 mm × 2.7 µ) with a flow rate of 0.4 ml/min. Bruker Daltonics MicroTOF-Q-II with electrospray ionization (ESI) was used for the HRMS data. Matrix-assisted laser desorption/ionization time of flight mass spectrometry was performed with Bruker Daltonics UltrafleXtreme Software-Flex control version 3.4. The proteins like ubiquitin, lysozyme C, insulin, \( \alpha \)-lactalbumin, RNase A, \( \beta \)-lactoglobulin, and myoglobin were analyzed with the sinapic acid matrix on MALDI-ToF-ToF. The \( \alpha \)-cyano-4-hydroxycinnamic acid (HCCA) matrix was used for peptide mapping of digest from all the proteins. Data analysis was performed using flexAnalysis 3.4. Peptide mass and fragment ion calculator were used for peptide mapping and sequencing (http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html). EDX analysis was performed using transmission electron microscopy (TEM) FEI TALOS 200S instrument. The samples for TEM analysis were prepared by drop casting a homogeneous dilute dispersion of sample over a carbon coated 400 mesh Cu grid.
2. Methods

2.1. Procedure for protein labeling
In a 1.5 ml Eppendorf tube, protein (7.5 nmol) in phosphate buffer (90 μl, pH 7.0, 0.1 M) was taken. The freshly prepared stock solution of 2-cyclohexenone (3m) (10 μl, 1.9 μmol) in phosphate buffer (pH 7.0, 0.1 M) was added. The final concentration of protein and reagent was 75 μM and 18.8 mM respectively. The reaction mixture was vortexed (350 rpm) in the incubator shaker at 25 °C for the given time. The reaction was followed by MALDI-ToF MS or ESI-MS. The excess of 2-cyclohexenone (3m) and salts were removed by using the centrifugal spin concentrators (Amicon® Ultra-0.5 mL, MWCO 3 kDa and 10 kDa) and then the sample was concentrated by lyophilization before subjecting it for the digestion, peptide mapping and sequencing by MS-MS.

2.2. Procedure for late-stage modification
In a 1.5 ml Eppendorf tube, modified lysozyme C (4n) (7.5 nmol) in phosphate buffer (90 μl, pH 7.0, 0.1 M) was taken. Tags containing alkoxyamine derivatives (5a-5c, 1.5 μmol) in DMSO (10 μl) was added from freshly prepared stock solutions. The final concentration of protein and reagent was 75 μM and 15 mM respectively. The reaction mixture was vortexed (350 rpm) in incubator shaker at 25 °C for the required time and the reaction was followed by ESI-MS. The protein solution was desalted by using the (Amicon® Ultra-0.5 mL, MWCO 3 kDa and 10 kDa) and then the sample was concentrated by lyophilization before subjecting it to digestion, peptide mapping and sequencing by MS-MS.

2.3 Procedure for protein expression
1 μl (100 ng/μl) of pET-28a(+)–Hs Ubc9 plasmid was added into the 100 μl of BL21 normal cells and was incubated on ice for 20 minutes. Then heat shock was given at 42 °C for 40 seconds. Cells were kept on ice for 2 minutes and 800 μl of Luria broth (LB) was added to cells for recovery. Cells were incubated at 37 °C, 200 rpm for 45 min. Then recovered cells were plated on LB plates containing kanamycin antibiotics (stock 60 mg/ml, working 1:1000). Then plates were incubated at 37 °C, overnight. Next day, one healthy colony was picked from the plate and
inoculated into 2 ml LB medium with kanamycin antibiotics and allow to grow overnight at 37 °C, 200 rpm. Secondary culture grown from primary culture in a 1:100 ratio at 37 °C, 200 rpm. At approximately 0.6 OD (600 nm), the secondary culture was induced with IPTG (200 μM) for 4 hour at 30 °C. Induced culture was spun at 8000 rpm for 10 minutes to pellet down. Re-suspended the cell pellet in lysis buffer [50 mM NaH₂PO₄, 300 mM NaCl, 10 mM Imidazole, 20 mM β-ME, 0.2% Triton X-100, 50 μg/ml lysozyme (1:1000), 1 mM PMSF, PIC (1:1000)]. The suspension was incubated in ice for 10-15 min followed by sonicated 45% amplitude 10 sec ON 10 sec OFF for 1 min and further utilized to purify Ubc9.

2.4. Procedure for purification
The cell lysate obtained from E. coli cells were subjected for modification with 2-cyclohexenone (3m, 36 mM) and the reaction mixture was vortexed (350 rpm) in incubator shaker at 25 °C for 3 h. In a 2 ml Eppendorf tube, alkoxyamine derivatized sepharose resin was mixed with modified cell lysate. The mixture was kept for end to end rotation for 3 h to capture the modified His-tagged recombinant protein on resin followed by unmodified proteins were washed with water 400 μl and then with KCl (3 M) solution 400 μl. The protein loaded on resin was subjected for cleavage either by imidazole 1 M (or 1 M NH₂OH) (phosphate buffer, pH 7.0, 0.1 M). After end to end rotation for 6 h the purified sample was collected with washing conditions as stated earlier. The aliquot was desalted by using (Amicon® Ultra-0.5 mL, MWCO 3 kDa and 10 kDa) and the pure protein was obtained by lyophilization.

2.4. Digestion protocol
Procedure for in-solution digestion of RNase A, lysozyme C, α-lactalbumin, β-lactoglobulin and insulin

All the solutions were prepared freshly before use in reactions.¹

**Step 1.** Protein (0.1 mg, 7.5 nmol) in 100 mM tris (10 μl, pH 7.8) with urea (6 M) was taken in 1.5 ml Eppendorf tube and incubated for 30 minutes.

**Step 2. Disulfide reduction:** To reduce the disulfide bonds, reducing agent (1 μL, 0.2 M DTT in 0.1 M tris) was added to the solution and sample was vortexed for 1 h at 25 °C or 37 °C.

**Step 3. Sulphydryl alkylation:** To block the free sulfhydryl groups, alkylating agent (4 μL, 0.2 M iodoacetamide in 0.1 M tris) was added to the solution and incubated (in the dark) for 1 h at ambient temperature.

**Step 4. Quenching alkylation reagent:** To quench the unreacted iodoacetamide, reducing agent (DTT, 4 μL) was added again to the mixture and the sample was vortexed at 25 °C for 1 h. Dilution of the reaction mixture with grade I water reduced the urea concentration to 0.6 M.

**Step 5. Enzymatic digestion:** To this solution, 10 μL of α-chymotrypsin or trypsin solution [2 μg, based on ratio of chymotrypsin or trypsin/protein (1:50); α-chymotrypsin or trypsin in 1 mM HCl was dissolved in 0.4 M tris and 0.01 M CaCl₂] was added and the mixture was incubated at 37 °C for 18 h. Trifluoroacetic acid (0.5 %) was used to adjust the pH of digested solution to < 6 (verified by pH paper). Subsequently, the sample was subjected to peptide mapping by MS and sequencing by MS-MS.
**Modified peak intensity enhancement**: In order to get intense peak of modified peptide fragment the sensitivity enhancement reagent (3-(aminoxy)-N,N,N-triethylpropan-1-aminium bromide (5d), 7.5 mM) was added while carrying out digestion. The quaternary amine (-N°Et3) center helped in enhancement of signal intensity and aminooxy functionality served as a partner for formation of oxime with ketone.2

**Procedure for in-solution digestion of ubiquitin and myoglobin**

Steps 1 and 5 were used for digestion of ubiquitin, and myoglobin. Steps 2, 3 and 4 are not desired as these proteins do not have disulfide bridges or free sulfhydryl groups.

3. **Synthesis of reagents**

![Scheme S1. Synthesis of 3-(aminoxy)propyl 3,5-bis(trifluoromethyl) benzoate 5a.](image)

**Synthesis of 2-(3-bromopropoxy)isoindoline-1,3-dione S3**

In a 250 ml round bottom flask charged with magnetic stir bar, N-hydroxyphthalimide S1 (4894 mg, 30 mmol) and triethyl amine (6.09 ml, 60 mmol) were dissolved in acetonitrile (60 ml). 1,3-dibromopropane S2 (8.34 ml, 60 mmol) was added in the reaction mixture and stirred at 25 °C for 16 h. The reaction mixture was concentrated in vacuo and then 1 N NaOH solution and ethyl acetate were added for extraction. The organic layer was separated, dried over anh. sodium sulfate, filtered, and concentrated in vacuo. The purification crude mixture was performed by
flash column chromatography using ethyl acetate: hexane (3:97) gave S3 in 50% yield; Rf 0.57, ethyl acetate: n-hexane 30:70; white solid. \(^1\)H NMR (400 MHz, CDCl\(_3\)) \(\delta\) 7.85-7.82 (m, 2H), 7.77-7.74 (m, 2H), 4.36 (t, \(J = 5.8\) Hz, 2H), 3.70 (t, \(J = 6.5\) Hz, 2H), 2.30-2.28 (m, 2H) ppm. \(^{13}\)C NMR (101 MHz, CDCl\(_3\)) \(\delta\) 163.7, 134.7, 129.0, 123.7, 76.2, 31.6, 29.4 ppm. MS (ESI) [M+H]\(^+\) calcd. C\(_{11}\)H\(_{11}\)BrNO\(_3\) 284.0, found 284.0 and calcd. C\(_{11}\)H\(_{11}\)BrNO\(_3\) 286.0, found 285.9.

**Synthesis of 3-((1,3-dioxoisooindolin-2-yl)oxy)propyl 3,5-bis(trifluoromethyl)benzoate S5**

![Chemical structure of S5](image)

In a 25 ml round bottom flask, 3,5-bis(trifluoromethyl)benzoic acid S4 (258 mg, 1 mmol), 2-(3-bromopropoxy)isoindoline-1,3-dione S3 (312 mg, 1.1 mmol) and TEA (418 µl, 3 mmol) were dissolved in acetonitrile (5 ml). The reaction mixture was refluxed and the progress of the reaction was followed by TLC. After 8 h, the reaction mixture was concentrated and purification by silica gel flash column chromatography (ethyl acetate: n-hexane, 2:98) led to the isolation of 3-((1,3-dioxoisooindolin-2-yl)oxy)propyl 3,5-bis(trifluoromethyl)benzoate S5 (335 mg, 73% yield; Rf 0.67, ethyl acetate: n-hexane 30:70; white solid). \(^1\)H NMR (500 MHz, CDCl\(_3\)) \(\delta\) 8.51 (s, 2H), 8.05 (s, 1H), 7.84-7.82 (m, 2H), 7.76-7.75 (m, 2H), 4.70 (t, \(J = 6.3\) Hz, 2H), 4.39 (t, \(J = 6.0\) Hz, 2H), 2.32-2.26 (m, 2H) ppm. \(^{13}\)C NMR (126 MHz, CDCl\(_3\)) \(\delta\) 164.0, 163.7, 134.7, 132.8, 132.3 (q, \(J = 34.1\) Hz, 2C), 130.1-129.8 (m, 2C), 129.0, 126.5-126.4 (m, 1C), 123.7, 123.0 (q, \(J = 272.8\) Hz, 2C), 74.9, 62.7, 27.8 ppm. \(^{19}\)F NMR (376 MHz, CDCl\(_3\)) \(\delta\) -62.94 (TFA was used as an internal standard, -75.70 ppm). HRMS (ESI) [M+H]\(^+\) calcd. C\(_{20}\)H\(_{14}\)F\(_6\)NO\(_5\) 462.0776, found 462.0775.

**Synthesis of 3-(aminoxy)propyl 3,5-bis(trifluoromethyl)benzoate 5a**

![Chemical structure of 5a](image)

In a 5 ml round bottom flask, 3-((1,3-dioxoisooindolin-2-yl)oxy)propyl 3,5-bis(trifluoromethyl)benzoate S5 (138 mg, 0.3 mmol) in DCM (3 ml) and hydrazine monohydrate (80%, 37 µl, 0.75 mmol) were stirred at room temperature. The progress of the reaction was followed by TLC. After 3 h, the reaction mixture was filtered and concentration of the filtrate in vacuo led to the isolation of 3-(aminoxy)propyl 3,5-bis(trifluoromethyl)benzoate 5a (80 mg, 81% yield; Rf 0.33, ethyl acetate: n-hexane 30:70; white solid). \(^1\)H NMR (500 MHz, CDCl\(_3\)) \(\delta\) 8.47 (s, 2H), 8.06 (s, 1H), 5.42 (bs, 2H), 4.49 (t, \(J = 6.5\) Hz, 2H), 3.82 (t, \(J = 6.1\) Hz, 2H), 2.13-2.07 (m, 2H) ppm. \(^{13}\)C NMR (126 MHz, CDCl\(_3\)) \(\delta\) 164.1, 132.6, 132.4 (q, \(J = 33.9\) Hz, 2C), 130.0-129.7 (m, 2C), 126.6-126.3 (m, 1C), 123.02 (q, \(J = 273.0\) Hz, 2C), 72.2, 63.6, 27.9 ppm.
Scheme S2. Synthesis of 7-((3-(aminooxy)propyl)thio)-4-methyl-2H-chromen-2-one 5b.

Synthesis of 2-((3-(4-methyl-2-oxo-2H-chromen-7-yl)thio)propoxy)isoindoline-1,3-dione S7

In a 25 ml round bottom flask, 7-mercapto-4-methylcoumarin S6 (192 mg, 1 mmol), K$_2$CO$_3$ (276 mg, 2 mmol), and 2-((3-bromopropoxy)isoindoline-1,3-dione S3 (568 mg, 2 mmol) were dissolved in degassed acetonitrile (5 ml) and refluxed for 16 h. The reaction mixture was concentrated in vacuo and purified by silica gel flash column chromatography using ethyl acetate:hexane (7:3) to give S7 (375 mg, 95% yield; R$_f$ 0.37, ethyl acetate:n-hexane 50:50; white solid). $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.44 (d, $J$ = 8.3 Hz, 1H), 7.26-7.13 (m, 2H), 6.18 (d, $J$ = 0.9 Hz, 1H), 3.72 (t, $J$ = 5.9 Hz, 2H), 3.05 (t, $J$ = 0.9 Hz, 2H), 2.41 (d, $J$ = 0.9 Hz, 3H), 2.16-2.13 (m, 2H) ppm. $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 163.6, 160.6, 153.9, 152.1, 142.5, 134.6, 128.9, 124.7, 123.6, 123.3, 117.4, 114.6, 114.0, 76.5, 28.5, 27.7, 18.6 ppm. HRMS (ESI) [M+H]$^+$ calcd. C$_{21}$H$_{18}$NO$_5$S 396.0906, found 396.0925.

Synthesis of 7-((3-(aminooxy)propyl)thio)-4-methyl-2H-chromen-2-one 5b

2-(3-((4-methyl-2-oxo-2H-chromen-7-yl)thio)propoxy)isoindoline-1,3-dione S7 (237 mg, 0.6 mmol) was dissolved in CH$_2$Cl$_2$ (12 ml) in a 50 ml round bottom flask. To this solution, hydrazine monohydrate (80%, 29 µl, 0.6 mmol) was added and stirred at 25 °C for 3 h. The reaction mixture was filtered and the filtrate was concentrated. The purification of crude reaction mixture was performed by reverse phase preparative HPLC to isolate 5b (76 mg, 45% yield; R$_f$ 0.6, ethyl acetate:n-hexane 50:50; pale green viscous liquid). $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.44 (d, $J$ = 8.3 Hz, 1H), 7.26-7.13 (m, 2H), 6.18 (d, $J$ = 0.9 Hz, 1H), 3.72 (t, $J$ = 5.9 Hz, 2H), 3.05 (t,
Scheme S3. Synthesis of 3-(aminooxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate 5c.

Synthesis of 3-((1,3-dioxoisindolin-2-yl)oxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate S9

In a 50 ml round bottom flask, biotin S8 (244 mg, 1 mmol), 2-(3-bromopropoxy)isoindoline-1,3-dione S3 (568 mg, 2 mmol), and DBU (304 µl, 2 mmol) were dissolved in acetonitrile (20 ml). The reaction mixture was refluxed and the progress of the reaction was monitored by TLC. After 16 h, the reaction mixture was concentrated in vacuo. Next, the solvent-solvent extraction was carried out by using ethyl acetate and water. The organic fractions were combined, dried over anh. sodium sulfate, filtered, and concentrated in vacuo. The purification of crude reaction mixture was performed by silica gel flash column chromatography (MeOH:DCM, 0.5-5%) to isolate 3-((1,3-dioxoisindolin-2-yl)oxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate S9 (224 mg, 50% yield; Rf 0.33, MeOH:DCM 05:95; white solid). $^1$H NMR (500 MHz, CDCl$_3$) δ 7.84-7.83 (m, 2H), 7.75-7.74 (m, 2H), 5.95 (s, 1H), 5.46 (s, 1H), 4.51-4.48 (m, 1H), 4.37-4.27 (m, 5H), 3.17-3.13 (m, 1H), 2.90 (dd, J = 12.8, 5.0 Hz, 1H), 2.74-2.71 (m, 1H), 2.34 (t, J = 7.4 Hz, 2H), 2.13-2.08 (m, 2H), 1.76-1.62 (m, 4H), 1.50-1.39 (m, 2H) ppm. $^{13}$C NMR (126 MHz, CDCl$_3$) δ 173.6, 163.6, 163.6, 134.6, 128.9, 123.6, 75.0, 61.9, 60.6, 60.1, 55.4, 40.4, 33.9, 28.3, 28.2, 27.7, 24.9 ppm. HRMS (ESI) [M+H]$^+$ calcd. C$_{21}$H$_{26}$N$_3$O$_6$S 448.1542, found 448.1548.

Synthesis of 3-(aminooxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate 5c

This compound was synthesized according to the procedure for synthesis of 5b.
Yield 71%; Rf 0.21, MeOH:DCM 05:95; white solid. $^1$H NMR (500 MHz, D$_2$O) $\delta$ 4.63 (dd, $J = 7.9, 4.9$ Hz, 1H), 4.45 (dd, $J = 7.9, 4.5$ Hz, 1H), 4.21 (t, $J = 6.3$ Hz, 2H), 3.90 (t, $J = 6.2$ Hz, 2H), 3.38-3.34 (m, 1H), 3.02 (dd, $J = 13.1, 5.0$ Hz, 1H), 2.82-2.79 (m, 1H), 2.44 (t, $J = 7.3$ Hz, 2H), 2.02-1.98 (m, 2H), 1.77-1.61 (m, 4H), 1.47-1.44 (m, 2H) ppm. $^{13}$C NMR (126 MHz, D$_2$O) $\delta$ 176.9, 165.3, 72.4, 62.1, 62.0, 60.3, 55.3, 39.7, 33.7, 27.9, 27.6, 26.8, 24.1 ppm. HRMS (ESI) [M+H]$^+$ calcd. C$_{15}$H$_{24}$N$_3$O$_4$S 318.1488, found 318.1467.

Scheme S4. Synthesis of 3-(aminooxy)-N,N,N-triethylpropan-1-aminium bromide 5d.

Synthesis of 3-((1,3-dioxoisoidolin-2-yl)oxy)-N,N,N-triethylpropan-1-aminium bromide S10

In a 5 ml round bottom flask, 2-(3-bromopropoxy)isoindoline-1,3-dione S3 (85 mg, 0.3 mmol), was dissolved in dichloromethane (3 ml) and trimethylamine (168 µl, 1.2 mmol) added to the reaction mixture. The reaction was refluxed and the progress of the reaction was monitored by TLC. After 24 h, the reaction mixture was concentrated in vacuo. This was followed by triturated with diethyl ether to obtain pure product 3-((1,3-dioxoisoidolin-2-yl)oxy)-N,N,N-triethylpropan-1-aminium bromide S10 (110 mg, 95% yield; as a white solid). $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.84-7.82 (m, 2H), 7.79-7.77 (m, 2H), 4.41 (t, $J = 5.8$ Hz, 2H), 3.80-3.77 (m, 2H), 3.61 (q, $J = 7.2$ Hz, 6H), 2.42-2.37 (m, 2H), 1.45 (t, $J = 7.1$ Hz, 9H) ppm. $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 163.7, 135.0, 128.7, 123.9, 53.9, 46.2, 22.2, 8.8, 8.2 ppm. HRMS (ESI) [M]$^+$ calcd. C$_{17}$H$_{25}$N$_2$O$_3$ 305.1865, found 305.1838.

Synthesis of 3-(aminooxy)-N,N,N-triethylpropan-1-aminium bromide 5d

This compound was synthesized according to the procedure for synthesis of 5a.

Yield 84%; white solid. $^1$H NMR (400 MHz, D$_2$O) $\delta$ 3.95 (t, $J = 5.7$ Hz, 2H), 3.35-3.29 (m, 8H), 2.08-2.04 (m, 2H), 1.29 (t, $J = 7.2$ Hz, 9H) ppm. $^{13}$C NMR (101 MHz, D$_2$O) $\delta$ 71.8, 53.6, 52.6, 20.4, 6.6 ppm. HRMS (ESI) [M]$^+$ calcd. C$_9$H$_{25}$N$_2$O 175.1810, found 175.1807.
4. Additional results and discussion

Table S1. Optimization of stoichiometry of the electrophile.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Protein</th>
<th>Equivalent</th>
<th>% Conversion$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>+1</td>
</tr>
<tr>
<td>1</td>
<td>Ubiquitin</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Ubiquitin</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>Ubiquitin</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>Ubiquitin</td>
<td>150</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>Ubiquitin</td>
<td>250</td>
<td>48</td>
</tr>
<tr>
<td>6</td>
<td>Ubiquitin</td>
<td>500</td>
<td>56</td>
</tr>
<tr>
<td>7</td>
<td>Ubiquitin</td>
<td>1000</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>RNase A</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>RNase A</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>RNase A</td>
<td>100</td>
<td>28</td>
</tr>
<tr>
<td>11</td>
<td>RNase A</td>
<td>150</td>
<td>33</td>
</tr>
<tr>
<td>12</td>
<td>RNase A</td>
<td>250$^b$</td>
<td>39</td>
</tr>
<tr>
<td>13</td>
<td>RNase A</td>
<td>500</td>
<td>45</td>
</tr>
<tr>
<td>14</td>
<td>RNase A</td>
<td>1000</td>
<td>35</td>
</tr>
</tbody>
</table>

$^a$% Conversions were determined by MALDI-MS. Extent of labeling: +1 to +3. $^b$38% conversion by ESI-MS.
Table S2. Effect of reaction medium pH on the protein labeling.

![Chemical structure of ubiquitin](image)

<table>
<thead>
<tr>
<th>Entry</th>
<th>Protein</th>
<th>pH</th>
<th>% Conversion&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>+1</td>
</tr>
<tr>
<td>1</td>
<td>Ubiquitin</td>
<td>5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>Ubiquitin</td>
<td>6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>Ubiquitin</td>
<td>7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>Ubiquitin</td>
<td>8&lt;sup&gt;b&lt;/sup&gt;</td>
<td>43</td>
</tr>
<tr>
<td>5</td>
<td>Ubiquitin</td>
<td>9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>Ubiquitin</td>
<td>10&lt;sup&gt;c&lt;/sup&gt;</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>RNase A</td>
<td>5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>RNase A</td>
<td>6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>33</td>
</tr>
<tr>
<td>9</td>
<td>RNase A</td>
<td>7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>RNase A</td>
<td>8&lt;sup&gt;b&lt;/sup&gt;</td>
<td>34</td>
</tr>
<tr>
<td>11</td>
<td>RNase A</td>
<td>9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>RNase A</td>
<td>10&lt;sup&gt;c&lt;/sup&gt;</td>
<td>28</td>
</tr>
</tbody>
</table>

<sup>a</sup> % Conversions determined by MALDI-MS. <i>Extent of labeling</i>: +1 to +3. <sup>b</sup> Phosphate buffer (0.1 M). <sup>c</sup> PBS (0.1 M).
Table S3. Screening of additives for inducing the reversibility.

![Image of molecular structures](image)

<table>
<thead>
<tr>
<th>Entry</th>
<th>Additive (7)</th>
<th>% Reversibility (6d to 2b)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No additive</td>
<td>&lt;5</td>
</tr>
<tr>
<td>2</td>
<td>Tertramethyl guanidine (7a)</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>DABCO (7b)</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>DBN (7c)</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>DBU (7d)</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>DMAP (7e)</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>N-methyl imidazole (7f)</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>Imidazole (7g)</td>
<td>93</td>
</tr>
</tbody>
</table>

$^a$ % Reversibility is determined by MALDI-MS.

Estimation of % conversion by LC, ESI-MS, and MALDI-MS:

The deviation between different techniques (LC, ESI-MS, MALDI-MS) is expected to be within ±5% particularly if we avoid the extreme changes in the absorption or ionization potential. As anticipated, insulin and cyclohexanone labelled insulin were not separable in HPLC under diverse conditions. To address the concern, we changed the polarity of the product (4o) by adding an additional tag (8a, 100% conversion). The tagged insulin (8a) and insulin (2e) were separable on HPLC and suggested 47% conversion (Figure S1). As anticipated, the estimation of conversion by ESI-MS (47% conversion) and MALDI (49% conversion) was within ±5% range. Also, the earlier investigations by us and others indicate that the MALDI$^4,5,6,7$ and ESI-MS$^8,9$ can be used for the estimation of % conversion.
**Figure S1.** Comparative estimation of % conversion. (a) HPLC spectrum for the reaction mixture, containing insulin (2c) and mono-labeled insulin (8a). (b) ESI-MS spectrum of the LC peak at 10.8 min (insulin 2c). (c) ESI-MS spectrum of the LC peak at 12.0 min (mono-labeled insulin 8a). (d) ESI-MS spectrum of both the LC peaks (2c+8a) suggest 47% conversion. (e) MALDI-ToF-MS spectra suggests 49% conversion.
Table S4. Effect of additives with competing groups on the efficiency of reaction.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Additive (S11, S12, S13)</th>
<th>Equiv.</th>
<th>Ubiquitin (2a to 4m) % Conversiona</th>
<th>Lysozyme C (2b to 4n) % Conversiona</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>None</td>
<td>-</td>
<td>46</td>
<td>77</td>
</tr>
<tr>
<td>2</td>
<td>Fmoc-Cys-NH2 (S11)</td>
<td>100</td>
<td>22</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>Fmoc-Cys-NH2 (S11)</td>
<td>200</td>
<td>20</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>Fmoc-Cys-NH2 (S11)</td>
<td>250</td>
<td>20</td>
<td>42</td>
</tr>
<tr>
<td>5</td>
<td>Fmoc-Lys-NH2 (S12)</td>
<td>100</td>
<td>37</td>
<td>62</td>
</tr>
<tr>
<td>6</td>
<td>Fmoc-Lys-NH2 (S12)</td>
<td>200</td>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>Fmoc-Lys-NH2 (S12)</td>
<td>250</td>
<td>34</td>
<td>58</td>
</tr>
<tr>
<td>8</td>
<td>Fmoc-His-NH2 (S13)</td>
<td>100</td>
<td>24</td>
<td>59</td>
</tr>
<tr>
<td>9</td>
<td>Fmoc-His-NH2 (S13)</td>
<td>200</td>
<td>21</td>
<td>58</td>
</tr>
<tr>
<td>10</td>
<td>Fmoc-His-NH2 (S13)</td>
<td>250</td>
<td>20</td>
<td>57</td>
</tr>
</tbody>
</table>

a %Conversions were estimated by MALDI-ToF.

We performed the control experiments to understand the potential inhibitory effect of additives. We selected Cys (S11), Lys (S12), and His (S13) as additives in the reaction of cyclohexenone with ubiquitin (2a) and lysozyme C (2b) (Table S4). Interestingly, the additives (S11-S13, up to 250 equivalents) reduce the conversions but do not stop the reaction completely.

**Stability of the product under reaction conditions:**

The bioconjugation providing different bond strengths are desired to cater diverse applications.10 The cyclohexenone adduct of a protein exhibits reversibility with time and is appropriate for reversible covalent chemistry.11 The stabilization of the labeled proteins can be achieved conveniently by their conversion to oxime derivatives. The latter exhibits no detectable reversibility even within 96 h (Figure S2).
**Figure S2.** Stability of the labeled protein under reaction conditions.

**Cyclohexenone versus Ni-NTA chromatography of the protein with a His-tag:**

The cell lysate containing overexpressed His-tagged Ubc9 was divided into two parts. The purification of the sample was done using cyclohexenone (for details, see section 2.4) and the Ni-NTA chromatography\textsuperscript{12,13} (gel image, Figure S3). We investigated the Ni contamination in both the purified samples. As anticipated, the sample purified by Ni-NTA chromatography was contaminated with Ni (0.04\% by weight, EDX analysis), whereas the cyclohexanone-based method delivered Ni-free Ubc9.

![Image of Coomassie stained SDS-PAGE](image)

**Figure S3.** Comparison of metal-free affinity chromatography and Ni-NTA chromatography: 12\% SDS-PAGE followed by Coomassie staining utilized for quantification. The Ubc9 (lane 1 and 2) appeared as band (\(~20\) kDa). MW - Molecular Weight, SDS-PAGE: lane 1 - Ubc9 purified by Ni-NTA chromatography, lane 2 - Ubc9 purified by 3m, lane 3 - ladder.
5. Structural analysis
5.1 Stability of 2-cyclohexenone (3m)

Figure S4. The $^1$H NMR recorded in the phosphate buffer (0.1 M, pH 7.0) and 10% D$_2$O. The results confirm the stability of 2-cyclohexenone (3m) in the aqueous buffer.
5.2. Circular dichroism spectroscopy

Figure S5. Unaltered secondary structure of protein in given reaction conditions in phosphate buffer (0.1 M, pH 7.0) confirmed by circular dichroism (CD) experiments. The spectra of (a) native ubiquitin (2a, black line) and monolabeled ubiquitin (4m, red line) at concentration 10.0 μM. (b) Native lysozyme C (2b, black line) and monolabeled lysozyme C (4n, red line) at concentration 10 μM. (c) Native insulin (2c, black line) and monolabeled insulin (4o, red line) at concentration 5.0 μM. (d) Native α-lactalbumin (2d, black line) and monolabeled α-lactalbumin (4p, red line) at concentration 10.0 μM. (e) Native RNase A (2e, black line) and monolabeled RNase A (4q, red line) at concentration 10.0 μM. (f) Native β-lactoglobulin (2f, black line) and monolabeled β-lactoglobulin (4r, red line) at concentration 5.0 μM. (g) Native myoglobin (2g, black line) and monolabeled myoglobin (4s, red line) at concentration 5.0 μM.
5.3. Enzymatic assay

Enzymatic assay of lysozyme C

An enzymatic activity assay was carried out with lysozyme C before and after labeling. The cell lysis of Micrococcus lysodeikticus (ML) was monitored by absorbance at 450 nm (A450) using quartz cuvette (path length, 1 cm at 25 °C). Potassium phosphate buffer (pH 6.2, 0.1 M) was prepared by dissolving potassium phosphate, monobasic in the Millipore Grade I water. The pH was adjusted to 6.2 at 25 °C using 1 M potassium hydroxide (KOH) solution. Micrococcus lysodeikticus (ML) cells suspension [0.01% (w/v)] was prepared in the potassium phosphate buffer. The change in absorbance at 450 nm of this ML cells suspension versus a buffer blank was in agreement with the literature (0.6-0.7) after adjustment using the appropriate amount of buffer. Freshly prepared lysozyme C (2b) and labeled lysozyme C (4n) solutions (10 µg/1 ml, in phosphate buffer) were used for the assay. Lysozyme C (2b) solution (100 µl) and micrococcus lysodeikticus suspension (1 ml) were mixed by inversion. The sample was immediately used for recording the absorbance at 450 nm. The procedure was repeated for measuring the absorption of micrococcus lysodeikticus (ML) cells and labeled lysozyme C (4n) mixture. The enzymatic activity of lysozyme C (4n) was retained after the site-selective chemical modification.

Enzymatic assay of RNase A

An enzymatic activity assay was performed with RNase A and compared with the labeled RNase A. The hydrolysis of ribonucleic acid (RNA) was monitored at 300 nm (A300) using quartz cuvette (path length, 1 cm at 25 °C). Sodium acetate buffer (pH 5.0, 0.1 M) was prepared by using the Millipore Grade I water (pH was adjusted with 2 M acetic acid). Ribonucleic acid [RNA, 0.1% (w/v), 1 mg/ml] solution was prepared in sodium acetate buffer. The freshly prepared solutions of RNase A (2e) and labeled RNase A (4q) solutions (10 µg/1 ml, in Millipore Grade I water) were used for an assay. Initially, change in the absorption of RNA was monitored at 300 nm using the RNA solution and blank. The absorbance was immediately recorded after a gentle mixing of RNA (500 µl) and RNase A (2e) (500 µl). Next, the absorbance was recorded for the mixture of RNA and labeled RNase A (4q). The activity of the labeled RNase A remains unperturbed.
6. Late-stage modification\textsuperscript{18}

Figure S6. $^{19}\text{F}$-NMR spectrum of labeled lysozyme C (6a). For MS and MS-MS data, see Figure S21. $^{19}\text{F}$-NMR probe attached lysozyme C (6a) shows a sharp signal at -62.65 ppm. Trifluoroacetic acid (0.02 mM) was used as an internal standard, -75.45 ppm. The $^{19}\text{F}$-NMR was recorded in phosphate buffer (0.1 M, pH 7.0):D$_2$O (9:1).
**Figure S7.** Steady-state fluorescence spectra of native lysozyme C (2b) and coumarin tagged lysozyme C (6b). In phosphate buffer (0.1 M, pH 7.0), 6b exhibits absorption and emission band peaked at 336 nm and 420 nm, respectively. For MS and MS-MS data, see Figure S22.

**Figure S8.** Exploration of biotin tagged lysozyme C (6c) by SDS-PAGE and streptavidin (Stv-n) induced band-shift analysis. The incubation of biotin tagged lysozyme C (6c) and Stv-n were done at 25 °C for 10 minutes before subjecting to SDS-PAGE (lane 4). For MS and MS-MS data, see Figure S23. 12% SDS-PAGE followed by Coomassie staining identifies the biotin tagged lysozyme C (6c). The complex (lane 4) of streptavidin (Stv-n) and 6c appeared as band (~90 kDa). MW - Molecular Weight, SDS-PAGE: lane 1 - native lysozyme C (2b), lane 2 - Stv-n and 2b, lane 3 - streptavidin (Stv-n), and lane 4 - Stv-n and 6c.
7. MS data of proteins

Figure S9. Screening of soft electrophiles. MALDI-ToF-MS spectra for screening of soft electrophiles (3a-3m) with ubiquitin (2a).
Figure S10. MS-MS spectrum of labeled ubiquitin (4a) with iodoacetamide (3a). (a) Labeling of ubiquitin at H68. (b) Labeling of ubiquitin at M1 (Nα-NH₂).

Figure S11. MS-MS spectrum of labeled ubiquitin (4b) with 4-bromophenacyl bromide (3b). (a) Labeling of ubiquitin at H68.
Figure S12. MS-MS spectrum of labeled ubiquitin (4d) with ethyl vinyl sulfone (3d). (a) Labeling of ubiquitin at K6. (b) Labeling of ubiquitin at K11. (c) Labeling of ubiquitin at K48.
Figure S13. MS-MS spectrum of labeled ubiquitin (4i) with maleimide (3i). (a) Labeling of ubiquitin at H68. (b) Labeling of ubiquitin at M1 (N<sup>α</sup>-NH<sub>2</sub>).
Figure S14. Site-selective labeling of ubiquitin. (a) ESI-MS spectra for labeled ubiquitin (4m) from the reaction of ubiquitin (2a) and 2-cyclohexenone (3m). (b) MS-MS spectrum of labeled ESTHLVLR (E64-R72). Single-site modification at H68 results in monolabeled ubiquitin (4m) and its oxime derivative with alkoxyamine (5d).
Figure S15. Site-selective labeling of lysozyme C. (a) ESI-MS spectra for labeled lysozyme C (4n) from the reaction of lysozyme C (2b) and 2-cyclohexenone (3m). (b) MS-MS spectrum of labeled HGLDNYR (H15-R21). Single-site modification at H15 results in monolabeled lysozyme C (4n) and its oxime derivative with alkoxyamine (5d).
Figure S16. Site-selective labeling of insulin. (a) ESI-MS spectra for labeled insulin (4o) from the reaction of insulin (2c) and 2-cyclohexenone (3m). (b) MS-MS spectrum of labeled VNQHLCGSHEALY (V2-Y16). Single-site modification at H10 results in monolabeled insulin (4o) and its oxime derivative with alkoxyamine (5d).
Figure S17. Site-selective labeling of α-lactalbumin. (a) ESI-MS spectra for labeled α-lactalbumin (4p) from the reaction of α-lactalbumin (2d) and 2-cyclohexenone (3m). (b) MS-MS spectrum of labeled NDQDPHSSNICNISCDK (N63-K79). Single-site modification at H68 results in monolabeled α-lactalbumin (4p) and its oxime derivative with alkoxyamine (5d).
Figure S18. Site-selective labeling of RNase A. (a) ESI-MS spectra for labeled RNase A (4q) from the reaction of RNase A (2e) and 2-cyclohexenone (3m). (b) MS-MS spectrum of labeled HIIVACEGPNYPVPVHFDASV (H105-V124). Single-site modification at H105 results in monolabeled RNase A (4q) and its oxime derivative with alkoxyamine (5d).
Figure S19. Site-selective labeling of β-lactoglobulin. (a) ESI-MS spectra for labeled β-lactoglobulin (4r) from the reaction of β-lactoglobulin (2f) and 2-cyclohexenone (3m). (b) MS-MS spectrum of labeled ALPMHIR (A142-R148). Single-site modification at H146 results in monolabeled β-lactoglobulin (4r) and its oxime derivative with alkoxyamine (5d).
Figure S20. Site-selective labeling of myoglobin. (a) ESI-MS spectra for labeled myoglobin (4s) from the reaction of myoglobin (2g) and 2-cyclohexenone (3m). (b) MS-MS spectrum of labeled ISDAIIHVL (I107-L115). Single-site modification at H113 results in monolabeled myoglobin (4s) and its oxime derivative with alkoxyamine (5d).
Figure S21. Late-stage modification. (a) ESI-MS spectra for lysozyme C (2b) and modified lysozyme C (6a) after oxime formation with alkoxyamine (5a). (b) MS-MS spectrum of labeled HGLDNYR (H15-R21). Single-site modification at H15 results in monolabeled lysozyme C (6a).
Figure S22. Late-stage modification. (a) ESI-MS spectra for lysozyme C (2b) and modified lysozyme C (6b) after oxime formation with alkoxyamine (5b). (b) MS-MS spectrum of labeled HGLDNYR (H15-R21). Single-site modification at H15 results in monolabeled lysozyme C (6b).
Figure S23. Late-stage modification. (a) ESI-MS spectra for lysozyme C (2b) and modified lysozyme C (6c) after oxime formation with alkoxyamine (5c). (b) MS-MS spectrum of labeled HGLDNYR (H15-R21). Single-site modification at H15 results in monolabeled lysozyme C (6c).
Figure S24. Screening of additives for chemically-triggered reversibility. MALDI-ToF-MS spectra for screening of additives (7a-7g) with labeled lysozyme C (6d).
8. NMR data
9. Acknowledgements

P.N.J. is recipient of research fellowship from CSIR, India. The authors are thankful to Mr. Sajeev T. K. and Dr. Ram K. Mishra for providing the cell lysate. The authors are thankful to SERB, DAE, DBT, Ramanujan Fellowship, and IISER Bhopal for research grants.

10. References