Electronic Supplementary Material (ESI) for Chemical Communications.

Residue-specific N-terminal glycine to aldehyde transformation renders analytically pure single-site labelled protein

Tularam Sahu,[a] Mohan Kumar,[a] Sajeev T. K.,[b] Manas Joshi,[a] Ram Kumar Mishra[b] and Vishal Rai[a]*

[a]Department of Chemistry and [b]Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, India.

Email: vrai@iiserb.ac.in

Contents

1. General information

The reagents, proteins, and enzymes were purchased from Sigma-Aldrich, Alfa Aeser, Spectrochem, and Merck. The organic solvents used were reagent grade. Aqueous buffers were prepared freshly using Millipore Grade I water (Resistivity > 5 MΩ cm, Conductivity < 0.2 μS/cm, TOC <30 ppb). Mettler Toledo (FE20) pH meter was used to adjust the final pH. The reaction mixture for the small molecules was stirred (Heidolph, 500-600 rpm). Proteins were either vortexed or incubated in Thermo Scientific MaxQ 8000 incubator shaker (350 rpm, 25-37 ^{o}C). Cellulose membrane (MWCO, 6-8 kD) from Spectrum labs was used for dialysis. Amicon® Ultra-0.5 mL 3-10 kDa MWCO centrifugal filters from Merck Millipore were used to remove small molecules from the protein mixture, desalting, and buffer exchange. Organic solvents were removed by BUCHI rotavapor R-210/215 whereas aqueous samples were lyophilized by CHRiST ALPHA 2-4 LD plus lyophilizer. Circular Dichroism (CD) measurements were recorded on JASCO J-815 CD spectropolarimeter equipped with peltier temperature controller. All the spectra were measured with a scan speed of 50 nm/min, spectral band width 1 nm using 1 cm path length cuvette at 25 ^{o}C. Steady-state fluorescence spectra was carried out in HORIBA JOBIN YVON, FLUOROLOG 3-111. The fluorescence spectra were measured with a quartz cuvette of 1 cm path length.

Chromatography: Thin-layer chromatography (TLC) was performed on silica gel coated aluminium TLC plates (Merck, TLC Silica gel 60 F_{254}) and visualized using a UV lamp (254 nm) and stains such as iodine, ninhydrin, 2,4-dinitrophenylhydrazine. The flash column chromatography was carried out on Combiflash NextGen 300+, Combiflash Rf 200, or gravity columns using Merck′s 230-400 or 100-200 mesh silica gel.

Nuclear magnetic resonance spectra: ^{1}H, ^{13}C, and ^{19}F NMR spectra were recorded on Bruker Avance III 400 and 500 MHz NMR spectrometer. ^{1}H NMR spectra were referenced to TMS (0 ppm), DMSO-d_6 (2.50 ppm) and MeOH-d_4 (3.31 ppm), ^{13}C NMR spectra were referenced to $CDCl_3$ (77.16 ppm), MeOH-d_4 (49.01 ppm) and DMSO-d_6 (39.52 ppm). Peak multiplicities are designated by the following abbreviations: s, singlet; bs, broad singlet; d,

doublet; t, triplet; q, quartet; m, multiplet; dd, doublet of doublets; ddd, doublet of doublet of doublets. Spectra were recorded at 298 K.

Mass spectrometry: SCIEX X500B QToF coupled with ExionLC AD UPLC and Agilent 6130 single quad coupled with Agilent 1200 series HPLC (ESI/APCI) was used for LC-MS and protein sequencing. Poroshell 300 SB-C18 HPLC column (2.1 × 75 mm × 5 μm, flow rate 0.4 ml/min) and XB-C18 UPLC column (2.5 x 150 mm, 1.7 μm, 100 Å, flow rate 0.3 ml/min) were used for small molecules and protein derived samples respectively. HRMS data were recorded on Bruker Daltonics MicroTOF-Q-II with electron spray ionization (ESI). Data analysis was performed using SCIEX Bio-pharma view Flex, Flex analysis, and Bruker data analysis software. Peptide mass and fragment ion calculator (http://db.systemsbiology.net:8080/proteomicsToolkit/ FragIonServlet.html) were used for peptide mapping and sequencing.

Acetonitrile and H_2O were buffered with 0.01% and 0.1% formic acid used as mobile phase in single quad and QTOF respectively. Method A was used to record the LC-ESI-MS data for proteins on Agilent 6130 Quadrupole paired with agilent 1200 infinity series HPLC and method B used for peptide mapping and MS/MS of digested proteins on SCIEX X500B paired with ExionLC AD UHPLC. Method C was used to record the LC-ESI-MS data for intact protein and mixture of proteins on SCIEX X500B paired with ExionLC AD UHPLC. Methods D was used for intact mass and MSMS of peptides on SCIEX X500B paired with ExionLC AD UHPLC.

Method A (Column: Agilent, Poroshell 300 SB-C18 5 μm 2.1 × 75 mm, flow rate 0.4 ml/min)

Time (min)	**Acetonitrile (%)**	**H_2O (%)**
0	10	90
1	10	90
8	60	40
12	90	10
15	90	10

Method B (Column: Phenomenex, Poroshell 100 Å, 2.5 x 150 mm, 1.7 μm, flow rate 0.3 ml/min)

Time (min)	**Acetonitrile (%)**	**H_2O (%)**
0	5	95
2	5	95
25	50	50
26	80	20
28	5	95

30	5	95

Method C (Column: Phenomenex, Poroshell 300 Å, 4.6 x 150 mm, 1.8 µm, flow rate 0.5 ml/min)

Time (min)	Acetonitrile (%)	H_2O (%)
0	5	95
2	5	95
35	30	70
45	50	50
47	80	20
48	80	20
49	5	95

Reaction conversion determination for protein labeling

ESI-MS: Conversion for protein labeling was calculated based on the relative peak intensity of native protein and labeled protein in the deconvoluted mass spectrum.

% Conversion = $I_{\text{desired product}}$ /$I_{\text{all relevant species}}$ where $I_{\text{desired product}}$ is the peak intensity of labeled protein, and $I_{\text{all relevant species}}$ is the sum of the peak intensities of native protein and labeled protein in the deconvoluted mass spectra.

Method D: (Column: Phenomenex, bioZen, Peptide PS-C18 (2.1 x 150 mm, 3 µm, 100 Å, flow rate 0.5 ml/min)

Time (minutes)	H_2O (%)	Acetonitrile (%)
0	90	10
3	90	10
5	50	50
10	50	50
12	10	90
15	90	10

2. Experimental methods

2.1 Procedure for protein labelling

In a 1.5 mL microcentrifuge tube, protein (**1a-c**, 3 nmol) was mixed with sodium bicarbonate buffer (120 µl, 0.1 M, pH 7.8). To this solution 2-(2-formylphenoxy) acetic acid (**2**, 1.5 mmol) in DMSO (30 µl) from a freshly prepared stock solution was added and vortexed at 25 °C. The overall concentration of protein and -(2-formylphenoxy) acetic acid was 20 µM and 10 mM respectively. After 24–48 h, NH_2OH (3 mM) was added and

vortexed. After 1 h the reaction mixture was diluted with acetonitrile:water (10:90, 1000 µl) and excess of reagent and salts were removed by using Amicon® Ultra-0.5 mL 3-kDa or 10-kDa MWCO centrifugal spin concentrator. The protein mixture was further washed with Millipore Grade I water (2 × 0.4 mL). The sample was analyzed by LC-ESI-MS. The sample was then buffer (PB, 0.02 M, pH 7.0, 130 µl,) exchanged. To this solution, $NaIO_4$ in buffer (20 µl, 0.4 mM) was added and vortexed at 25 °C. After 15 min excess of $NaIO_4$ was quenched by adding ethylene glycol (1 mM) and vortexed for 5 min. The excess of reagent was removed by using Amicon® Ultra-0.5 mL 3-kDa or 10-kDa MWCO centrifugal spin concentrator and sample was re-suspended in phosphate buffer (140 µl, 0.02 M, pH 7.0). To this solution, hydroxylamine (10 mM) in DMSO (10 µl) from a freshly prepared stock solution was added for late-stage modification (Oxime formation) and vortexed for 1 h. The excess of O-hydroxylamine and salts were removed by spin concentrator (0.5 mL 3-kDa or 10-kDa MWCO) and the sample was collected in an aqueous medium. Modification of protein was analyzed by LC-ESI-MS. The aqueous sample was concentrated by lyophilization before subjecting it to digestion, peptide mapping, and sequencing by MS-MS.

Procedure for labelling of insulin in DMEM cell culture media:

In a 1.5 mL microcentrifuge tube, protein (**1a**, 5 nmol) was mixed with DMEM cell culture media (L0104-500, Biowest) (200 µl, pH 7.8). To this solution 2-(2-formylphenoxy) acetic acid (**2**, 5 mmol) in DMSO (50 µl) from a freshly prepared stock solution was added and vortexed at 25 °C. After 24 , NH_2OH (5 mM) was added and vortexed. After 1 h the reaction mixture was diluted with acetonitrile:water (10:90, 1000 µl) and excess of reagent and salts were removed by using Amicon® Ultra-0.5 mL 3-kDa centrifugal spin concentrator. The protein mixture was further washed with Millipore Grade I water (2 × 0.4 mL). The sample was analyzed by LC-ESI-MS. The sample was then buffer (PB, 0.02 M, pH 7.0, 0.25 M NaCl 100 µl,) exchanged. To this solution, $NaIO_4$ in buffer (10 µl, 50 mM) and L-Met (2.5 µl, 100 mM)was added and vortexed at 25 °C. After 5 min excess of $NaIO_4$ was quenched by addition of PB buffer (400 µl, 0.1 M, pH 7.0). The excess of reagent was removed by using Amicon® Ultra-0.5 mL 3-kDa or 10-kDa MWCO centrifugal spin concentrator and sample analysed by LC-ESI-MS.

2.2 Purification of the labeled insulin

Hydrazide beads **8** (200 µl, hydrazide resin loading: 16 µmol/ml) were taken in a 5 ml fritted polypropylene chromatography column with end tip closures. Phosphate buffer (0.5 M, pH 7.0, 5 × 1 ml) was used to wash the beads and re-suspended in phosphate buffer (100 µl, 0.1 M, pH 7.0, 0. 5 M NaCl). Protein mixture (**1a** and N-glyoxamide insulin **4a**, 200 µM) in phosphate buffer (200 µl, 1 M, pH 6.0) and aniline (10 mM) in phosphate buffer (100 µl, 0.1 M, pH 7.0, 0. 5 M NaCl) were added to the beads followed by end-to-end rotation (30 rpm, rotary mixer) at 25 ^{o}C for 12 h. The immobilization of the labeled protein on hydrazide resin was monitored by UV-absorbance of the supernatant (complete loading of labeled protein was also confirmed by ESI-MS of supernatant). After collecting the supernatant, the beads were washed with phosphate buffer (0.1 M, pH 7.0, 5 x 1 ml) and KCl (1 M, 4 x 1 ml) to remove the adsorbed insulin from resin. The beads were further washed with buffer (0.1 M, pH 7.0, 3 x 1 ml) and re-suspended in ammonium acetate buffer (200 µl, 0.1 M, pH 4.5). To release the labeled protein from its immobilized derivative was released by adding hydroxylamine derivatives (50 µl, DMSO) (end-to-end rotation at 25 ^{o}C for 2-12 h). The supernatant was collected and then the salts and O-hydroxylamine were removed using the spin concentrator (3 kDa or 10-kDa MWCO). The purity of the labeled insulin **5a, 7a', 7d** and **7e** was confirmed by ESI-MS.

2.3 Digestion protocol

All the solutions were prepared freshly before use in reactions.

Reduction of disulfide bond in modified insulin

Protein (0.1 mg) in 100 µL of H_2O was taken in a 1.5 ml Eppendorf tube. To reduce the disulfide bond between chain A and chain B, dithiothreitol (10 µL, 0.2 M DTT in H_2O) was added. The reaction mixture was incubated at 37°C for 3-6 h. The N-terminal modification of chain A was confirmed by LC-MS/MS analysis.

Procedure for peptide mapping and MS/MS of modified melittin

Protein (0.1 mg) in 200 µL of H_2O was taken in a 1.5 ml Eppendorf tube and recorded peptide mapping and LC-MS/MS data which confirms the N-terminal modification.

Procedure for in-solution digestion of myoglobin

Protein (0.1 mg) was taken in a 1.5 ml microcentrifuge tube. To this solution, urea (6 M, 10 µl) and tert-butanol (10 µl) was added and incubated for 1 h at 37 ^{o}C. The urea concentration of sample was reduced desalted by using Amicon® Ultra-0.5 mL 3-kDa or 10-kDa MWCO centrifugal spin concentrator and sample was re-suspended in 100 grade I water. To this

solution, 5 µl of enzyme (α-chymotrypsin/trypsin) solution was added and the mixture was incubated at 37 °C for 18 h. The pH of digested solution was adjusted to < 6 (verified by pH paper) with trifluoroacetic acid (0.5%). Subsequently, the sample was used for peptide mapping by MS and sequencing by MS-MS.

2.4 Insulin bioactivity assay

Immunofluorescence

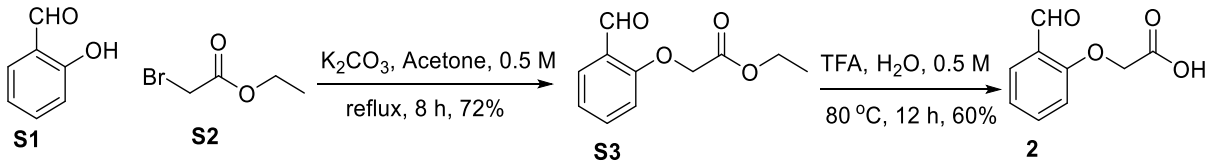
HeLa cells were grown on coverslips in a six-well plate format in Dulbecco's modified Eagle's medium (DMEM) and serum-starved for 24 h in 1% serum-containing DMEM medium. Subsequently, cells were treated with 50 µM LY294002 for 30 minutes. After LY294002 treatment, cells were washed twice with PBS and treated with either native insulin or purified insulin (3 µg each) in 1 ml of 1% FBS containing DMEM media for 30 minutes. Post-treatment, cells were again washed twice with PBS and fixed using 100% chilled methanol for 15 min at 4 °C. The cells were then rehydrated and permeabilized with rehydration buffer (10 mM Tris, 150 mM NaCl, 0.1% TritonX-100) for 10 min. Cells were blocked with 5% Normal Goat Serum (NGS) for 1 h at 4 °C after rehydration. The cells were stained for 2 h with pAkt antibodies (1:600, anti-pAKT-S473, CST #4060) and Actin antibodies (1:1000, BD #612656) at 4 °C. After primary antibody incubation, cells were washed three times with PBS-T (5 minutes each) and incubated with Alexa Fluor-488 conjugated goat anti-rabbit IgG against pAkt (1:1000, Thermo #A11034) and Alexa Fluor-568 conjugated goat anti-mouse IgG against Actin (1:1000, Thermo #A11004) for 1 h. Cells were washed thrice with PBS-T and mounted on slides using DAPI containing mounting medium (SIGMA F6057). Fluorescence signals were captured on Zeiss LSM 780 confocal microscope, and images were analyzed using Image J software.

Western blotting

HEK293T cells were cultured and treated as mentioned in the immunofluorescence section. Post-treatment, cells were washed twice with PBS and lysed directly in the 1X Laemmli buffer (containing 1 mM Sodium orthovanadate and 1mM PMSF) by boiling at 100 °C for 10 minutes. The cleared lysates were obtained by centrifugation and analysed on 8% SDS-PAGE. Following wet transfer protocols, proteins were transferred onto the methanol-activated 0.2 µm PVDF membrane (Merck, Cat #ISEQ85R) using 1X transfer buffer (2.5 mM Tris-HCl pH 7.5, 19.2 mM Glycine). The membrane was blocked for 1 h in 5% BSA. Further, the membrane was incubated overnight with anti-pAkt-S473 (1:2000, CST #4060)

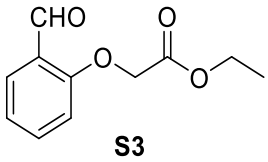
and anti-Actin (1:8000, BD #612656) antibodies. The membrane was later washed three times for 10 minutes each with TBS-T buffer (20 mM Tris HCl pH 7.5, 150 mM NaCl, 0.1% Tween-20). The membrane was incubated with Alexa fluor Plus 680 secondary antibodies (Invitrogen# A32734) for 1 hour. Later it was washed with TBS-T buffer (three times for 10 minutes each). Images were taken using LI-COR IR system. Anti-Actin (1:8000, BD #612656) antibodies were used for loading control detection. Alexa fluor Plus 800 secondary antibodies were used for Actin detection.

3. Synthesis of reagents



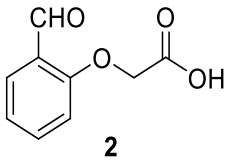
Scheme S1. Synthesis of 2-(2-formylphenoxy) acetic acid (**2**).[1]

Synthesis of ethyl 2-(2-formylphenoxy) acetate (S3)



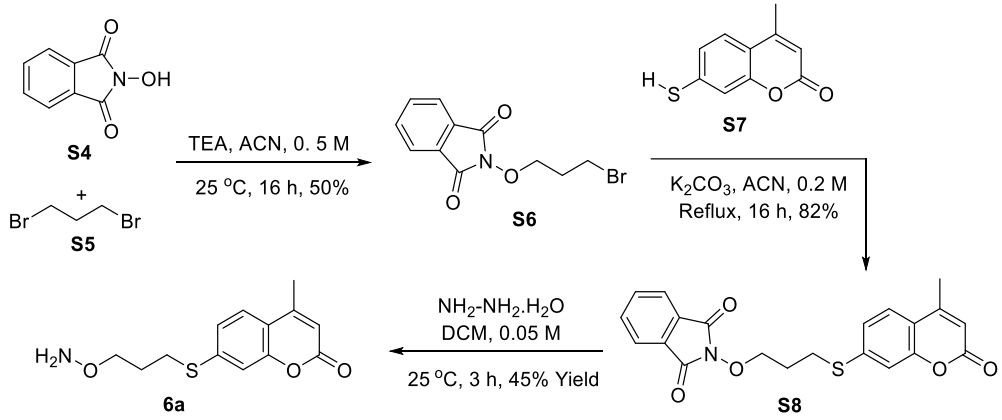
In a 25 ml round bottom flask, 2-hydroxy benzaldehyde **S1** (366 mg, 3 mmol), ethyl 2-bromoacetate **S2** (1 g, 6 mmol) and K_2CO_3 (828 mg, 6 mmol) was dissolved in acetone (6 ml). The reaction mixture was allowed to reflux for 8 h. The reaction was monitored using thin layer chromatography and upon completion, the reaction mixture was filtered to remove potassium carbonate. The solution was concentrated under vacuum and the product was purified using flash column chromatography using ethyl acetate:n-hexane (2:98) to afford ethyl 4-(4-formylphenoxy) acetate **S3** (72% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 10.55 (s, 1H), 7.85 (dd, *J* = 7.7, 1.8 Hz, 1H), 7.51 (m, 1H), 7.06 (t, *J* = 7.5 Hz, 1H), 6.85 (d, *J* = 8.4 Hz, 1H), 4.74 (s, 2H), 4.26 (q, *J* = 7.1 Hz, 2H), 1.28 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 189.5, 168.1, 160.1, 135.7, 128.6, 125.4, 121.8, 112.6, 65.6, 61.6, 14.1 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{11}H_{12}NaO_4$ 231.0633, found 231.0656.

Synthesis of 2-(2-formylphenoxy) acetic acid (2)



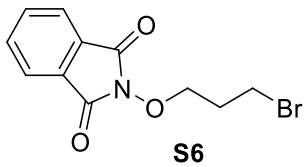
2

In a 25 ml round bottom flask, ethyl 4-(4-formylphenoxy) acetate **S3** (440 mg, 2.1 mmol), was mixed with water (4 ml). To this solution trifluoro acetic acid (964 mg, 8.4 mmol) was added drop wise at 0-5 ^{o}C. The reaction mixture was allowed to reflux for 12 h. The reaction was monitored using thin layer chromatography and upon completion of the reaction, the solution was concentrated under vacuum and the product was purified using flash column chromatography using ethyl acetate:n-hexane (50:50) to afford ethyl 4-(4-formylphenoxy) acetate **2** (63% yield). ^{1}H NMR [500 MHz, $(CD_3)_2CO$] δ 10.55 (s, 1H), 7.78 (dd, *J* = 7.7, 1.8 Hz, 1H), 7.63 (m, 1H), 7.19 (d, *J* = 8.5 Hz, 1H), 7.15-7.09 (m, 1H), 4.94 (s, 2H) ppm. ^{13}C NMR [126 MHz, $(CD_3)_2CO$] δ 189.6, 169.7, 161.4, 136.7, 128.5, 126.3, 122.3, 114.5, 65.9 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_9H_8NaO_4$ 203.0320, found 203.0330.



Scheme S2. Synthesis of 7-((3-(aminooxy)propyl)thio)-4-methyl-2H-chromen-2-one (**6a**).[1]

Synthesis of 2-(3-bromopropoxy)isoindoline-1,3-dione (S6)

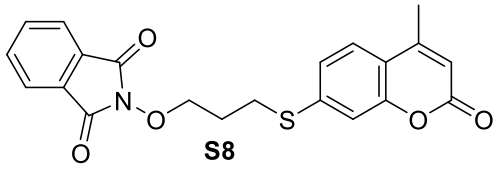


S6

In a 250 ml round bottom flask, N-hydroxyphthalimide **S4** (4894 mg, 30 mmol) was taken and dissolved in ACN (60 ml). To this solution triethyl amine (6.1 ml, 60 mmol) and 1,3-dibromo propane **S5** (8.3 ml, 60 mmol) was added and stirred at 25 ^{o}C for 16 h. After this reaction mixture was concentrated in vacuo and 1 N NaOH solution and ethyl acetate was added. The organic layer was separated, dried over anhydrous sodium sulphate and

concentrated in vacuo. The crude mixture was purified by flash column chromatography using ethyl acetate:n-hexane (3:97) gave **S6** (50% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ ^{1}H NMR (400 MHz, $CDCl_3$) δ 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, $CDCl_3$) δ 163.7, 134.7, 129.0, 123.7, 76.2, 31.6, 29.4 ppm. MS (ESI) $[M+H]^+$ calcd. $C_{11}H_{11}{}^{79}BrNO_3$ 284.0, found 284.0 and calcd. $C_{11}H_{11}{}^{81}BrNO_3$ 286.0, found 285.9.

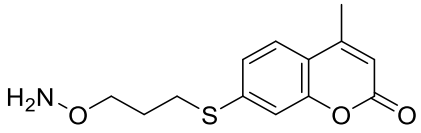
Synthesis of 2-(3-((4-methyl-2-oxo-2H-chromen-7-yl)thio)propoxy)isoindoline-1,3-dione (S8)



S8

In a 25 ml round bottom flask, 7-mercapto-4-methylcoumarin **S7** (192 mg, 1 mmol), K_2CO_3 (276 mg, 2 mmol) and 2-(3-bromopropoxy)isoindoline-1,3-dione **S6** (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:n-hexane (7:3) to give **S8** (82% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.87-7.84 (m, 2H), 7.78-7.75 (m, 2H), 7.48 (d, *J* = 8.2 Hz, 1H), 7.22 (m, 2H), 6.22 (d, *J* = 0.8 Hz, 1H), 4.36 (t, *J* = 5.8 Hz, 2H), 3.35 (t, *J* = 7.1 Hz, 2H), 2.41 (d, *J* = 0.9 Hz, 3H), 2.16-2.13 (m, 2H) ppm. ^{13}C NMR (101 MHz, $CDCl_3$) δ 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, 76.6, 28.5, 27.7, 18.6 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{21}H_{18}NO_5S$ 396.0906, found 396.0925.

Synthesis of 7-((3-(aminooxy)propyl)thio)-4-methyl-2H-chromen-2-one (6a)



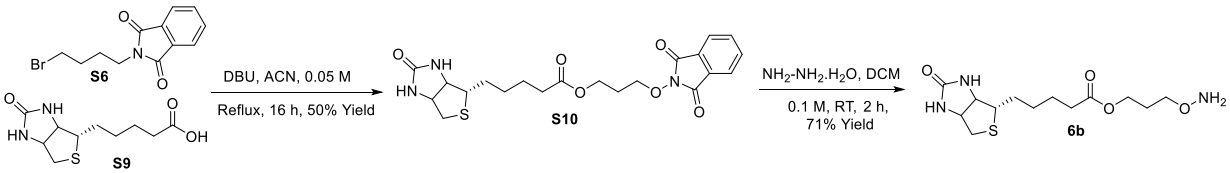
6a

2-(3-((4-methyl-2-oxo-2H-chromen-7-yl)thio)propoxy)isoindoline-1,3-dione **S8** (237 mg, 0.6 mmol) in 5 ml round bottom flask was dissolved in CH_2Cl_2 (12 ml). To this solution, hydrazine monohydrate (80%, 29 µl, 0.6 mmol) was added and stirred at 25 ^{o}C for 3 h. The reaction mixture was filtered and the filtrate was concentrated. Purification of crude mixture by reverse phase preparative HPLC gave **6a** (45% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.44 (d, *J* = 8.3 Hz, 1H), 7.26-7.13 (m, 2H), 6.18 (d, *J* = 0.9 Hz, 1H), 3.77 (t, *J* = 5.9 Hz, 2H), 3.05 (t, *J* = 7.3 Hz, 2H), 2.38 (d, *J* = 0.8 Hz, 3H), 2.02-1.93 (m, 2H) ppm. ^{13}C NMR

(101 MHz, $CDCl_3$) δ 160.7, 154.0, 152.3, 143.3, 124.7, 123.1, 117.2, 114.1, 113.9, 73.9, 29.0, 27.8, 18.6 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{13}H_{16}NO_3S$ 266.0851, found 266.0841.

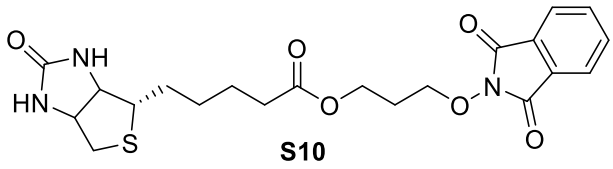
Synthesis of 7-(3-(aminooxy)propoxy)-4-methyl-2H-chromen-2-one (6a')

Procedure for synthesis of **6a'** is same as **6a.** ^1H NMR (500 MHz, CDCl3) δ 7.49 (d, J = 8.8 Hz, 1H), 6.86 (dd, J = 8.8, 2.5 Hz, 1H), 6.82 (d, J = 2.0 Hz, 1H), 6.13 (s, 1H), 4.11 (t, J = 6.3 Hz, 2H), 3.86 (t, J = 6.1 Hz, 2H), 2.40 (s, 3H), 2.11 (m, J = 6.2 Hz, 2H) ppm. ^{13}C NMR (126 MHz, CDCl3) δ 162.21 (s), 161.54 (s), 155.47 (s), 152.73 (s), 125.67 (s), 113.74 (s), 112.80 (s), 112.12 (s), 101.61 (s), 72.24 (s), 65.62 (s), 28.33 (s), 18.85 (s) ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{13}H_{16}NO_3S$ 249.1001, found 249.1010.



Scheme S3. Synthesis of 3-(aminooxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate (**6b**).[2]

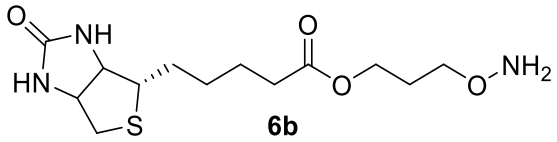
Synthesis of 3-((1,3-dioxoisoindolin-2-yl)oxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate (S10)



In 50 ml round bottom flask, biotin **S9** (244 mg, 1mmol), 2-(3-bromopropoxy)isoindoline-1,3-dione **S6** (568 mg, 2 mmol) and DBU (304 µl, 2 mmol) were dissolved in acetonitrile (20 ml) to reflux. The progress of the reaction was analyzed by TLC. After 16 h, reaction mixture was concentrated on vacuum and carried out for ethyl acetate and water work up. The organic layer was collected and dried on anhydrous sodium sulfate, filtered and concentrated on rotary evaporator. Purification of crude reaction mixture by flash chromatography (MeOH/DCM, 0.5-5%) gave 3-((1,3-dioxoisoindolin-2-yl)oxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate **S10** (50% yield). ^1H NMR (500 MHz, $CDCl_3$) δ 7.87-7.80 (m, 2H), 7.78-7.71 (m, 2H), 5.95 (s, 1H), 5.46 (s, 1H), 4.48 (dd, *J* = 15.0, 9.8 Hz, 1H), 4.38-4.23 (m, 5H), 3.20-3.11 (m, 1H), 2.89 (dd, *J* = 12.8, 5.0 Hz,

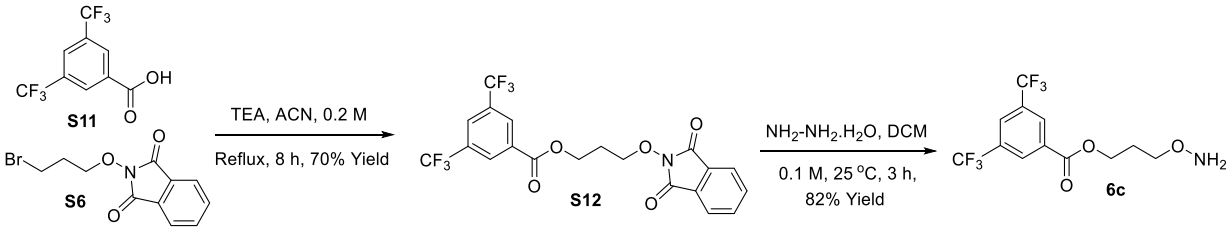
1H), 2.72 (d, *J* = 12.8 Hz, 1H), 2.34 (t, *J* = 7.5 Hz, 2H), 2.16-2.05 (m, 2H), 1.80-1.60 (m, 4H), 1.53-1.37 (m, 2H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 173.7, 163.7, 163.7, 134.7, 128.9, 123.7, 75.1, 62.0, 60.7, 60.2, 55.5, 40.7, 33.9, 28.4, 28.3, 27.8, 24.9 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{21}H_{26}N_3O_6S$ 448.1542, found 448.1548.

Synthesis of 3-(aminooxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate (6b)



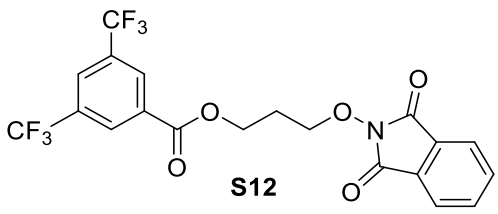
6b

In a 5 ml round bottom flask, 3-((1,3-dioxoisoindolin-2-yl)oxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate **S10** (134 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 2 h, the reaction mixture was filtered and concentration of the filtrate in vacuo led to the isolation of 3- (aminooxy)propyl 5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate **6b** (71% yield). ^{1}H NMR (500 MHz, D_2O) δ 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_2O) δ176.9, 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. $C_{13}H_{24}N_3O_4S$ 318.1488, found 318.1467.



Scheme S4. Synthesis of 3-(aminooxy)propyl 3,5-bis(trifluoromethyl)benzoate **(6c)**[2]**.**

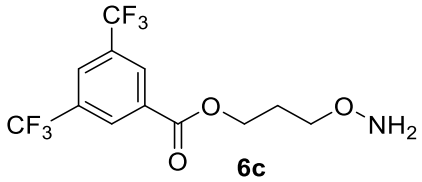
Synthesis of 3-((1,3-dioxoisoindolin-2-yl)oxy)propyl 3,5-bis(trifluoromethyl)benzoate (S12)



S12

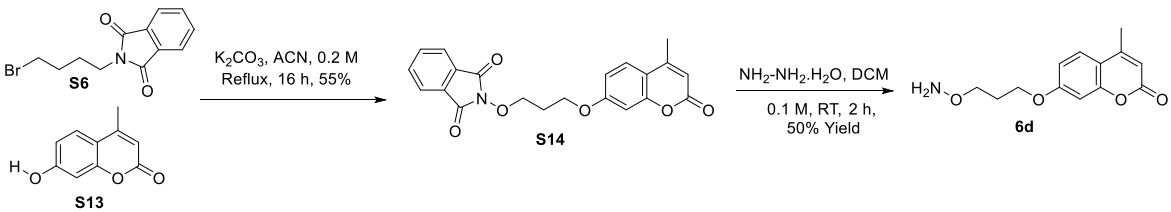
In a 25 ml round bottom flask, 3,5-bis(trifluoromethyl)benzoic acid **S11** (258 mg, 1 mmol), 2-(3-bromopropoxy)isoindoline-1,3-dione **S6** (312 mg, 1.1 mmol) and TEA (418 µl, 3 mmol) were dissolved in acetonitrile (5 ml) to reflux. Progress of the reaction was followed by TLC. After 8 h, reaction mixture was concentrated and purification of crude by flash column chromatography (ethyl acetate:n-hexane, 2:98) gave 3-((1,3-dioxoisoindolin-2-yl)oxy)propyl 3,5-bis(trifluoromethyl)benzoate **S12** (73% yield).). ^{1}H NMR (500 MHz, $CDCl_3$) δ 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, $CDCl_3$) δ 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, $CDCl_3$) δ -62.94 (TFA was used as an internal standard, -75.70 ppm). HRMS (ESI) $[M+H]^+$ calcd. $C_{20}H_{14}F_6NO_5$ 462.0776, found 462.0775.

Synthesis of 3-(aminooxy)propyl 3,5-bis(trifluoromethyl)benzoate (6c)



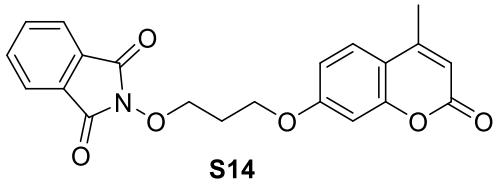
6c

In a 5 ml round bottom flask, 3-((1,3-dioxoisoindolin-2-yl)oxy)propyl 3,5-bis(trifluoromethyl)benzoate **S12** (138 mg, 0.3 mmol) in DCM (3 ml) was added hydrazine monohydrate (80%, 37 µl, 0.75 mmol) and stirred at room temperature. The progress of the reaction was followed by TLC. After 3 h, reaction mixture was filtered and concentration of filtrate in vacuo gave 3-(aminooxy)propyl 3,5-bis(trifluoromethyl)benzoate **6c** (81% yield). 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. ^{19}F NMR (376 MHz, $CDCl_3$) δ -62.54 (TFA was used as an internal standard, -75.70 ppm). HRMS (ESI) $[M+H]^+$ calcd. $C_{12}H_{12}F_6NO_3$ 332.0721, found 332.0699.



Scheme S4. 7-(3-(aminooxy)propoxy)-4-methyl-2H-chromen-2-one **(6d)**[2].

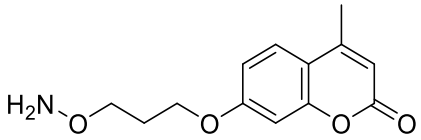
Synthesis of 2-(3-((4-methyl-2-oxo-2H-chromen-7-yl)oxy)propoxy)isoindoline-1,3-dione (S14)



S14

In a 25 ml round bottom flask, 7-hydroxy-4-methyl-2H-chromen-2-one **S13** (176 mg, 1 mmol), K2CO3 (276 mg, 2 mmol), and 2-(3-bromopropoxy)isoindoline-1,3-dione **S6** (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 **S14** (454 mg, 60% yield). ^{1}H NMR (500 MHz, CDCl3) δ 7.84 (m, 2H), 7.76 (m, 2H), 7.51 (d, J = 8.8 Hz, 1H), 6.90 (dd, J = 8.7, 2.5 Hz, 1H), 6.87 (d, J = 2.4 Hz, 1H), 6.14 (s , 1H), 4.43 (t, J = 6.0 Hz, 2H), 4.33 (t, J = 6.1 Hz, 2H), 2.40 (d, J = 1.0 Hz, 3H), 2.29 (m, 2H) ppm. ^{13}C NMR (126 MHz, CDCl3) δ 163.8, 162.1, 161.6, 155.4, 152.7, 134.8, 129.1, 125.8, 123.8, 113.9, 112.6, 112.2, 102.0, 75.1, 64.8, 28.3, 18.9.ppm. (ESI) [M+H]+ calcd. C21H18NO6 380.1134, found 380.1129.

Synthesis of 7-(3-(aminooxy)propoxy)-4-methyl-2H-chromen-2-one (6d)



6d

7-(3-(aminooxy)propoxy)-4-methyl-2H-chromen-2-one **S14** (0.6 mmol) in 5 ml round bottom flask was dissolved in CH_2Cl_2 (12 ml). To this solution, hydrazine monohydrate (80%, 29 µl, 0.6 mmol) was added and stirred at 25 ^{o}C for 3 h. The reaction mixture was filtered and the filtrate was concentrated. Purification of crude mixture by reverse phase preparative HPLC gave **6d** (50% yield)**.** ^{1}H NMR (500 MHz, CDCl3) δ 7.49 (d, J = 8.8 Hz, 1H), 6.86 (dd, J = 8.8, 2.5 Hz, 1H), 6.82 (d, J = 2.0 Hz, 1H), 6.13 (s, 1H), 4.11 (t, J = 6.3 Hz, 2H), 3.86 (t, J = 6.1 Hz, 2H), 2.40 (s, 3H), 2.11 (m, J = 6.2 Hz, 2H) ppm. ^{13}C NMR (126 MHz, CDCl3) δ 162.21 (s), 161.54 (s), 155.47 (s), 152.73 (s), 125.67 (s), 113.74 (s), 112.80 (s), 112.12 (s), 101.61 (s), 72.24 (s), 65.62 (s), 28.33 (s), 18.85 (s) ppm. HRMS (ESI) $[M+H]^{+}$ calcd. $C_{13}H_{16}NO_3S$ 250.1001, found 249.1010.

4. Additional results and discussion

4.1 Spectral Analysis

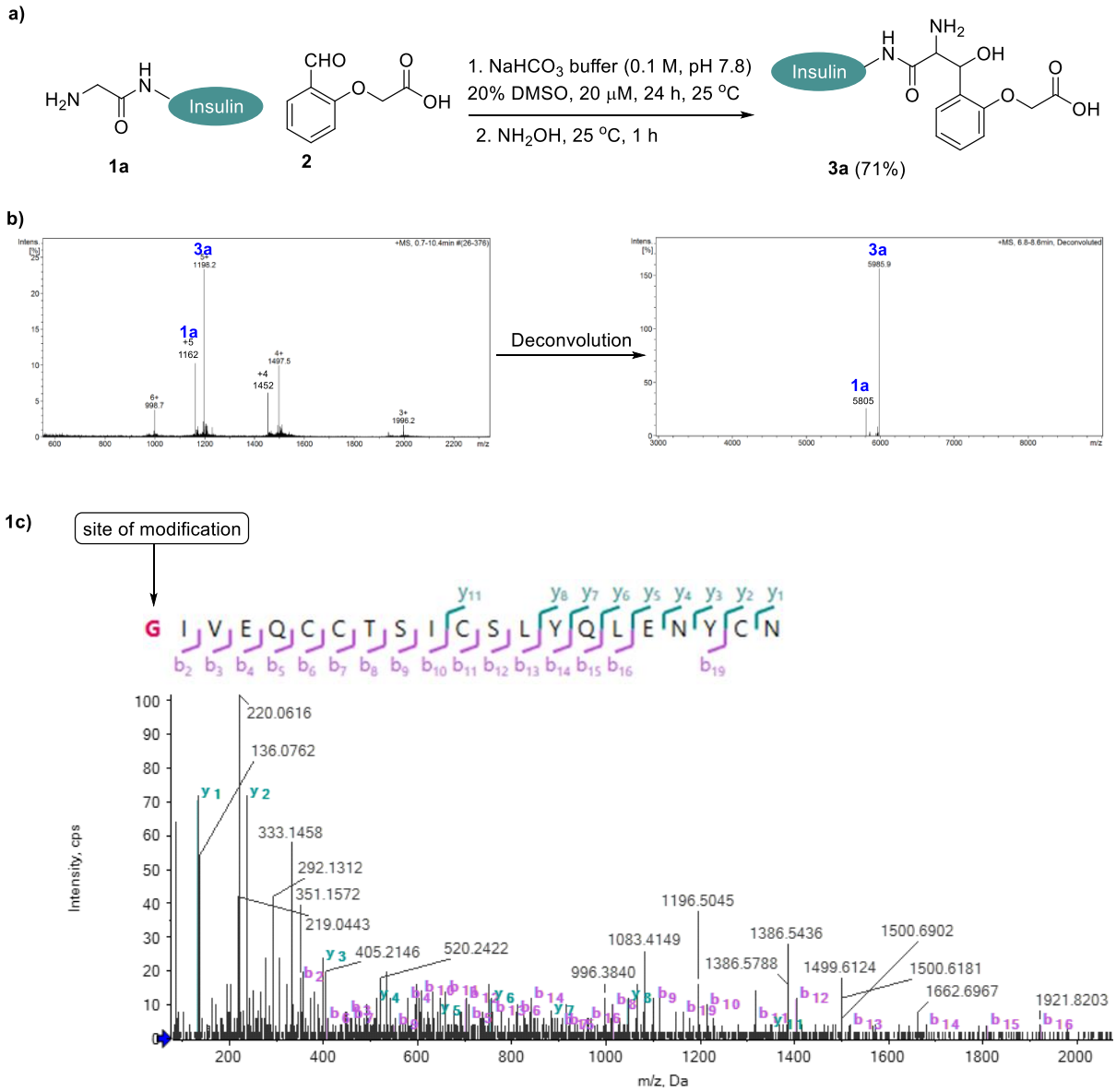
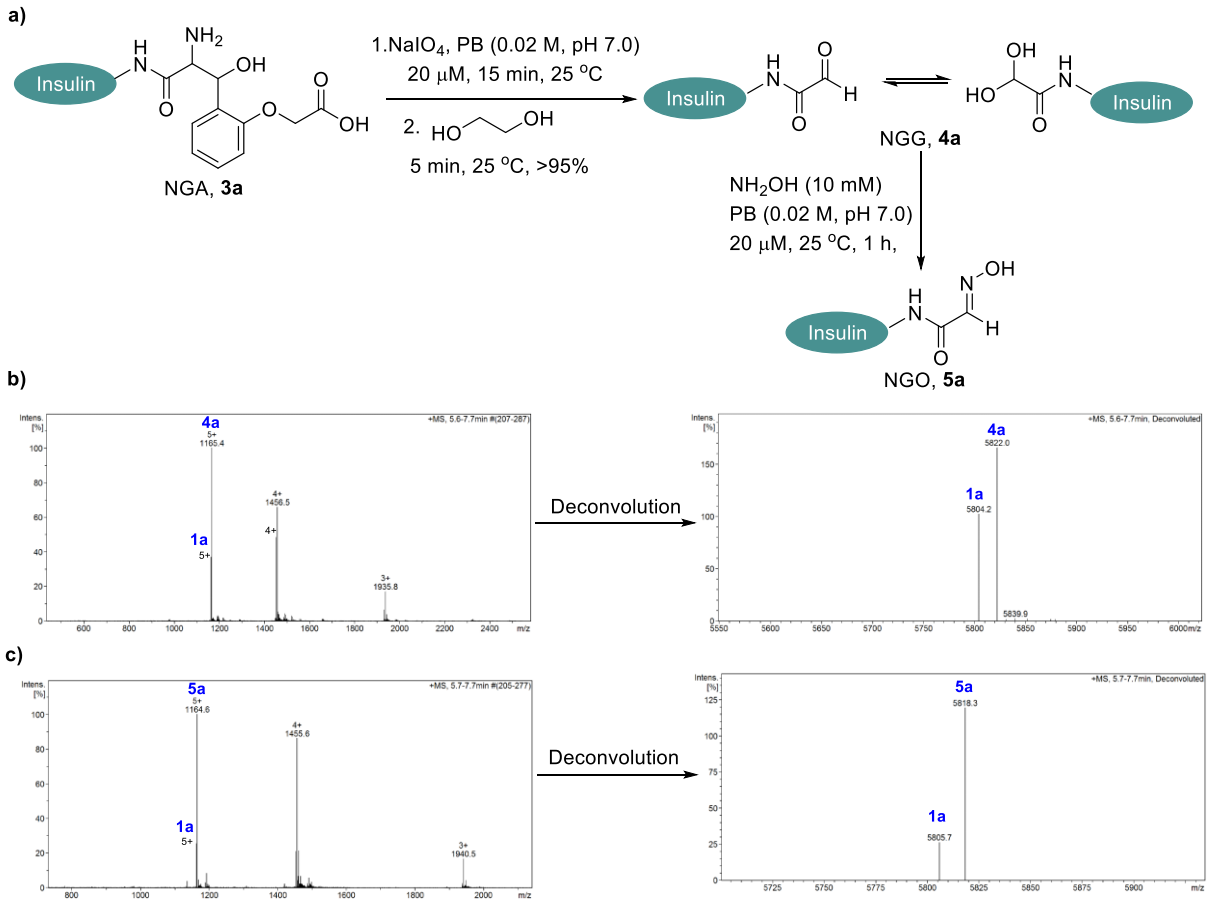


Figure S1. (a) Site-selective N-Gly-aminoalcohol insulin formation with Gly-tag reagent **2**. (b) LC-ESI-MS spectra for insulin (**1a**) and N-Gly-aminoalcohol insulin (**3a**). (c) MS-MS spectrum of labeled peptide GIVEQCCTSICSLYQLENYCN (G1-N21, *m/z* 855.0, $[M+3H]^{3+}$) confirms the site of modification N-Gly (G1).


a)
Insulin
NGA, 3a
1.NaIO4, PB (0.02 M, pH 7.0)
20 μM, 15 min, 25 °C
2.
5 min, 25 °C, >95%
NGG, 4a
NH2OH (10 mM)
PB (0.02 M, pH 7.0)
20 μM, 25 °C, 1 h,
NGO, 5a
b)
4a
1a
Deconvolution
c)
5a
1a
Deconvolution

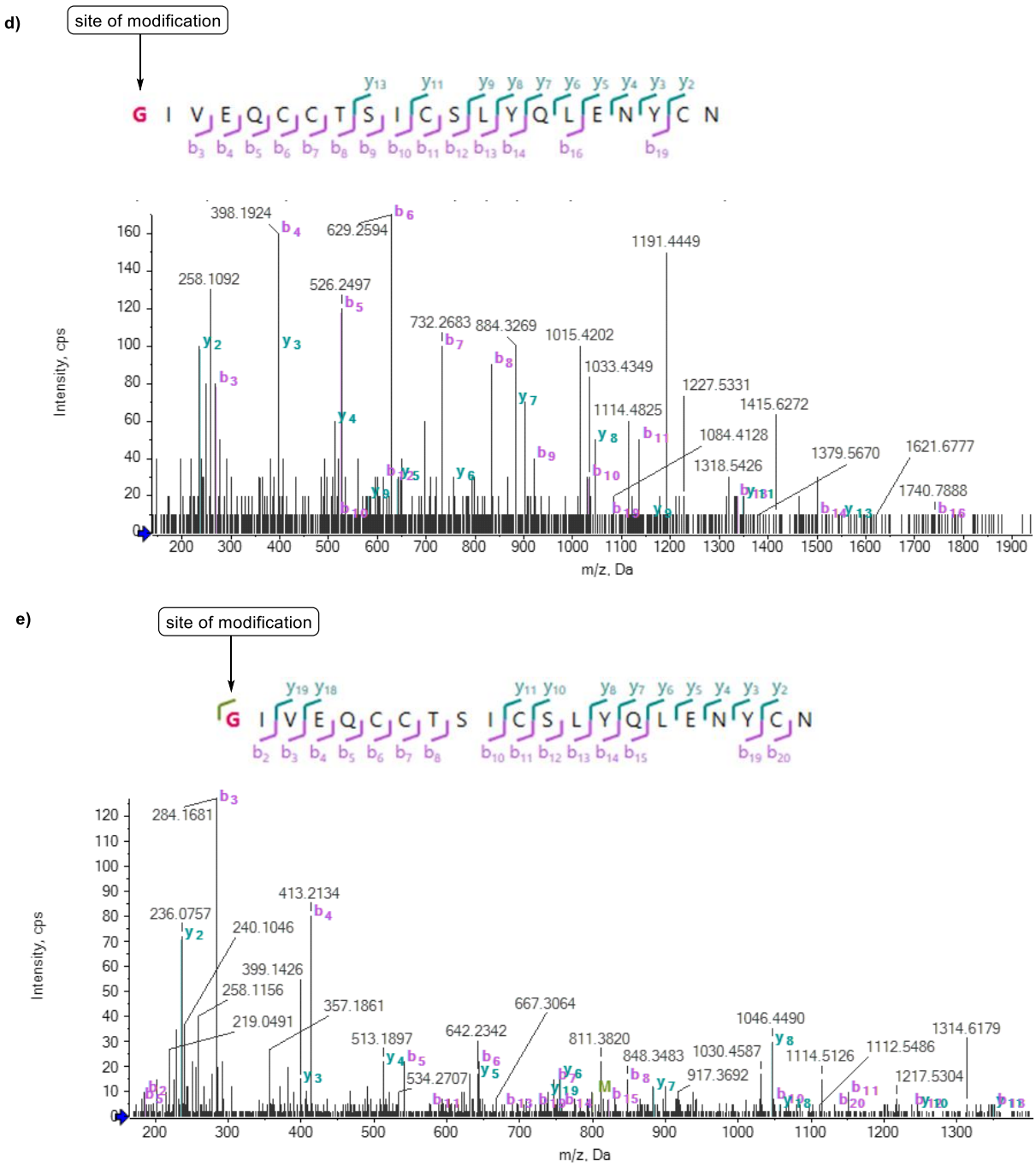


Figure S2. (a) NGA-NGG-NGO: Transformation of N-Gly-aminoalcohol insulin (**3a**) to N-Gly-glyoxamide (**4a**) and subsequently to N-Gly-oxime (**5a**) (b) LC-ESI-MS spectra for native insulin (**1a**) and N-Gly-glyoxamide insulin (**4a**). (c) LC-ESI-MS spectra for native insulin (**1a**) and N-Gly-oxime insulin (**5a**). (d) MS-MS spectrum of N-Gly-glyoxamide peptide GIVEQCCTSICSLYQLENYCN (G1-N21, *m/z* 1191.4 $[M+2H]^{2+}$) confirms the site of modification (G1). (e) MS-MS spectrum of N-Gly-oxime peptide GIVEQCCTSICSLYQLENYCN (G1-N21, *m/z* 799.6 $[M+3H]^{3+}$) confirms the site of modification N-Gly (G1).

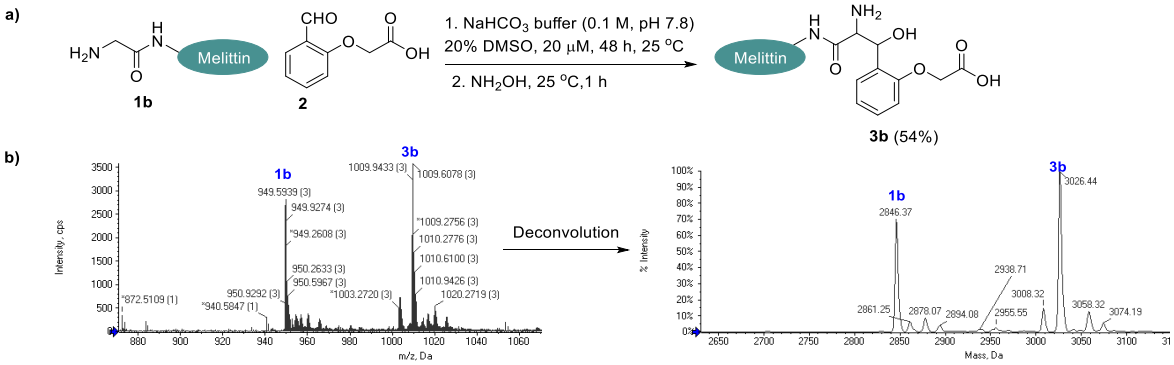


Figure S3. (a) Site-selective N-Gly-aminoalcohol melittin formation with Gly-tag reagent **2**. (b) LC-ESI-MS spectra for native melittin (**1b**) and N-Gly-aminoalcohol melittin (**3b**).

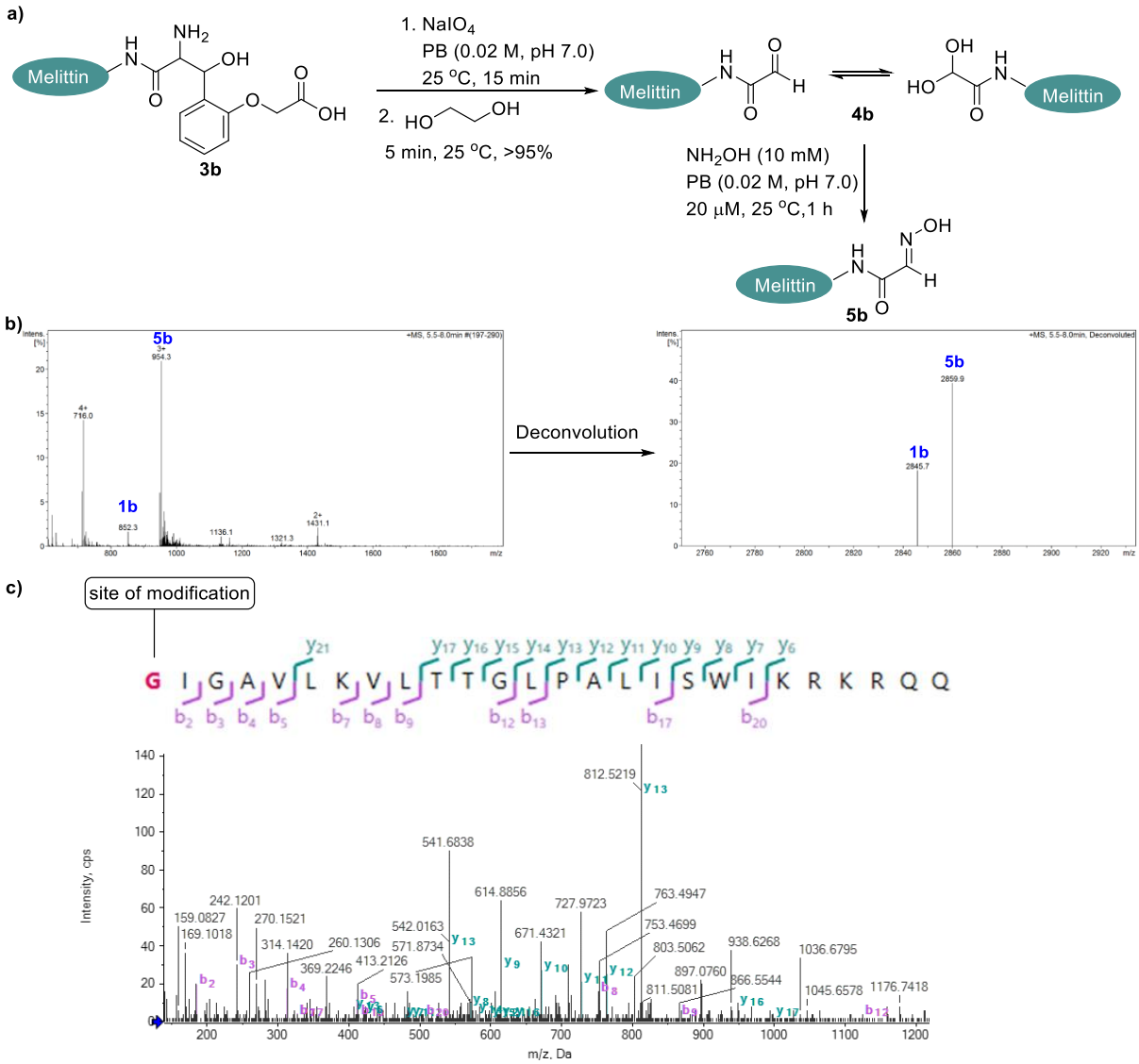


Figure S4. (a) Periodate cleavage of N-Gly-aminoalcohol melittin (**3b**) and late-stage modification with NH_2OH. (b) LC-ESI-MS spectra for native melittin (**1b**) and N-Gly-oxime melittin (**5b**). (c) MS-MS spectrum of N-Gly-oxime peptide GIGAVLKVLTTGLPALISWIKRKRQQ (G1-Q26, *m/z* 572.9 $[M+5H]^{5+}$) confirms the site of modification N-Gly (G1).

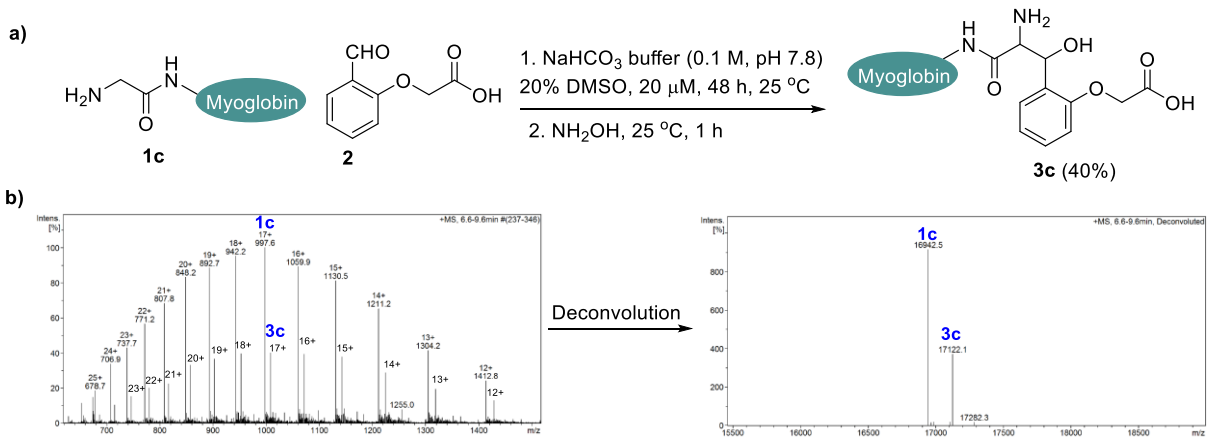


Figure S5. (a) Site-selective N-Gly-aminoalcohol myoglobin formation with Gly-tag reagent **2**. (b) LC-ESI-MS spectra for native myoglobin (**1c**) and N-Gly-aminoalcohol myoglobin (**3c**).

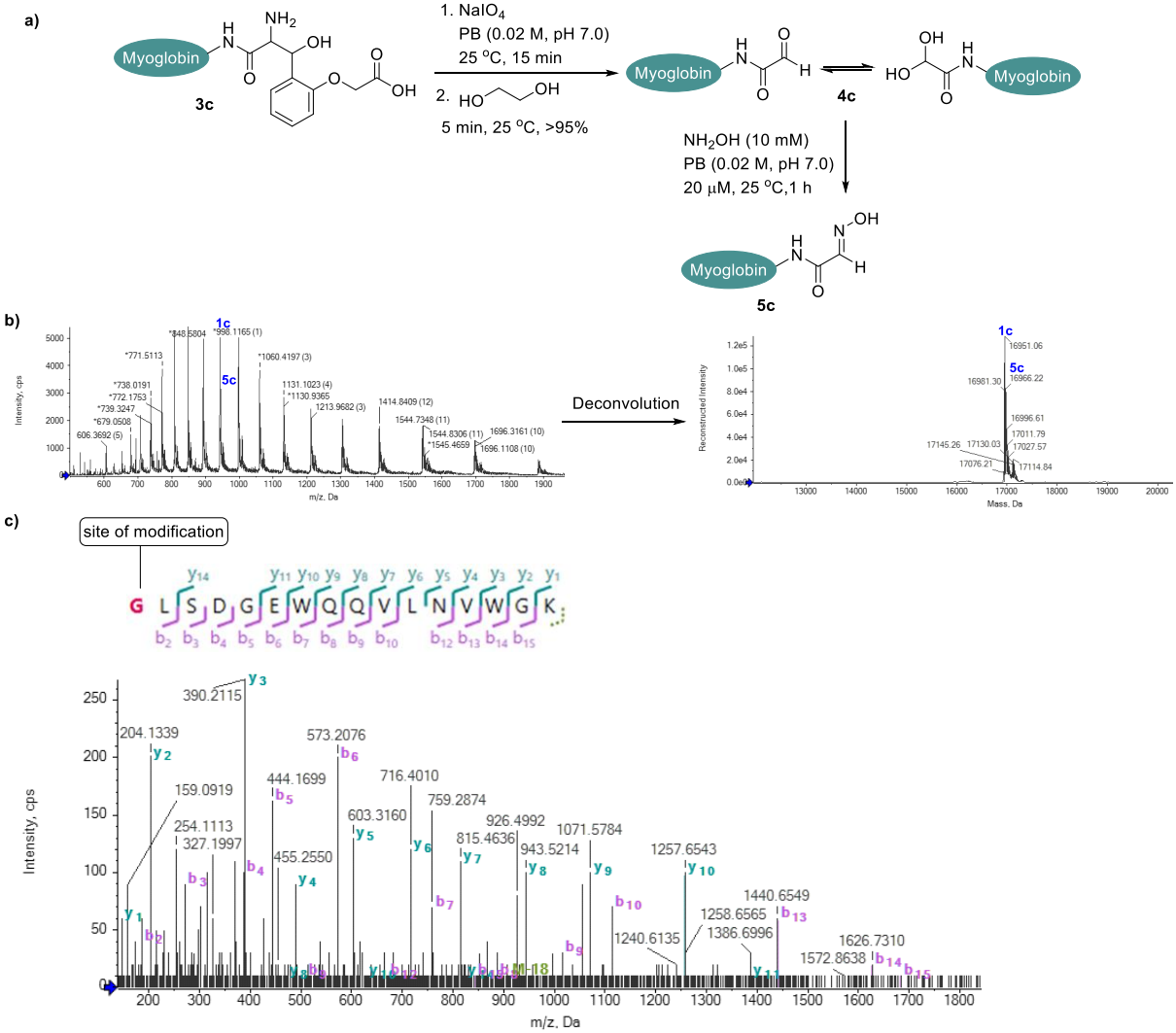


Figure S6. (a) Periodate cleavage of N-Gly-aminoalcohol myoglobin (**3c**) and late-stage modification with NH_2OH. (b) LC-ESI-MS spectra for native myoglobin (**1c**) and N-Gly-oxime myoglobin (**5c**). (c) MS-MS spectrum of N-Gly-oxime peptide GLSDGEWQQVLNVWGK (G1-K16, *m/z* 915.4 $[M+2H]^{2+}$) confirms the site of modification N-Gly (G1).

4.2 Late-stage modification

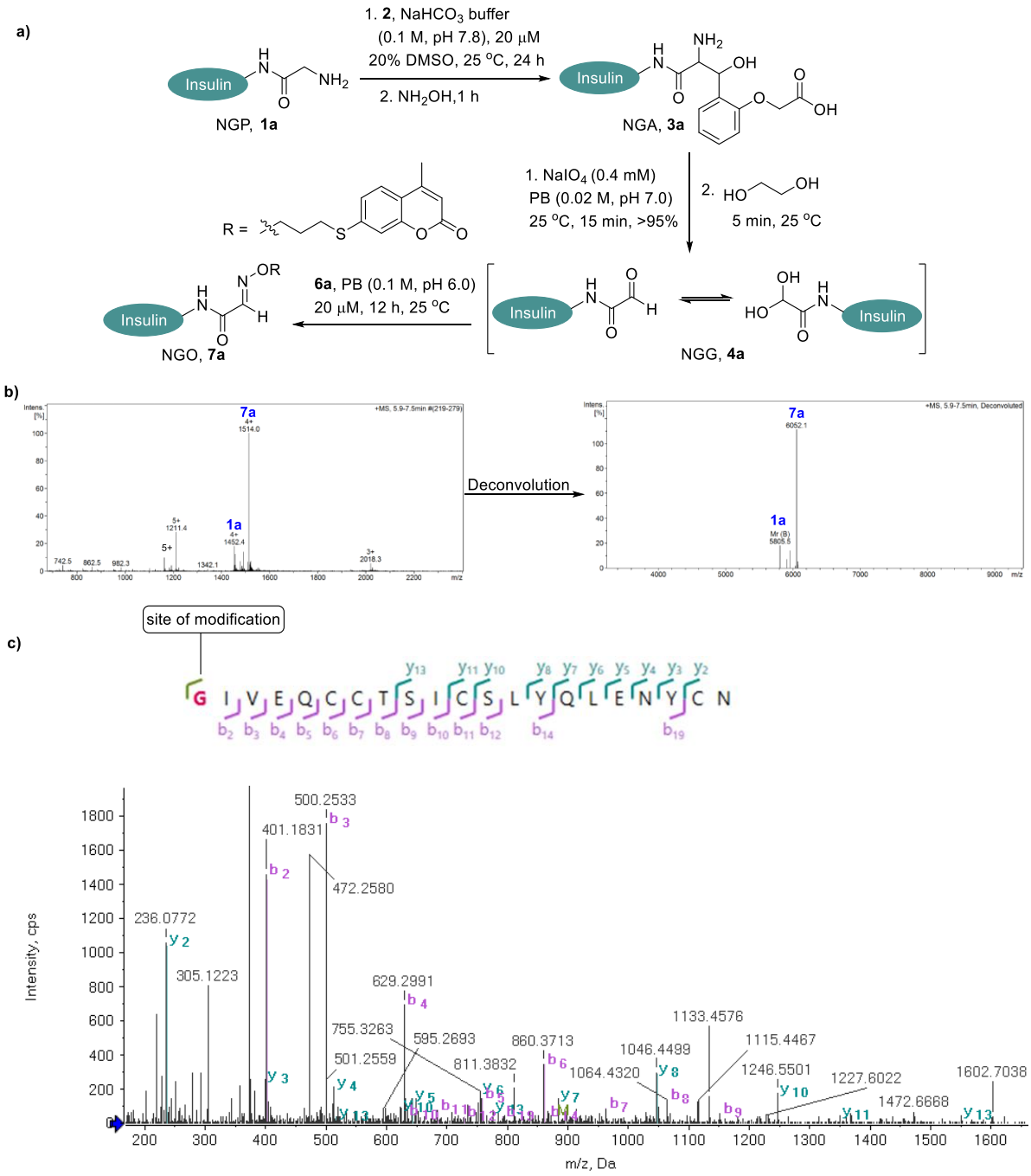


Figure S7. (a) Periodate cleavage of N-Gly-aminoalcohol insulin (**3a**) and late-stage modification with coumarin (**6a**). (b) LC-ESI-MS spectra for native insulin (**1a**) and N-Gly-oxime insulin (**7a**). (c) MS-MS spectrum of N-Gly-oxime peptide GIVEQCCTSICSLYQLENYCN (G1-N21, *m/z* 876.3 $[M+3H]^{3+}$) confirms the N-Gly modification.

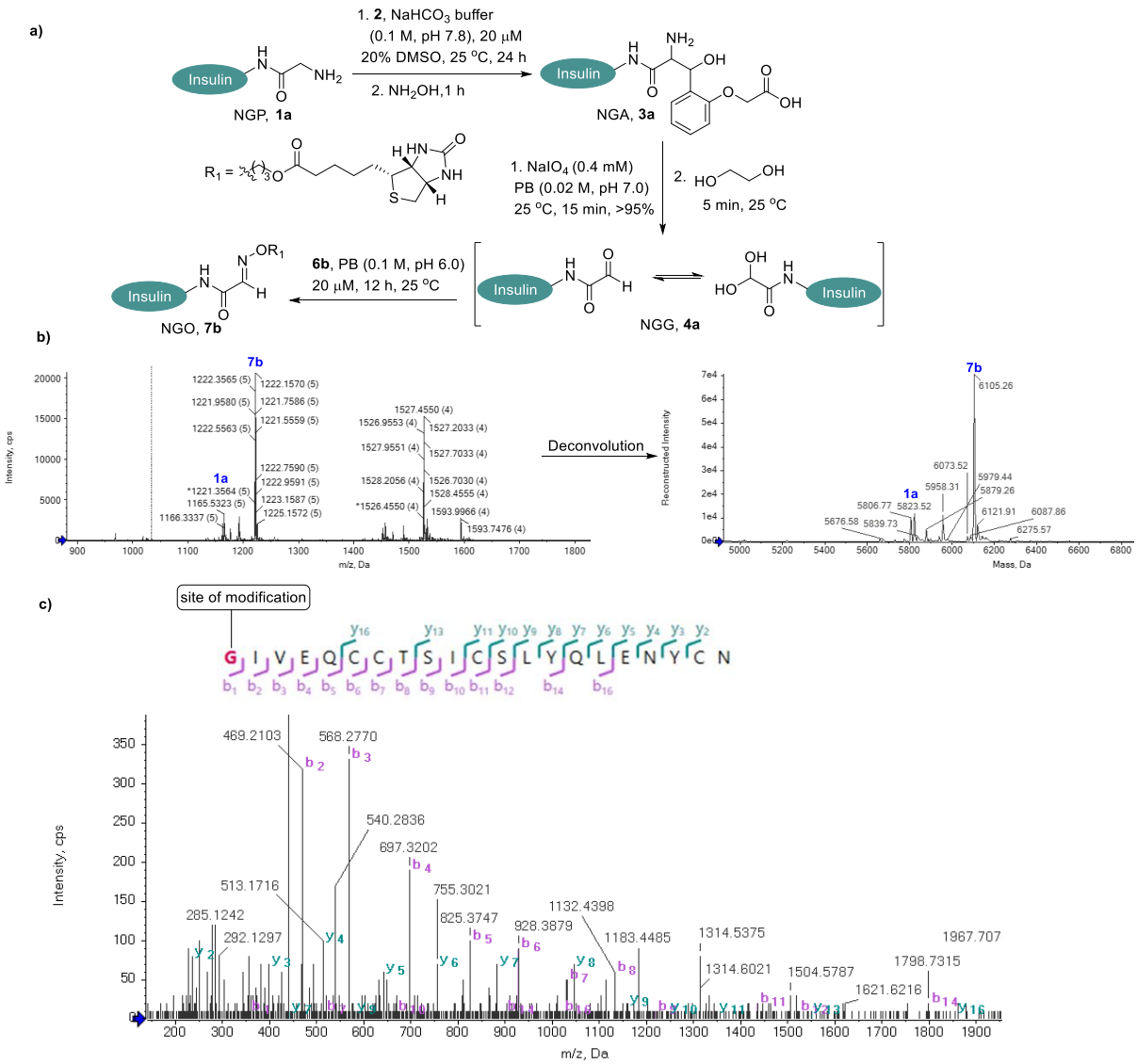


Figure S8. (a) Periodate cleavage of N-Gly-aminoalcohol insulin (**3a**) and late-stage modification with biotin (**6b**). (b) LC-ESI-MS spectra for native insulin (**1a**) and N-Gly-oxime insulin (**7b**). (c) MS-MS spectrum of N-Gly-oxime peptide GIVEQCCTSICSLYQLENYCN (G1-N21, *m/z* 1341.0 $[M+2H]^{2+}$) confirms the N-Gly modification.

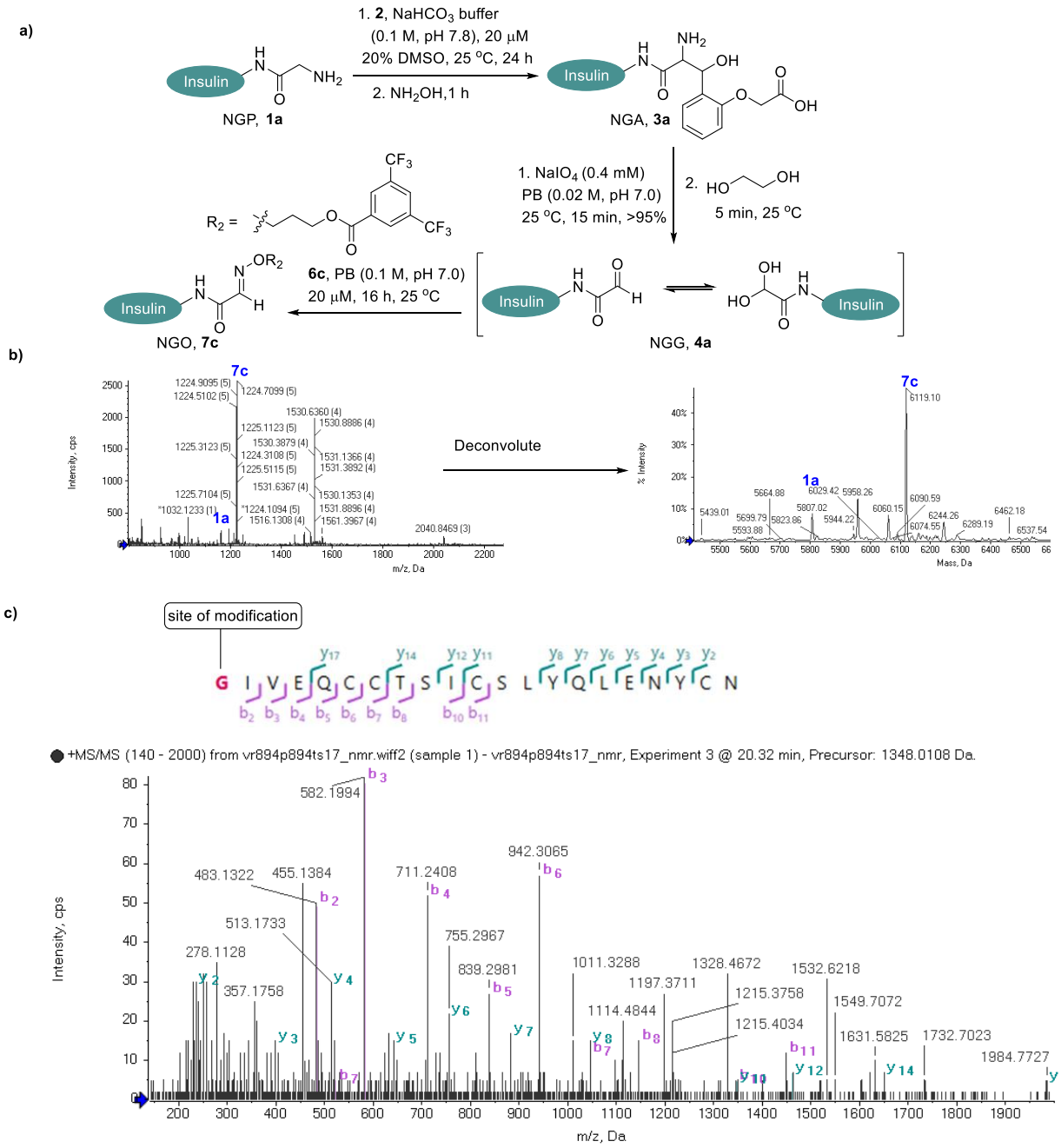
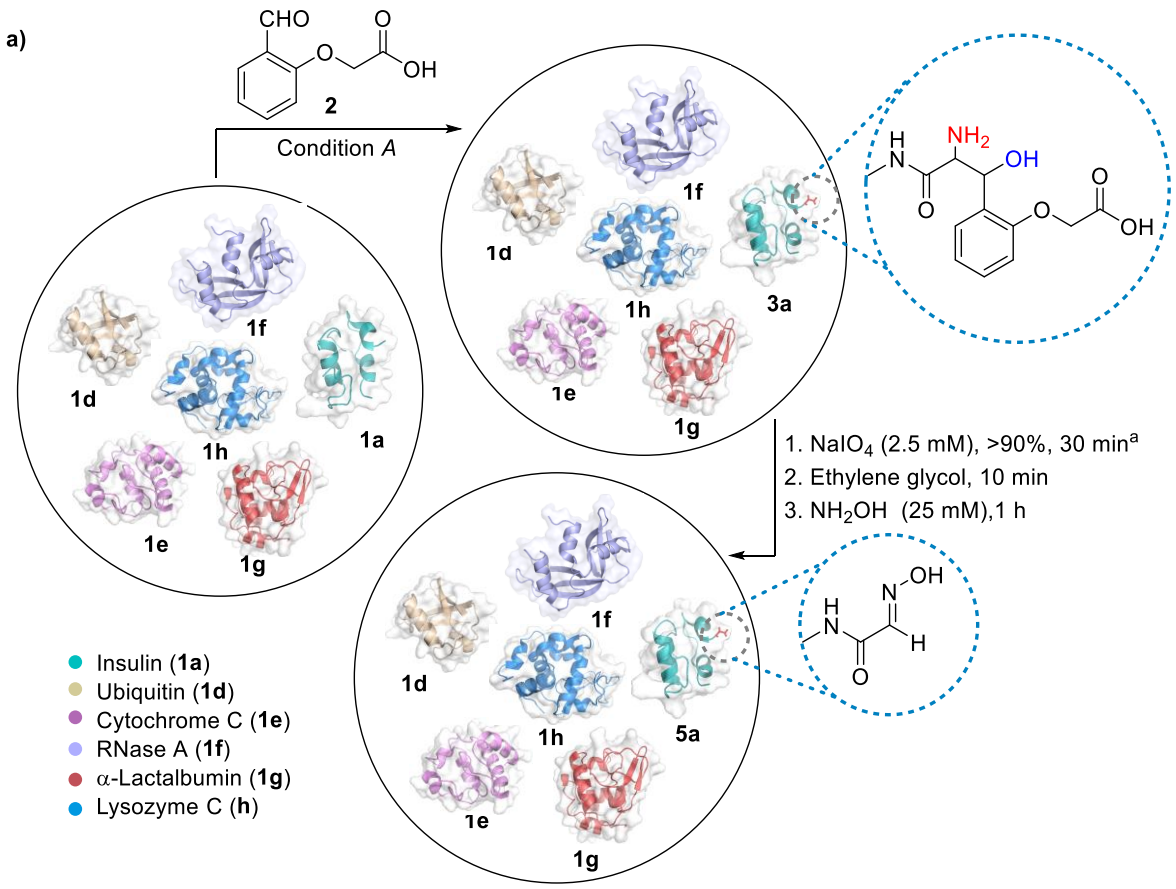


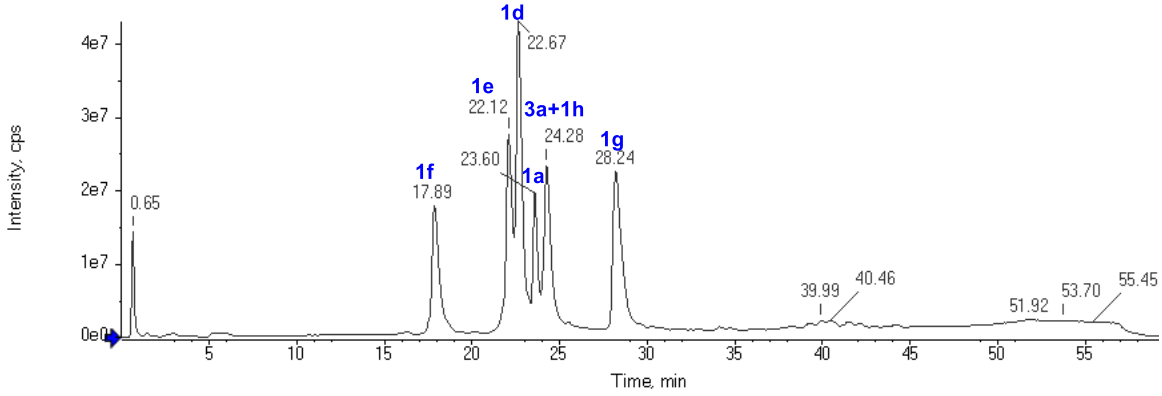
Figure S9. (a) Periodate cleavage of N-Gly-aminoalcohol insulin (**3a**) and late-stage modification with ^{19}F-NMR probe (**6c**). (b) LC-ESI-MS spectra for native insulin (**1a**) and N-Gly-oxime insulin (**7c**). (c) MS-MS spectrum of N-Gly-oxime peptide GIVEQCCTSICSLYQLENYCN (G1-N21, *m/z* 1348.0 $[M+2H]^{2+}$) confirms the N-Gly modification.

4.3 Transformation of N-Gly to N-Gly-glyoxamide in complex mixture of proteins, DMEM cell culture media and in human blood serum



Condition *A* : 1) $NaHCO_3$ buffer (pH 7.8, 0.1 M), **2** (25 mM), 50 μM, 48 h, 25 °C
2) NH_2OH, 1 h, 25 °C, 44% conversion. [a]Phosphate buffer (pH 7.0, 0.02 M), 20 μM, 25 °C

b)



RNase A (**1f**), Cytochrome C (**1e**), Ubiquitin (**1d**), Native insulin (**1a**),
N-Gly-aminoalcohol insulin (**3a**), Lysozyme C (**1h**), α-Lactalbumin (**1g**)

Figure S10. (a) NGP – NGA – NGG – NGO workflow in complex mixture of proteins (**1a, 1d-1h**). (b) LC-ESI-MS spectra for the N-Gly-aminoalcohol insulin (**3a**) formation in complex mixture of proteins after treatment with reagent **2**.

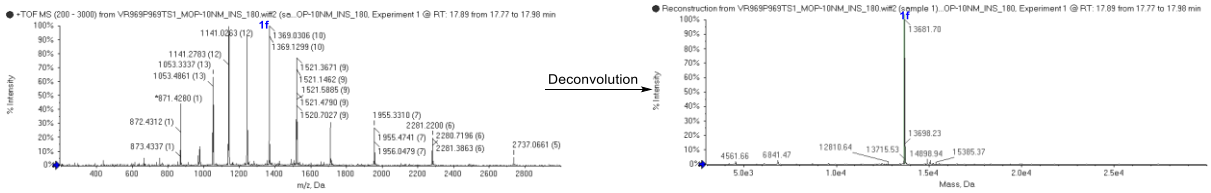


Figure S11. LC-ESI-MS spectra for the unlabeled RNase A (**1f**) in complex mixture of proteins after treatment with reagent **2**.

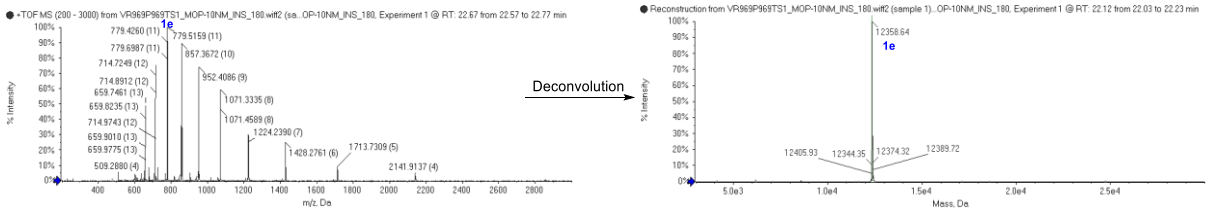


Figure S12. LC-ESI-MS spectra for the unlabeled cytochrome C (**1e**) in complex mixture of proteins after treatment with reagent **2**.

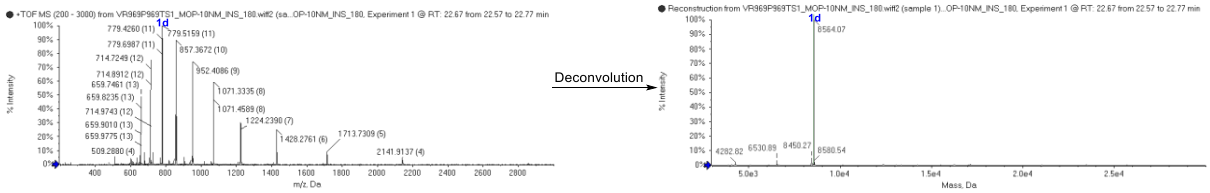


Figure S13. LC-ESI-MS spectra for the unlabeled ubiquitin (**1d**) in complex mixture of proteins after treatment with reagent **2**.

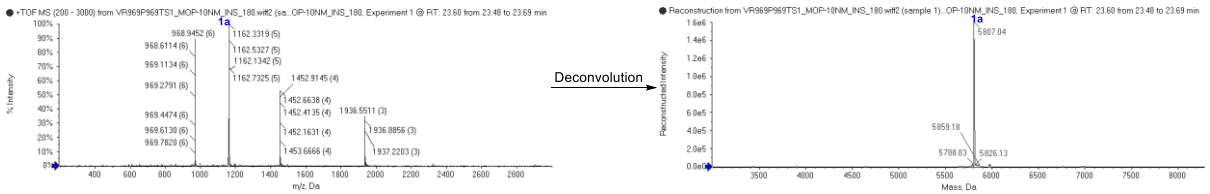


Figure S14. LC-ESI-MS spectra for the unlabeled insulin (**1a**) in complex mixture of proteins after treatment with reagent **2**.

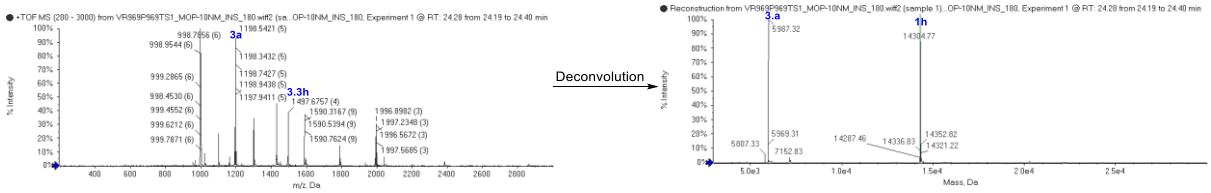


Figure S15. LC-ESI-MS spectra for the N-Gly-aminoalcohol insulin (**3a**) and unlabeled lysozyme C (**1h**) in complex mixture of proteins after treatment with reagent **2**.

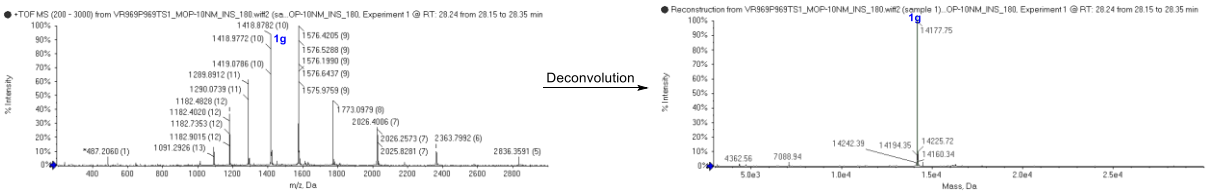
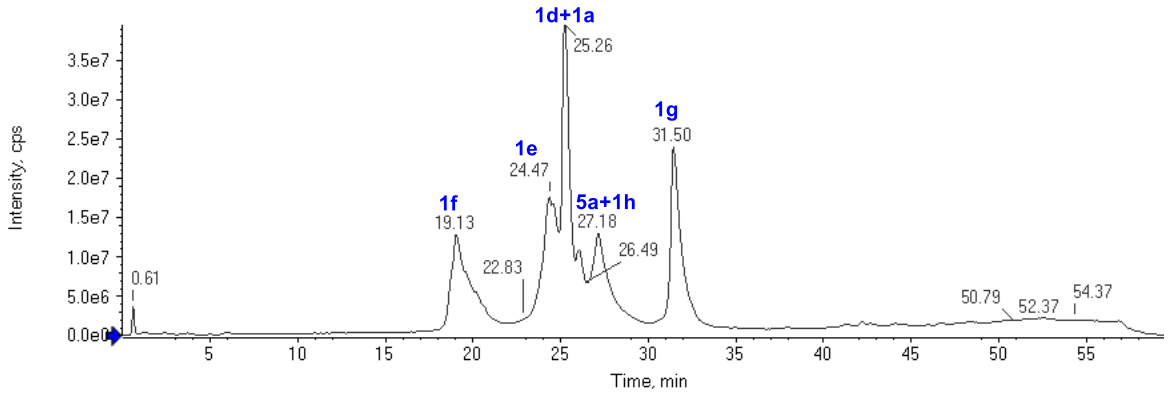


Figure S16. LC-ESI-MS spectra for the unlabeled α-lactalbumin (**1g**) in complex mixture of proteins after treatment with reagent **2**.



RNase A (**1f**), Cytochrome C (**1e**), Ubiquitin (**1d**), native insulin (**1a**), N-Gly-oxime insulin (**5a**), Lysozyme C (**1h**), α-Lactalbumin (**1g**)

Figure S17. LC-ESI-MS spectra for the N-Gly-oxime insulin (**5a**) formation in complex mixture of proteins after treatment with sodium periodate and NH_2OH.

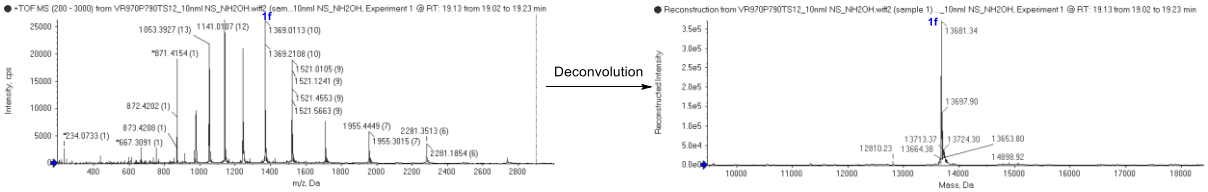


Figure S18. LC-ESI-MS spectra for the unlabeled RNase A (**1f**) in complex mixture of proteins after treatment with sodium periodate and NH_2OH.

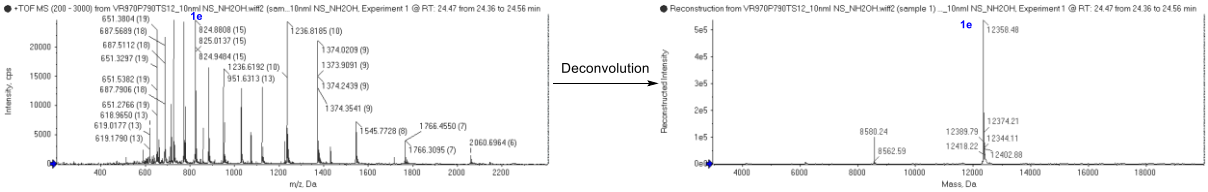


Figure S19. LC-ESI-MS spectra for the unlabeled cytochrome C (**1e**) in complex mixture of proteins after treatment with sodium periodate and NH_2OH.

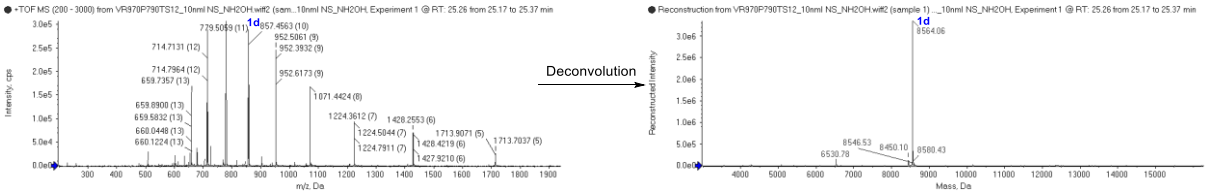


Figure S20. LC-ESI-MS spectra for the unlabeled ubiquitin (**1d**) in complex mixture of proteins after treatment with sodium periodate and NH_2OH.

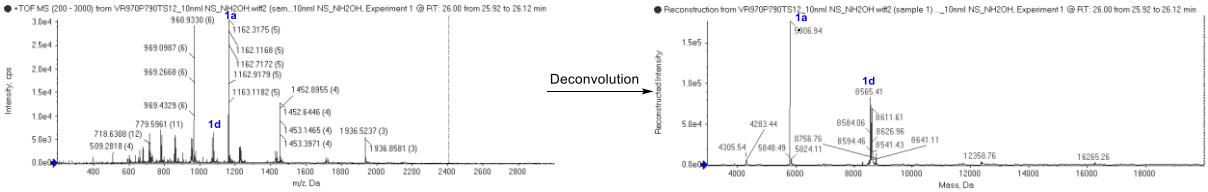


Figure S21. LC-ESI-MS spectra for the unlabeled insulin (**1a**) and unlabeled ubiquitin (**1d**) in complex mixture of proteins after treatment with sodium periodate and NH_2OH.

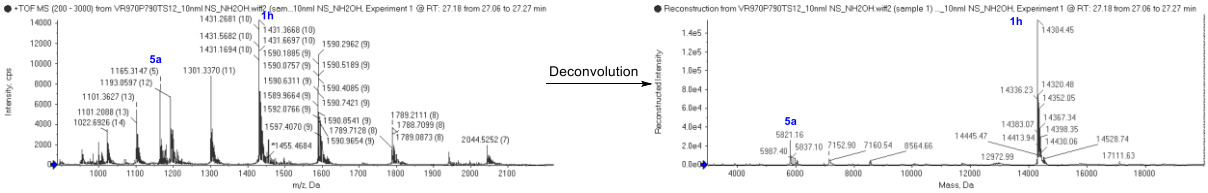


Figure S22. LC-ESI-MS spectra for the unlabeled lysozyme C (**3h**) in complex mixture of proteins after treatment with sodium periodate and NH_2OH.

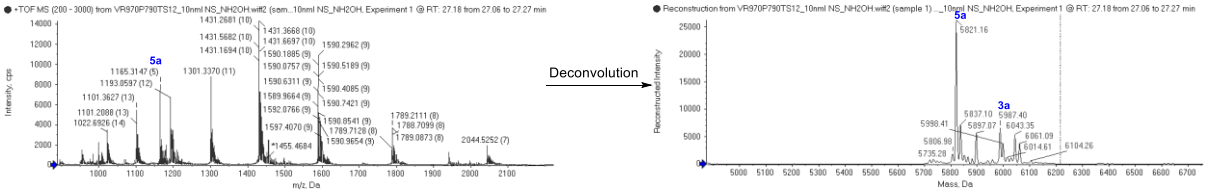


Figure S23. LC-ESI-MS spectra for the N-Gly-oxime insulin (**5a**) in complex mixture of proteins after treatment with sodium periodate and NH_2OH.

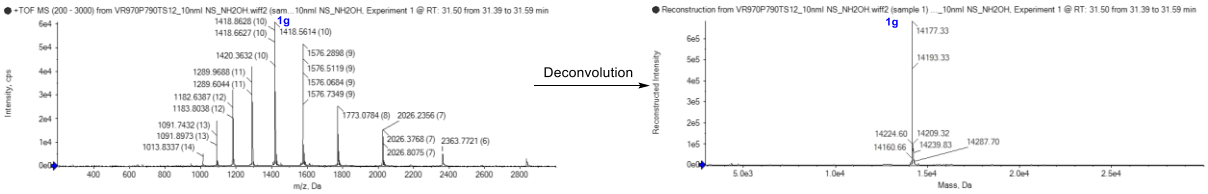


Figure S24. LC-ESI-MS spectra for the unlabeled α-lactalbumin (**1g**) in complex mixture of proteins after treatment with sodium periodate and NH_2OH.

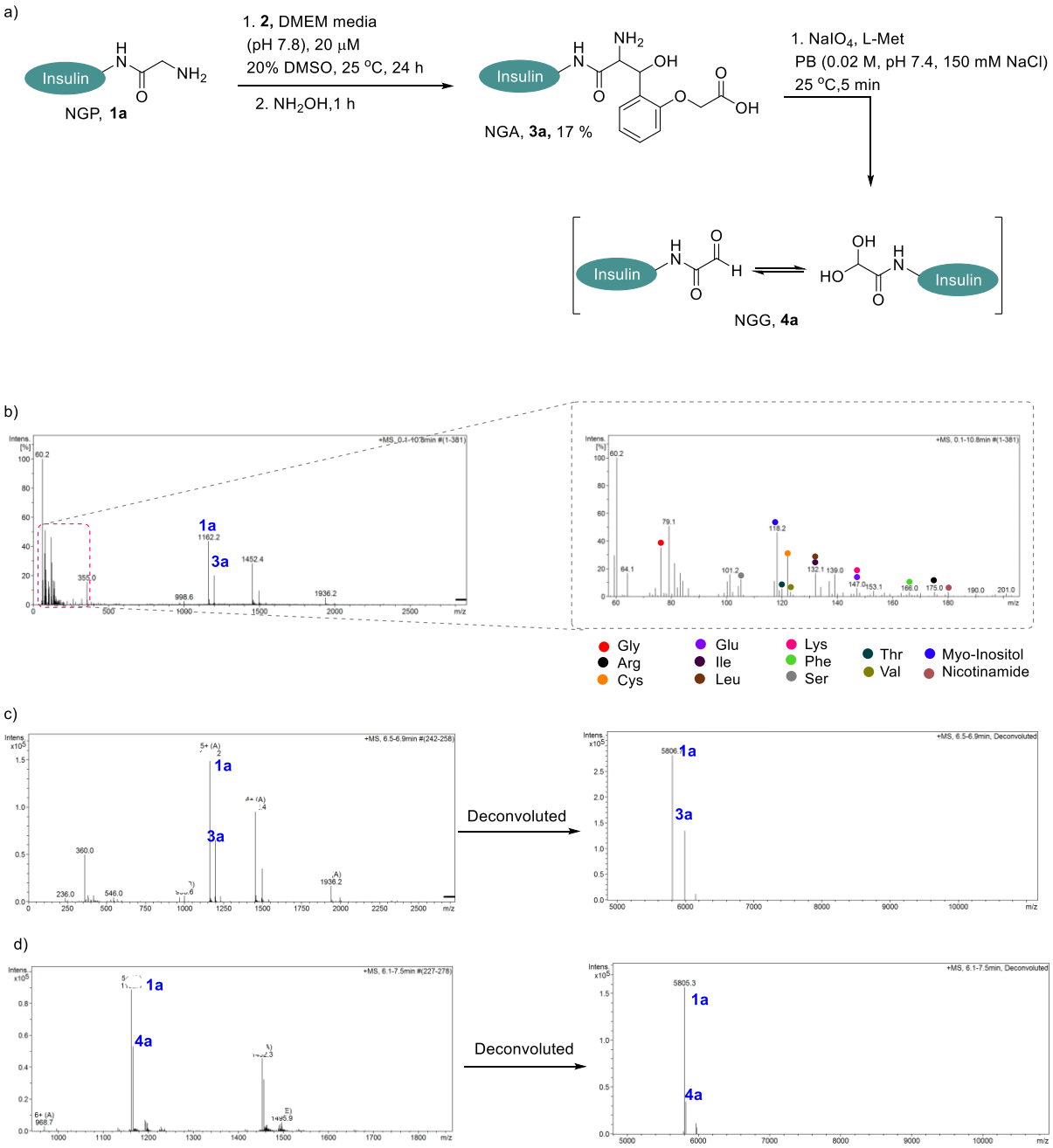


Figure S25. (a) N-Gly aminoalcohol formation with Insulin in DMEM cell culture media. (b) LC-ESI-MS spectra for the reaction mixture of N-Gly-aminoalcohol insulin (**3a**) displaying component from cell culture media. (c) LC-ESI-MS spectra for insulin (**1a**) and N-Gly-aminoalcohol insulin (**3a**). (d) LC-ESI-MS spectra for native insulin (**1a**) and N-Gly-glyoxamide insulin (**4a**).

4.4 Purification of labeled protein

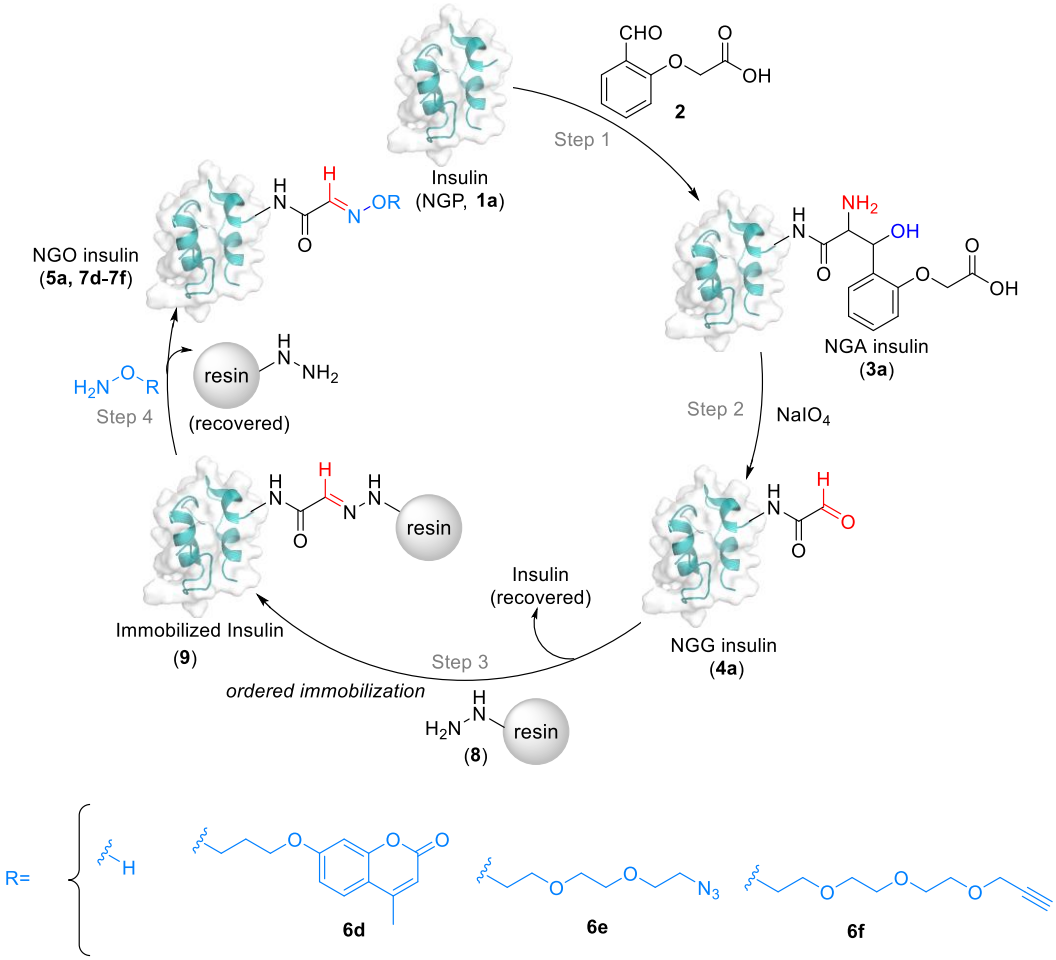


Figure S26. Integration of purification protocol to NGP – NGA – NGG – NGO workflow: *Step 1*: N-Gly-aminoalcohol formation (**3a**), *Step 2*: N-Gly-glyoxamide generation (**4a**), *Step 3*: Ordered immobilization of N-Gly-glyoxamide insulin (**4a)** on hydrazide resin (**8**), *Step 4*: Trans-oximization with O-hydroxylamine and its derivatives (**6d-6f**) to release N-Gly-oxime insulin (**5a**, **7d-7f**) in analytically pure single-site tagged form.

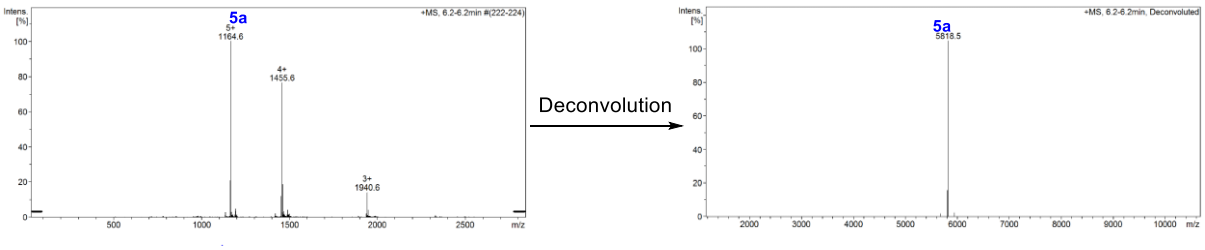


Figure S27. LC-ESI-MS spectra of purified N-Gly-oxime insulin (**5a**) labeled with NH_2OH.

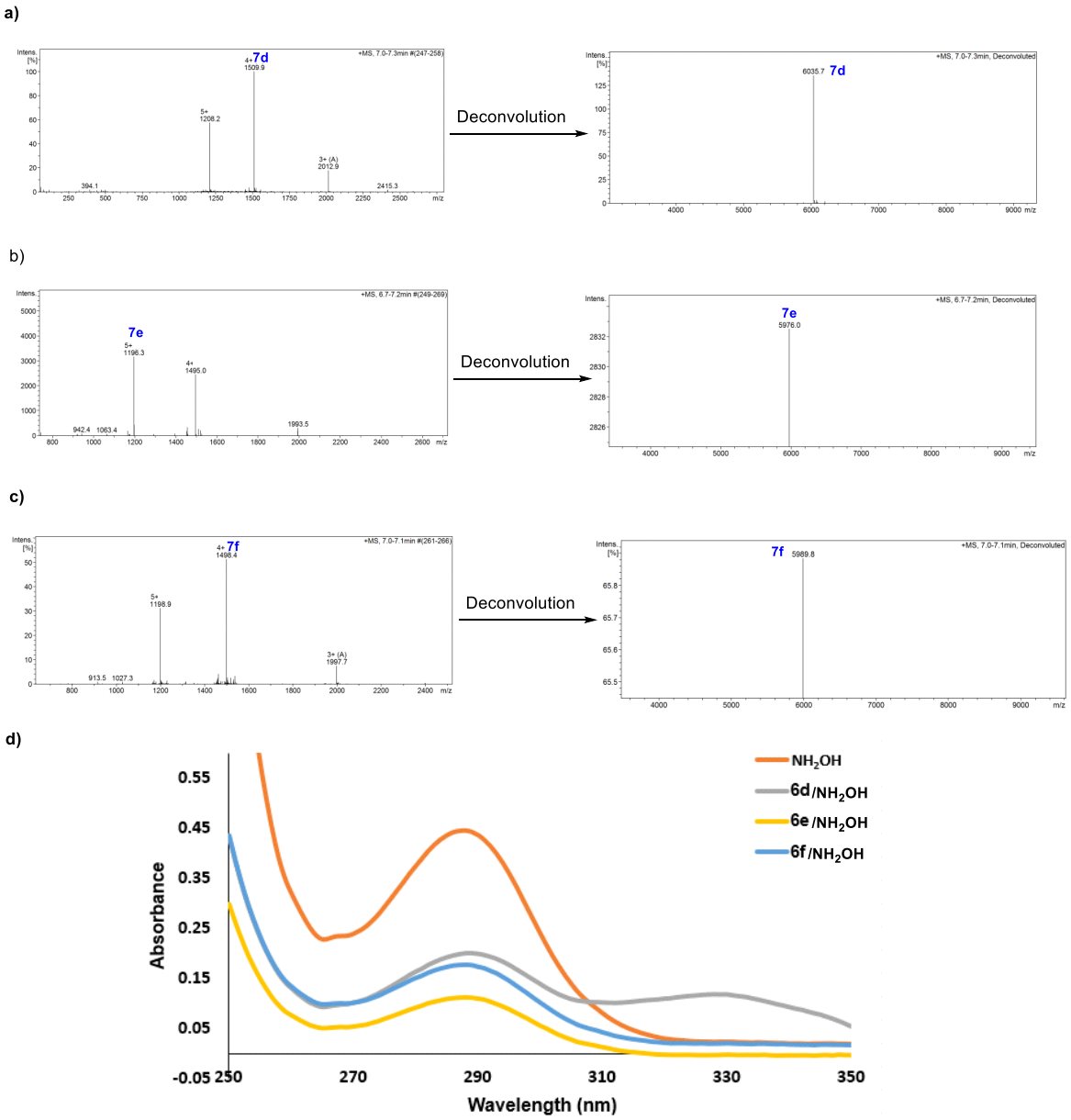
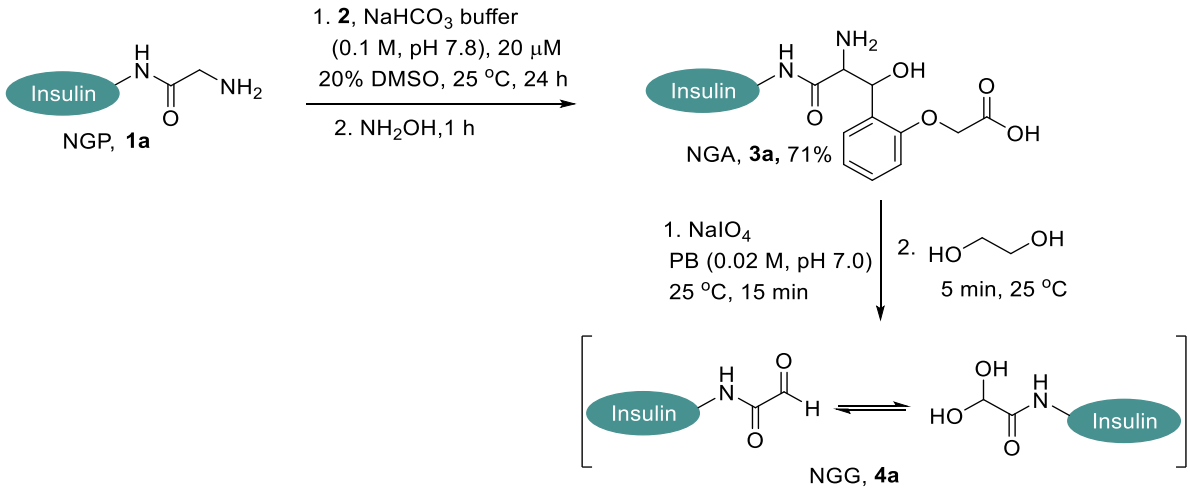


Figure S28. (a) LC-ESI-MS spectra of purified N-Gly-oxime insulin (**7d**) labeled with **6d**. (b) LC-ESI-MS spectra of purified N-Gly-oxime insulin (**7e**) labeled with **6e**. (c) LC-ESI-MS spectra of purified N-Gly-oxime insulin (**7f**) labeled with **6f**. (d) The efficiency of release by **6d** (55%), **6e** (75%), and **6f** (60%) were estimated with respect to the hydroxylamine.

4.5 Optimization of reaction conditions

Table S1. Optimization of the relative stoichiometry of $NaIO_4$



Entry	$NaIO_4$ (equiv.)	% Conversion (4a)[a]
1	5	13
2	10	52
3	15	75
4	20	>95

[a] % Conversions are estimated based on LC-ESI-MS.

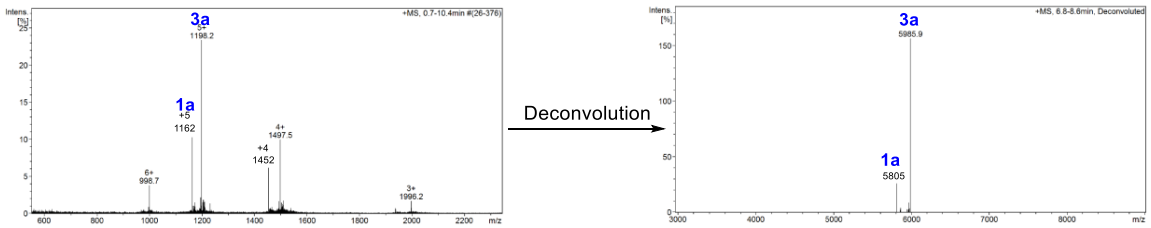


Figure S29. LC-ESI-MS spectra for insulin (**1a**) and N-Gly-aminoalcohol insulin (**3a**).

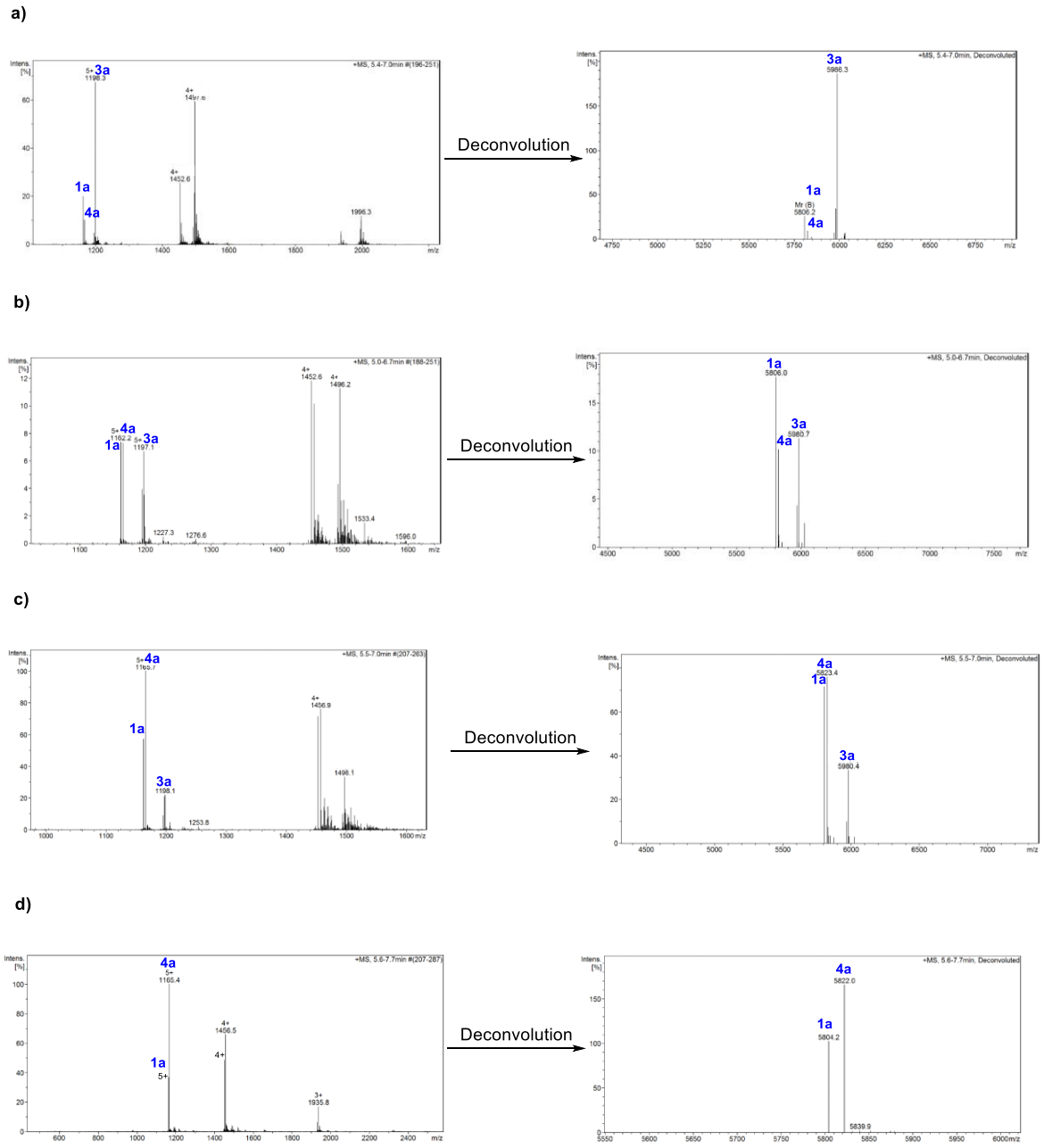
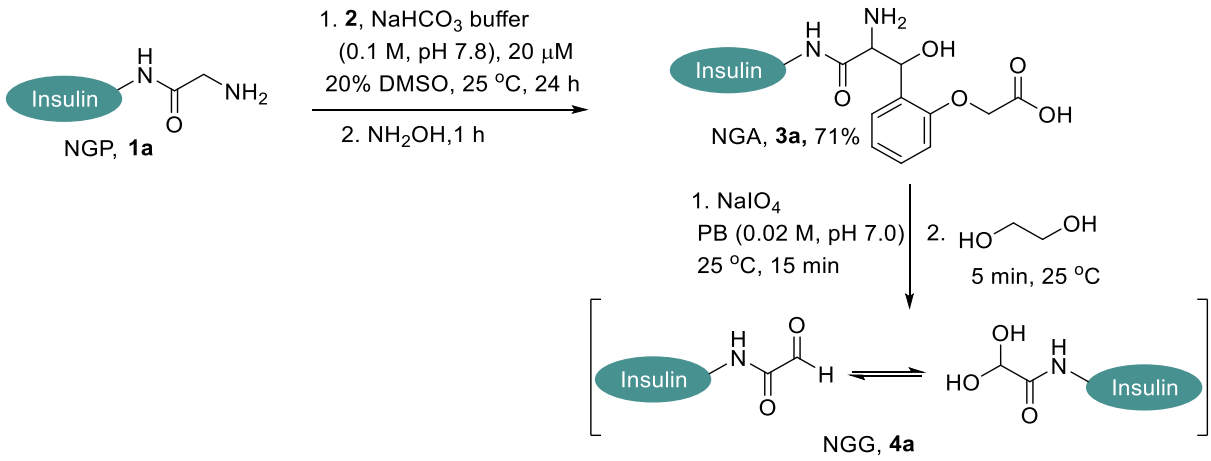


Figure S30. LC-ESI-MS spectra for native insulin (**1a**) and N-Gly-glyoxamide insulin (**4a**) N-Gly-aminoalcohol insulin (**3a**) for optimization of of the relative stoichiometry of $NaIO_4$ (a) 5 equivalents, (b) 10 equivalents, (c) 15 equivalents, and (d) 20 equivalents.

Table S2. Optimization of the reaction temperature for N-Gly-aminoalcohol (**3a)** to N-Gly-glyoxamide (**4a**) formation



Entry	T (°C)	% Conversion (4a)[a]
1	25	>95
2	4	>95

[a] % Conversions are estimated based on LC-ESI-MS.

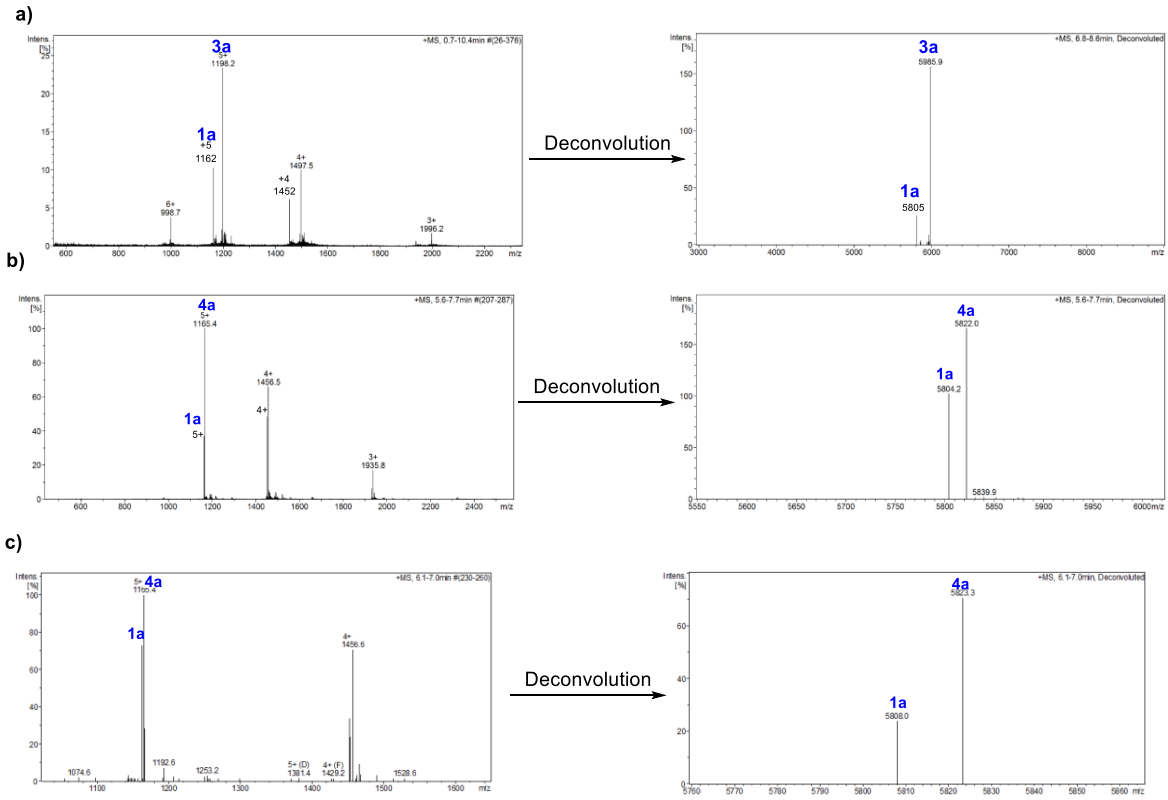
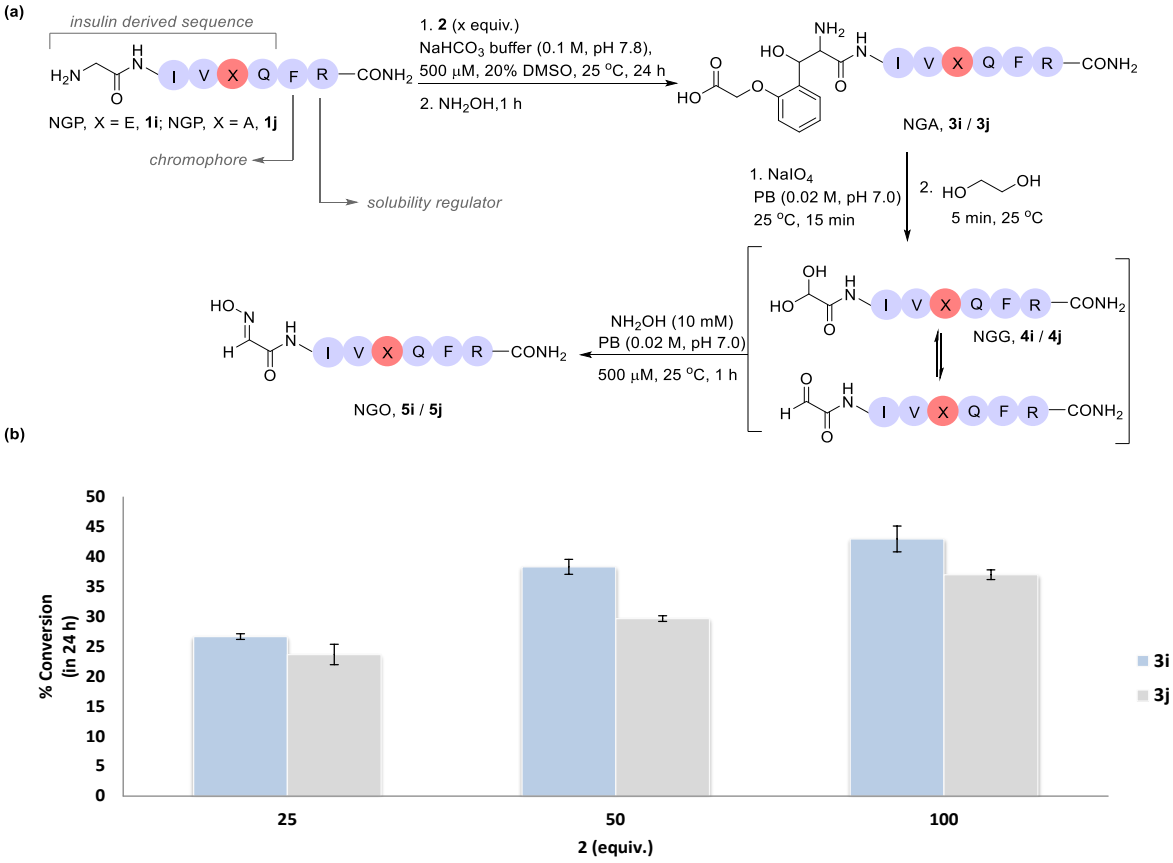


Figure S31. LC-ESI-MS spectra for insulin (**1a**) and N-Gly-aminoalcohol insulin (**3a**). (b) LC-ESI-MS spectra for insulin (**1a**) and N-Gly-glyoxamide insulin (**4a**) at 25 °C. (c) LC-ESI-MS spectra for insulin (**1a**) and N-Gly-glyoxamide insulin (**4a**) at 4 °C.

4.6 NGP – NGA – NGG – NGO workflow for insulin derived peptides

Table S3. Impact of proximal Glu on Gly-tag chemistry



Entry	2 (equiv.)	% Conversion (3i)[a]	% Conversion (3j)[a]
1	25 (trial 1)	26	23
2	25 (trial 2)	27	22
3	25 (trial 3)	27	26
4	50 (trial 1)	38	30
5	50 (trial 2)	40	29
6	50 (trial 3)	37	30
7	100 (trial 1)	42	37
8	100 (trial 2)	46	36
9	100 (trial 3)	41	38

[a] % Conversions are estimated based on LC-ESI-MS.

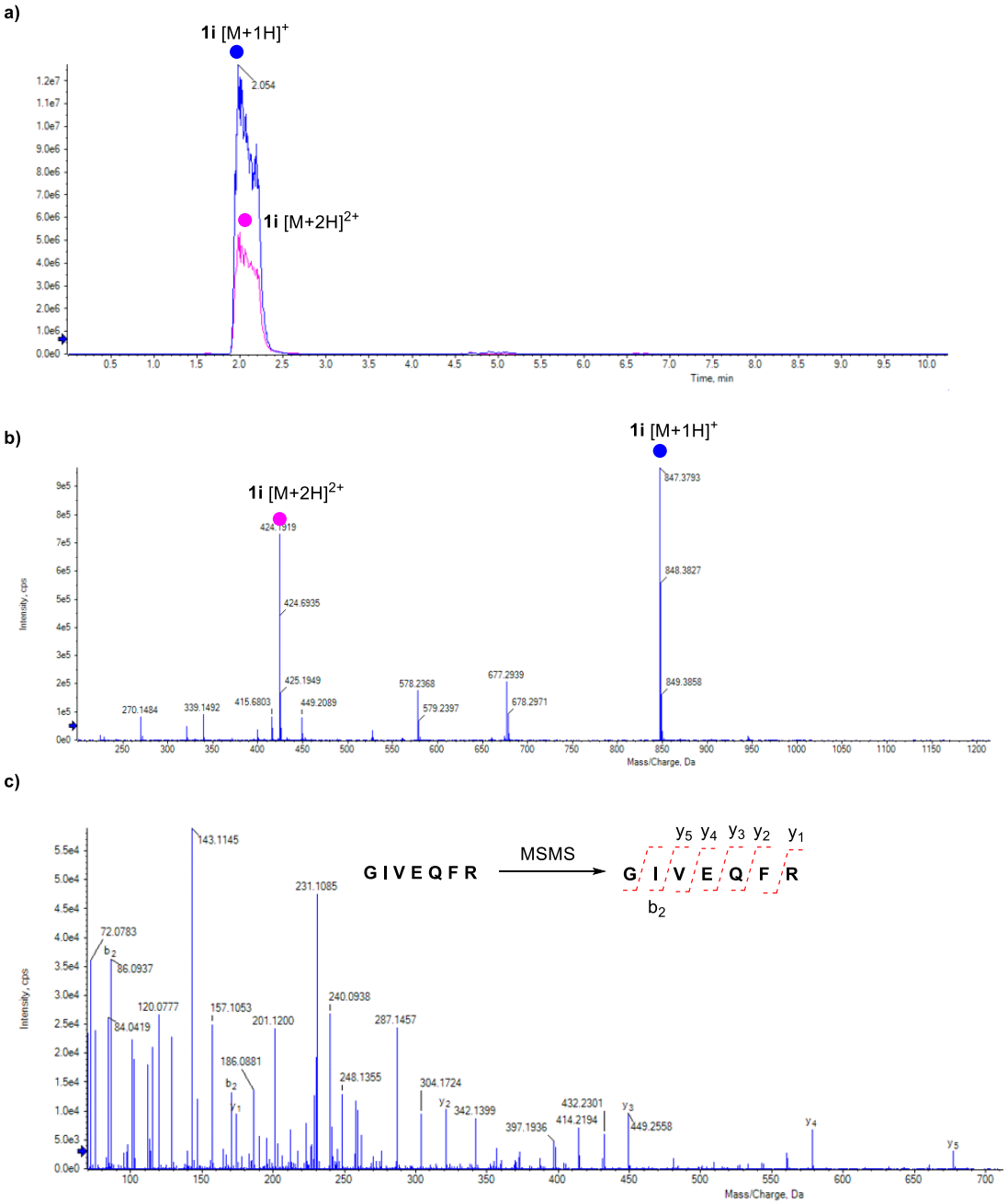


Figure S32. (a) XIC data for peptide GIVEQFR (**1i**). (b) ESI-MS data for peptide **1i**. (c) MS-MS spectrum of **1i**.

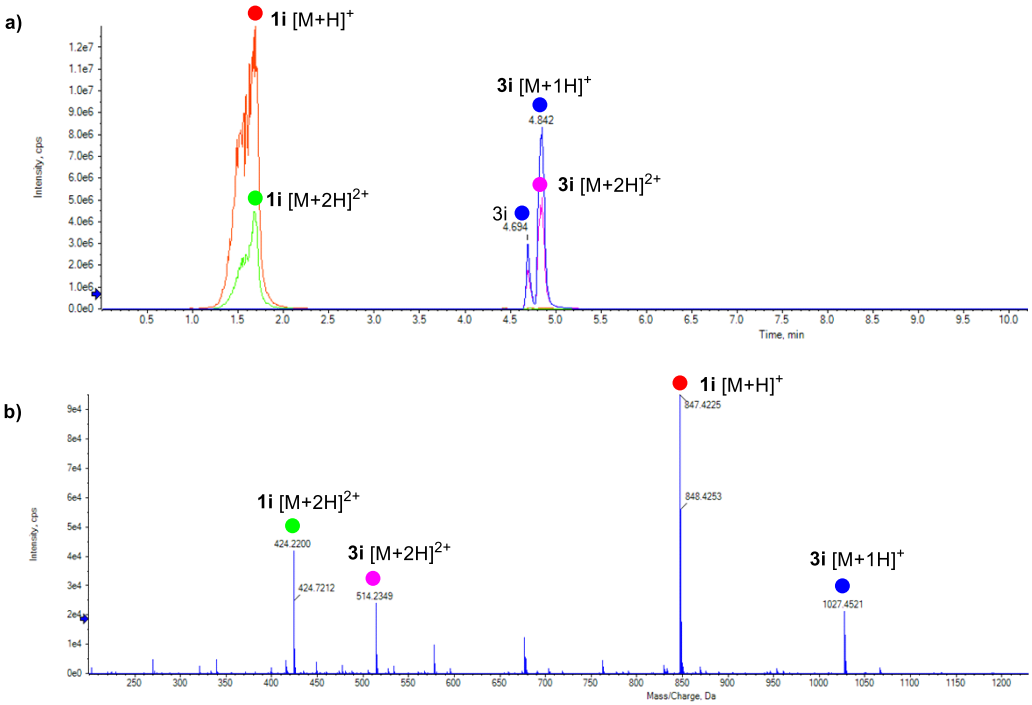


Figure S33. Trial 1: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (25 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

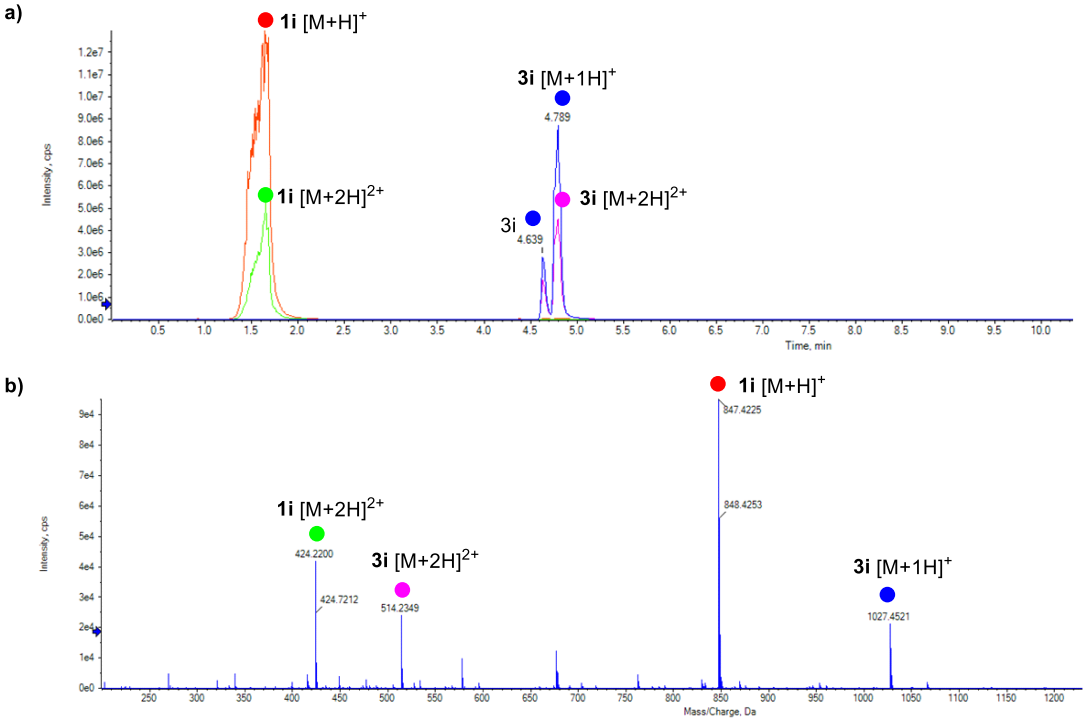


Figure S34. Trial 2: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (25 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

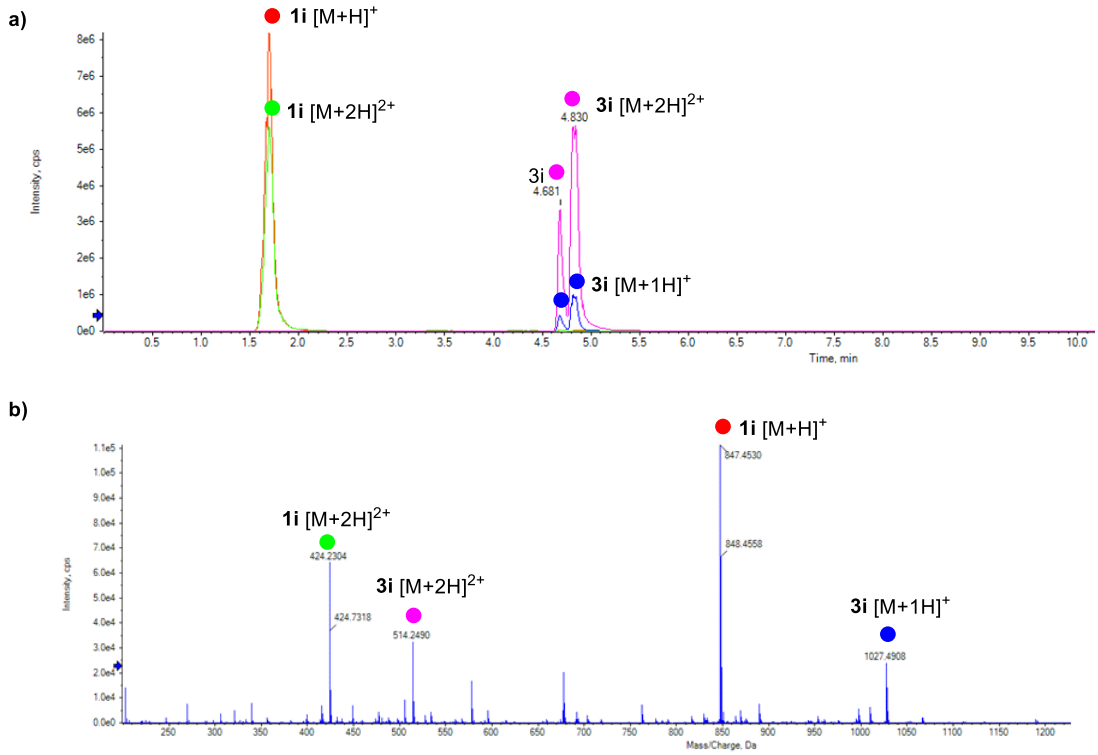


Figure S35. Trial 3: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (25 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

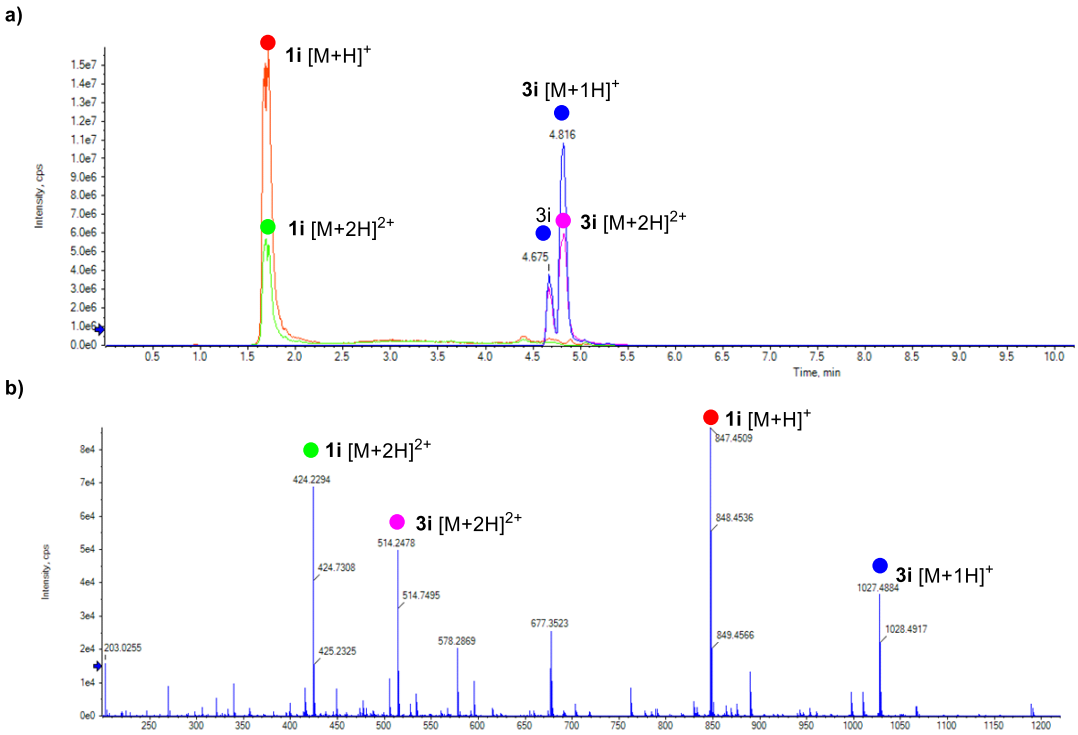


Figure S36. Trial 1: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (50 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

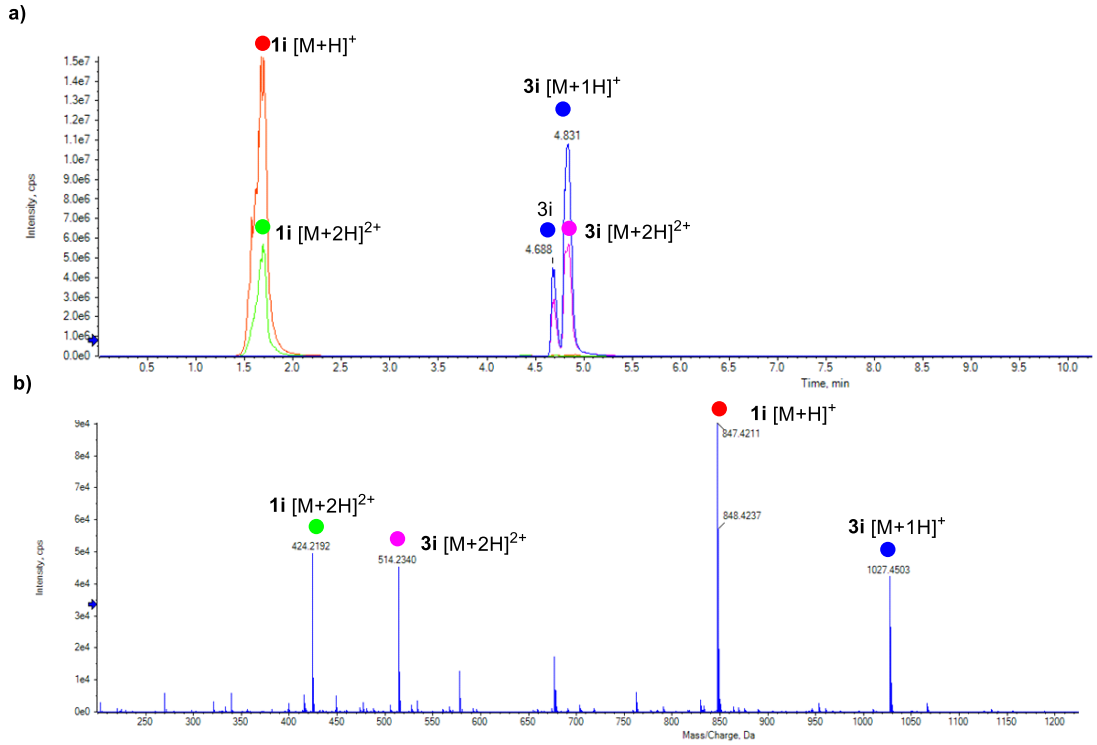


Figure S37. Trial 2: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (50 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

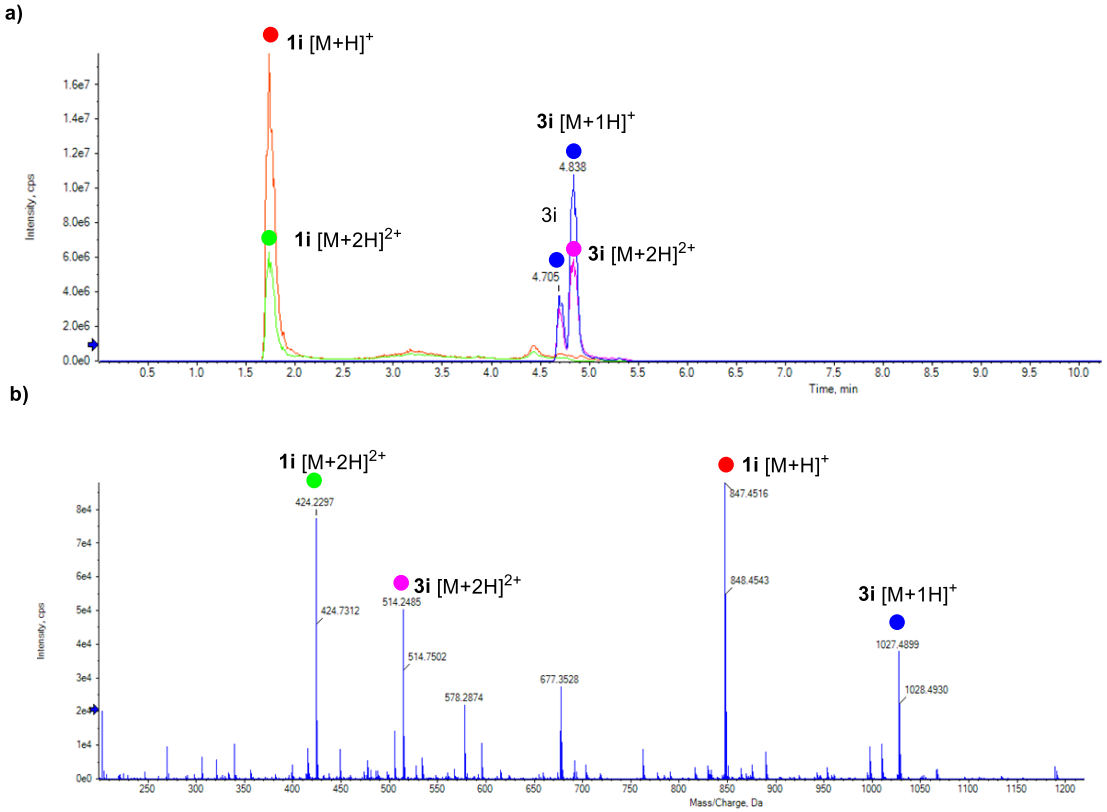


Figure S38. Trial 3: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (50 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

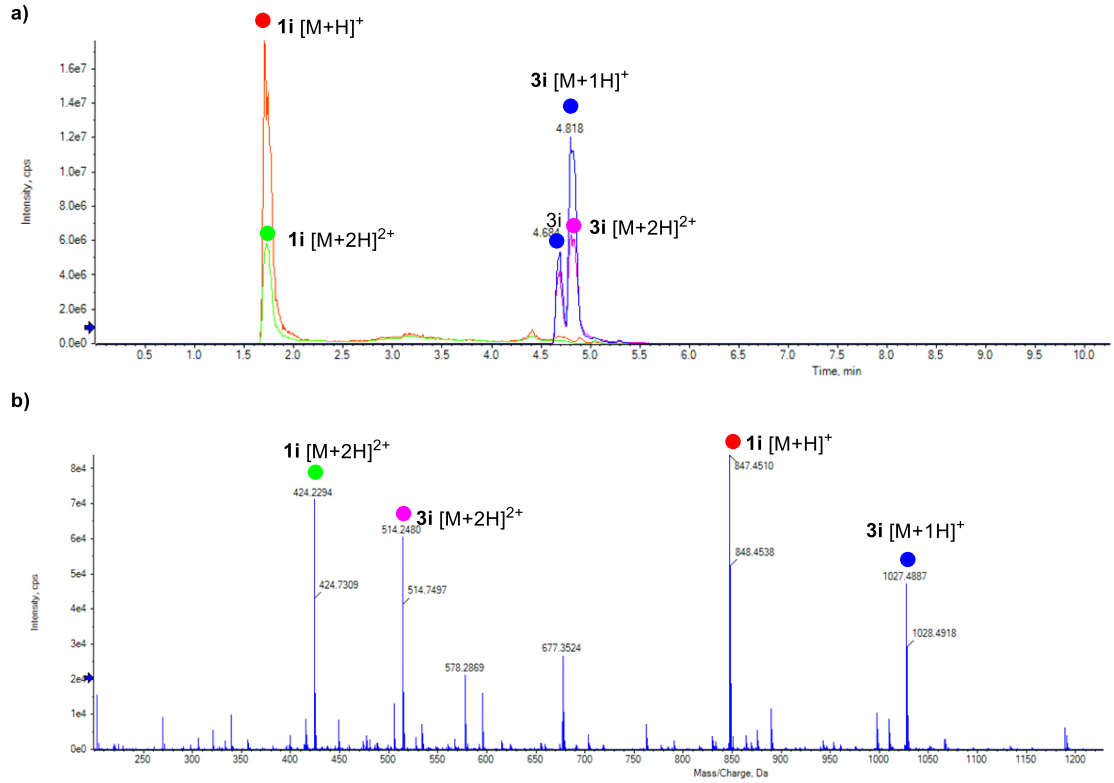


Figure S39. Trial 1: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (100 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

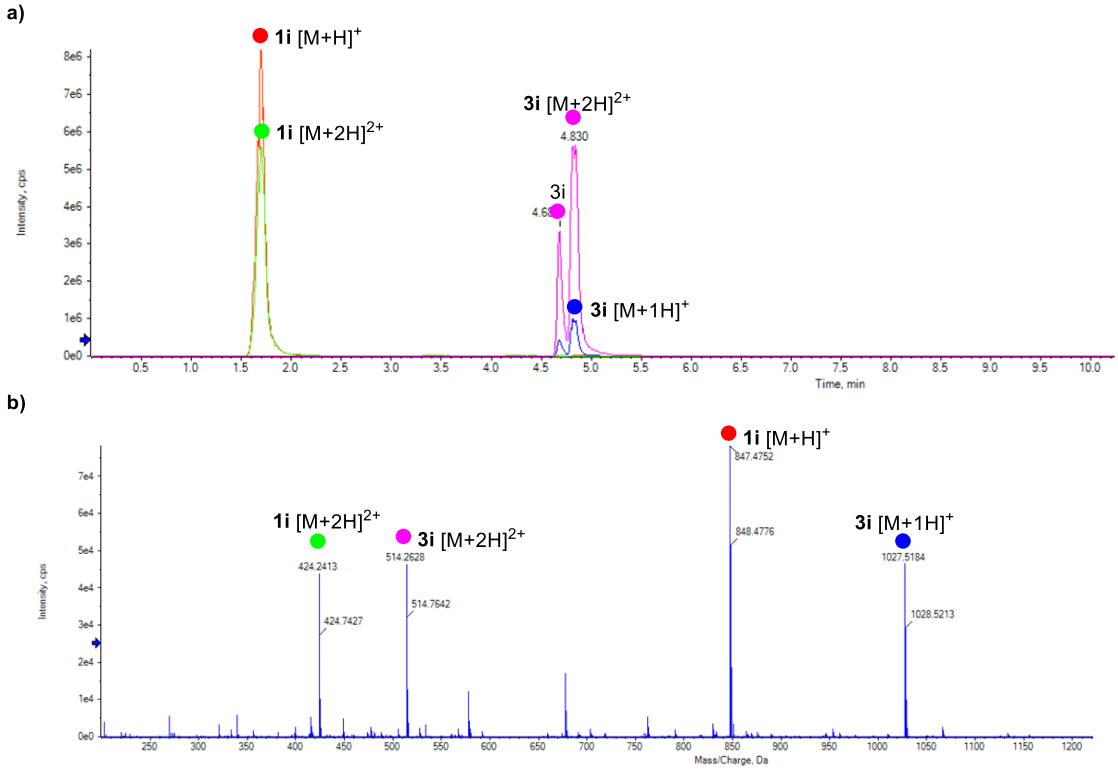


Figure S40. Trial 2: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (100 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

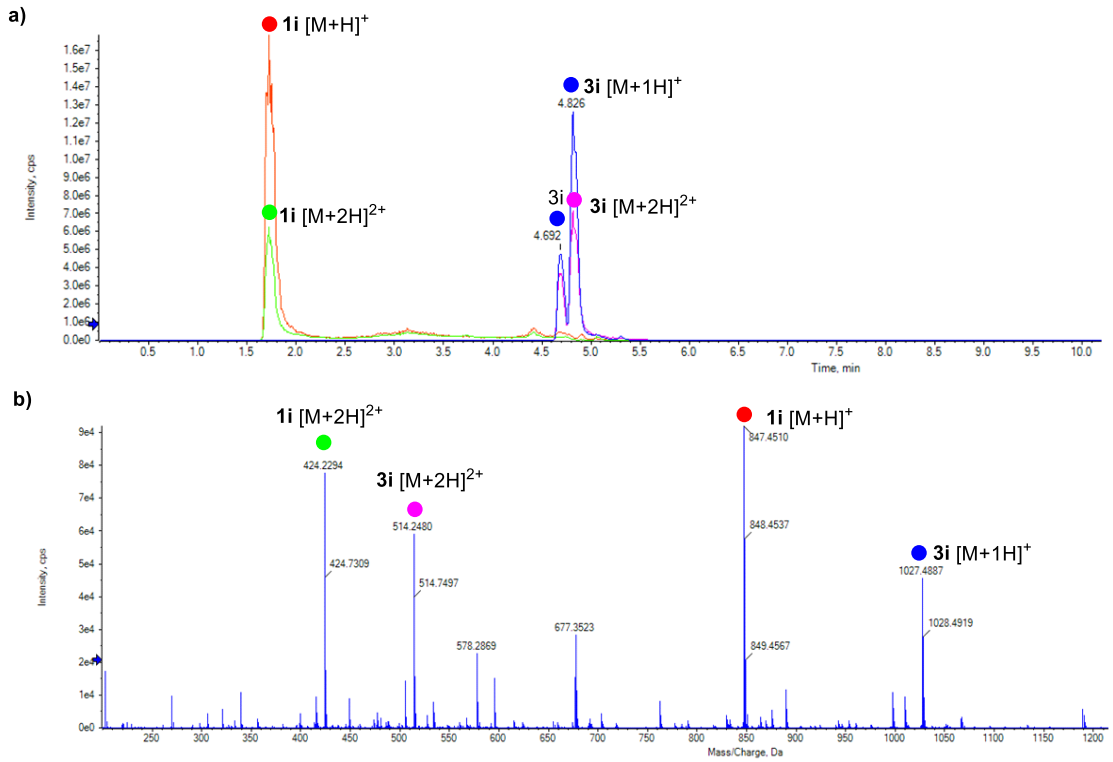


Figure S41. Trial 3: N-Gly-aminoalcohol (**3i**) formation with peptide (**1i**) using Gly-tag reagent **2** (100 equiv.). (a) XIC data for **1i** and **3i**. (b) ESI-MS data for **1i** and **3i**.

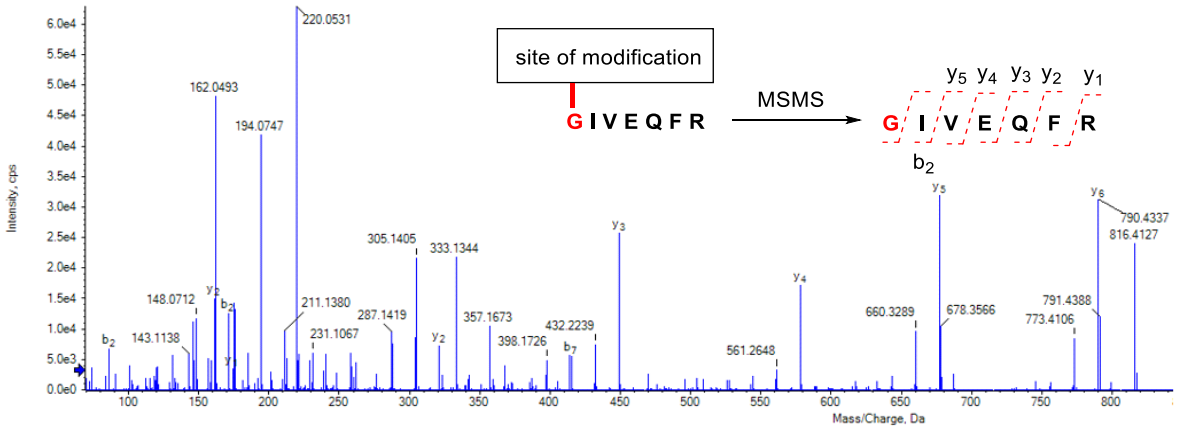


Figure S42. MS-MS spectrum of N-Gly-aminoalcohol (**3i**) peptide GIVEQFR (G1-R7, *m/z* 1027.4887 $[M+H]^+$) confirms the N-Gly modification.

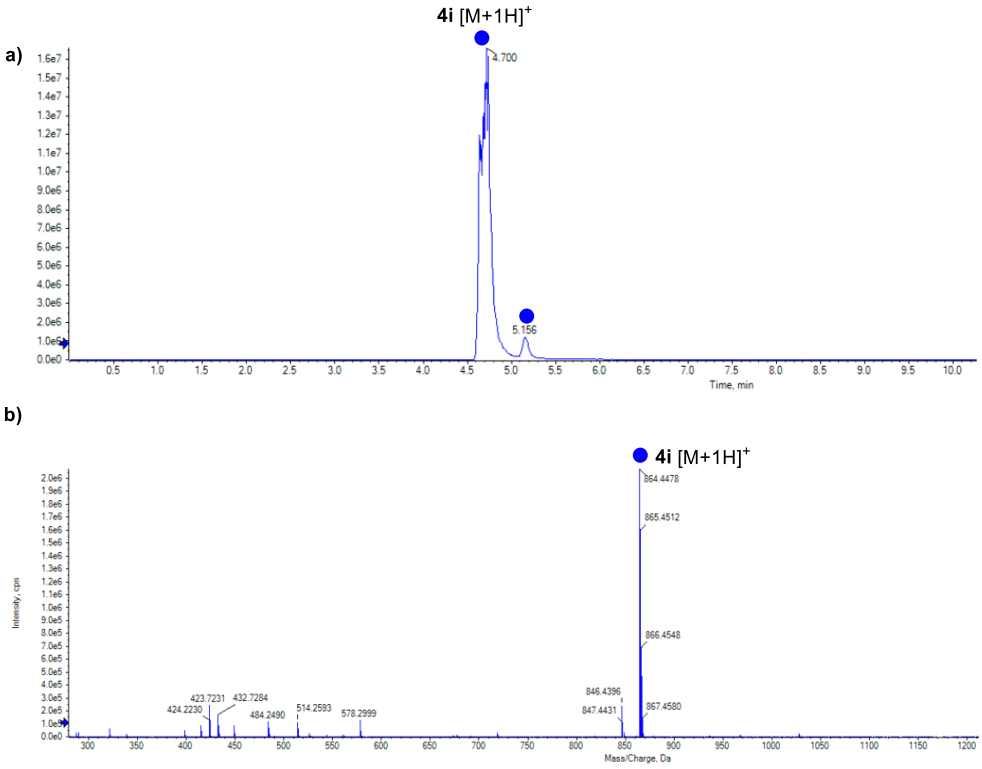


Figure S43. Periodate cleavage of N-Gly-aminoalcohol GIVEQFR(**3i)**. (a) XIC data for **4i**. (b) ESI-MS data for **4i**.

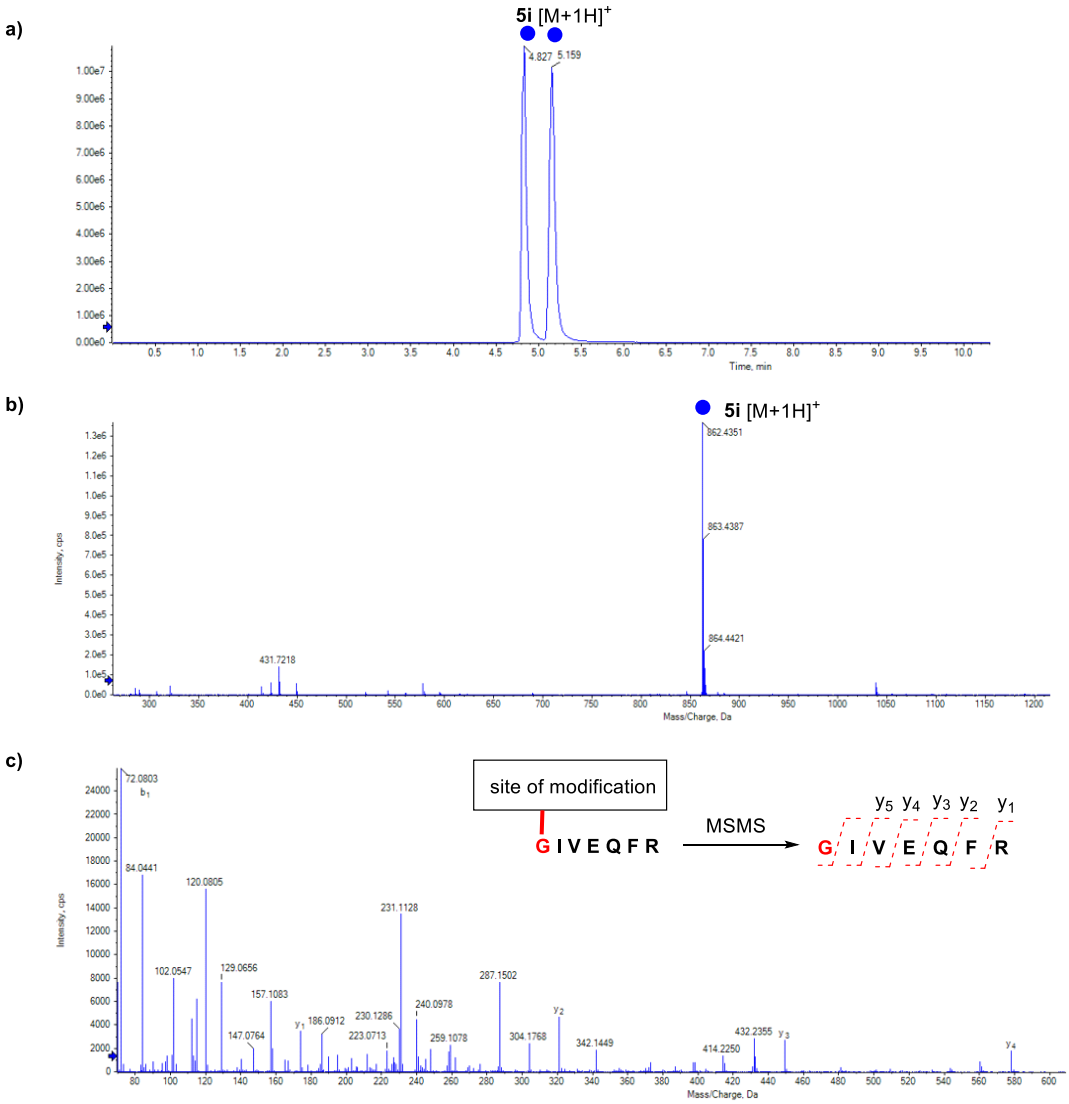


Figure S44. (a) XIC data for N-Gly-oxime GIVEQFR (**5i**). (b) ESI-MS data for **5i**. (c) MS-MS spectrum of N-Gly-oxime GIVEQFR (*m/z* 862.435 $[M+H]^+$) confirms the N-Gly modification.

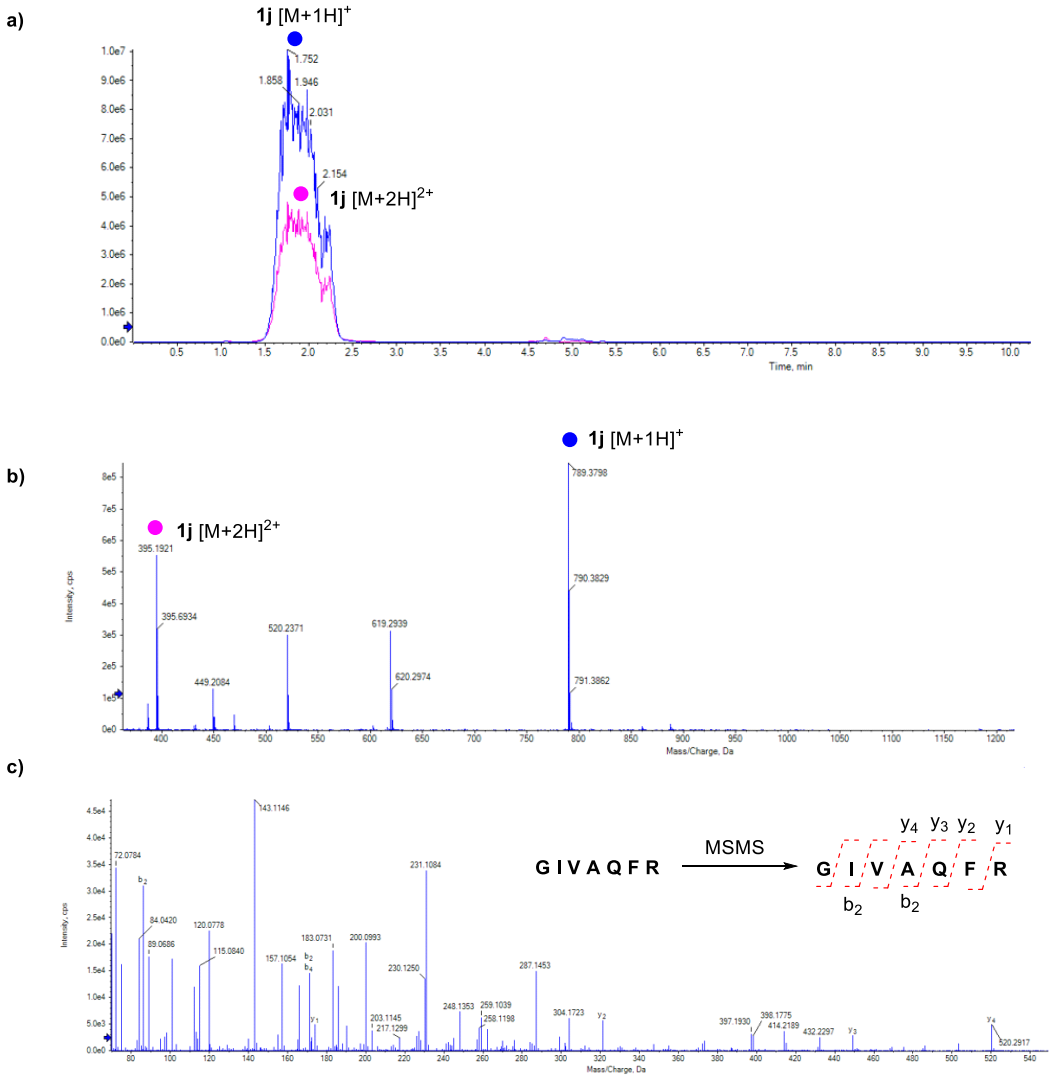


Figure S45. (a) XIC data for peptide GIVAQFR (**1j**). (b) ESI-MS data for peptide **1j**. (c) MS-MS spectrum of **1j**.

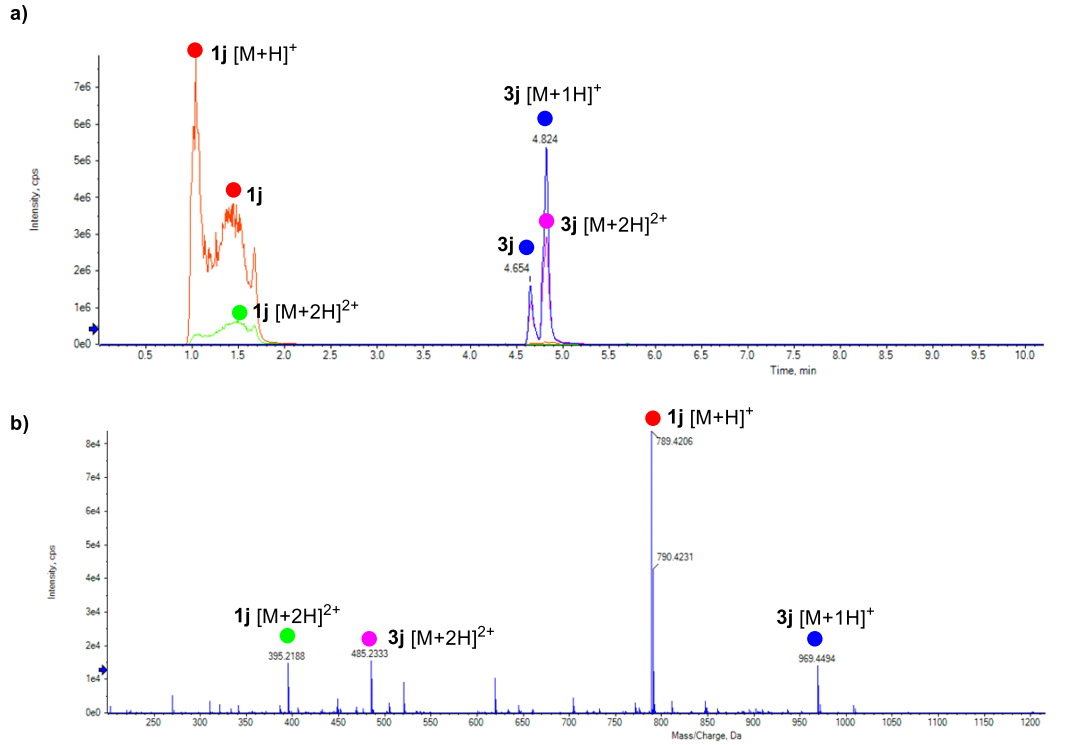


Figure S46. Trial 1: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (25 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

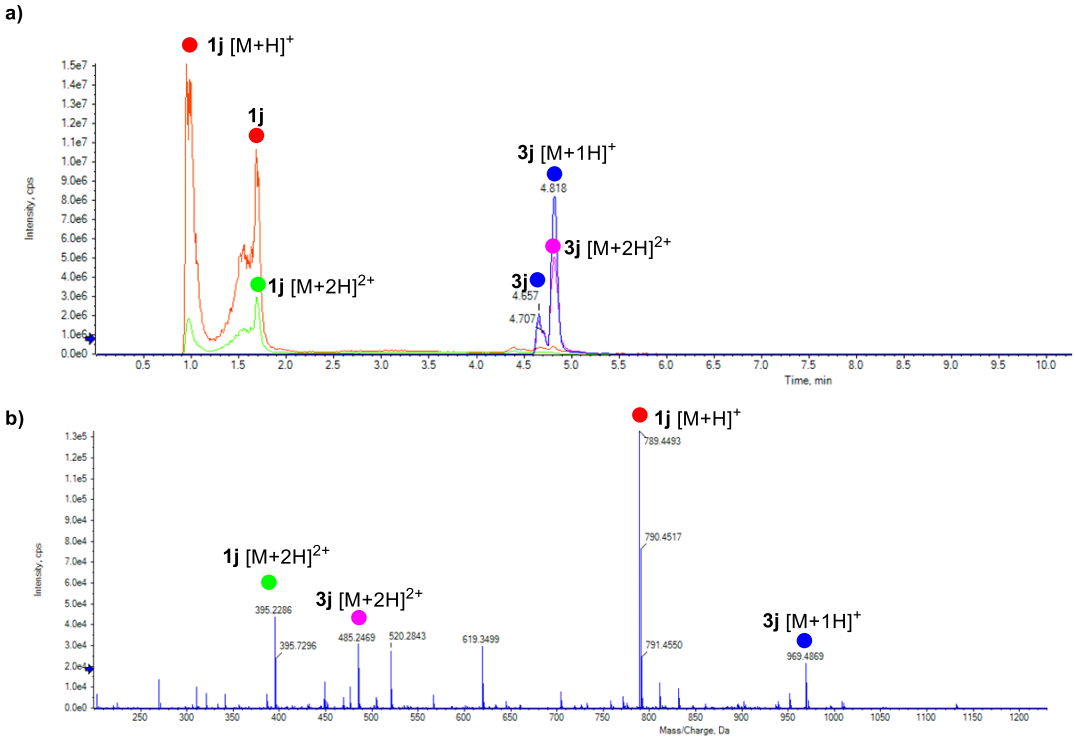


Figure S47. Trial 2: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (25 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

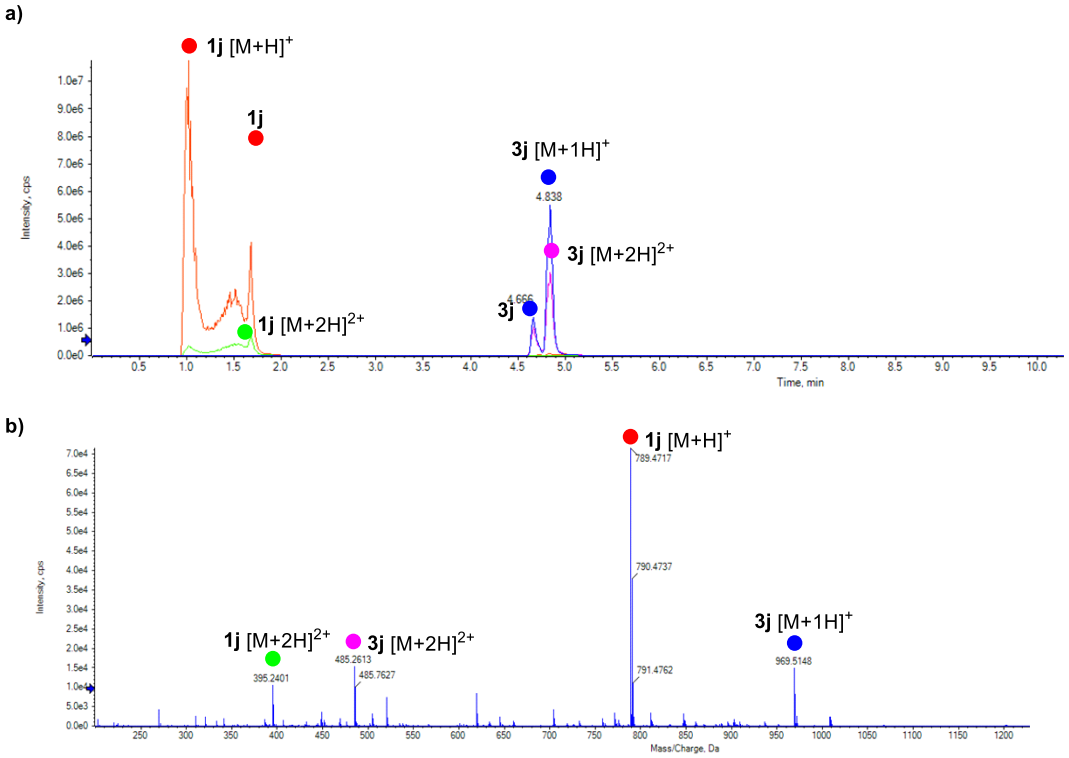


Figure S48. Trial 3: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (25 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

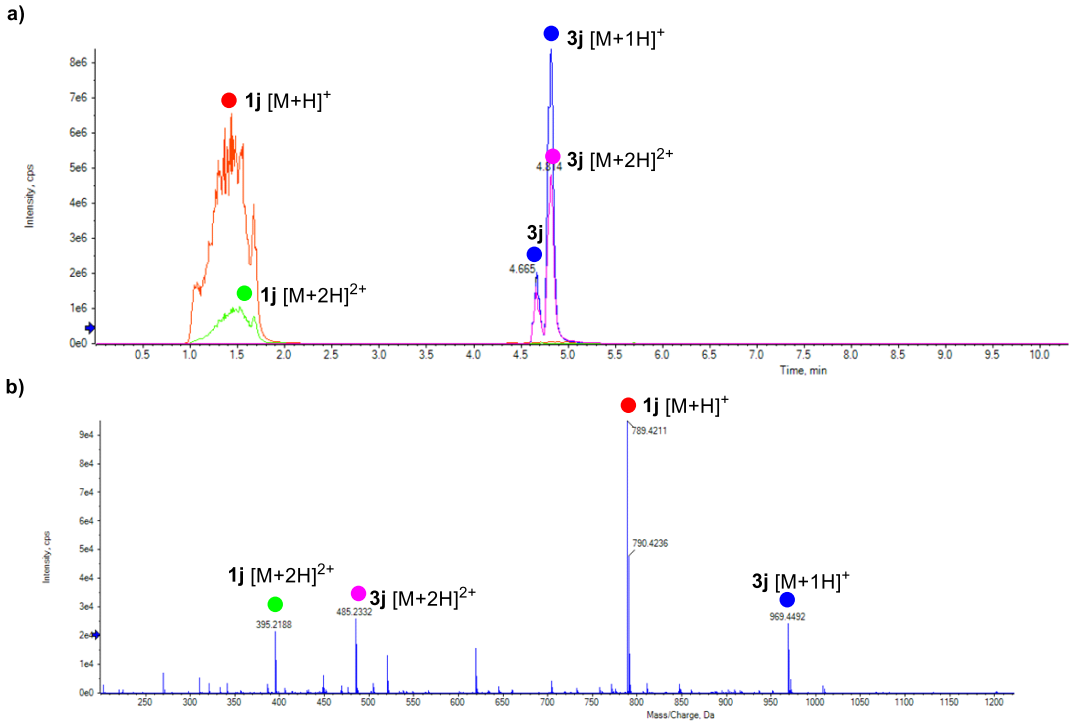


Figure S49. Trial 1: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (50 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

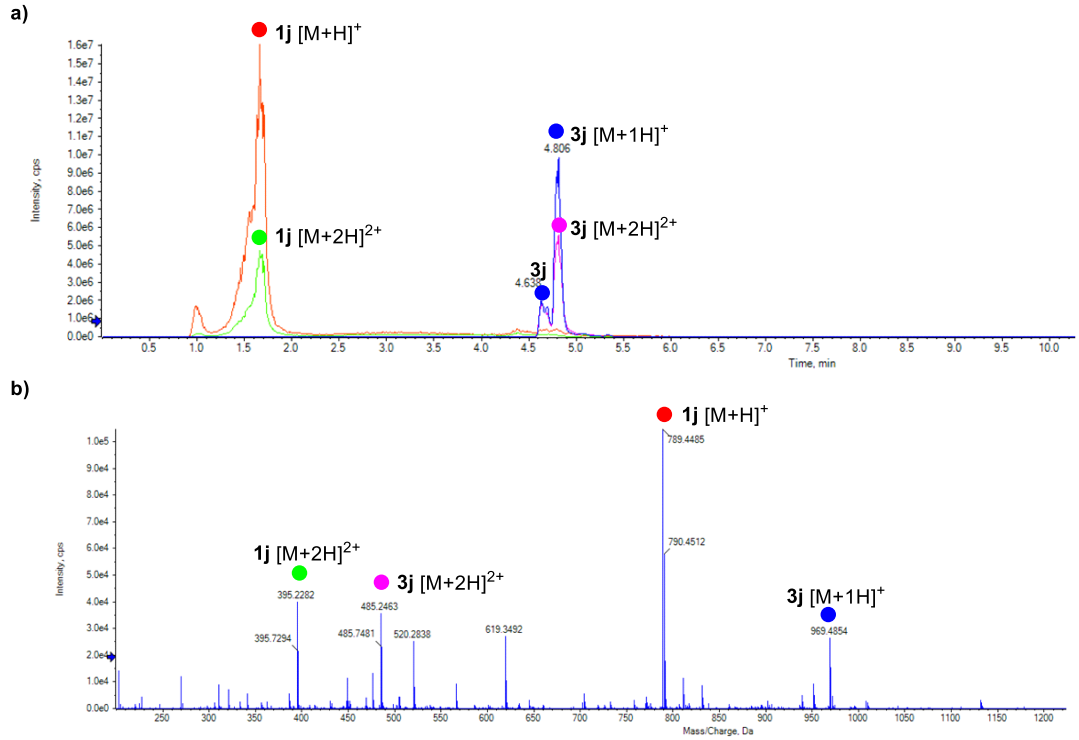


Figure S50. Trial 2: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (50 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

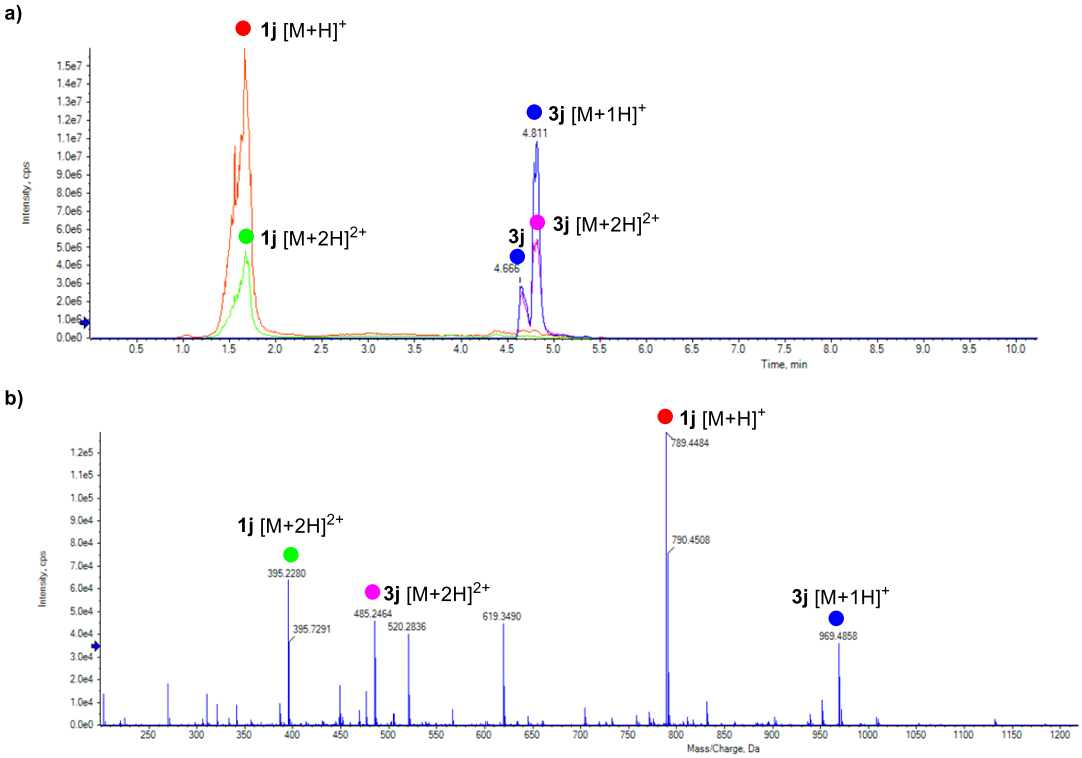


Figure S51. Trial 3: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (50 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

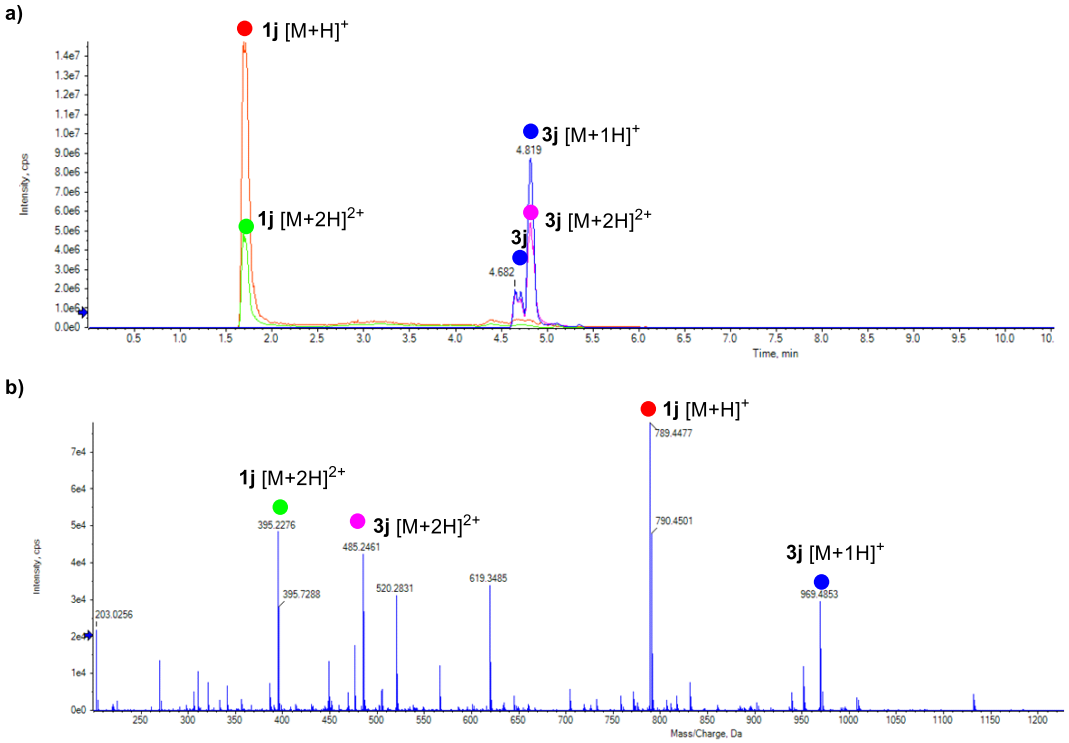


Figure S52. Trial 1: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (100 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

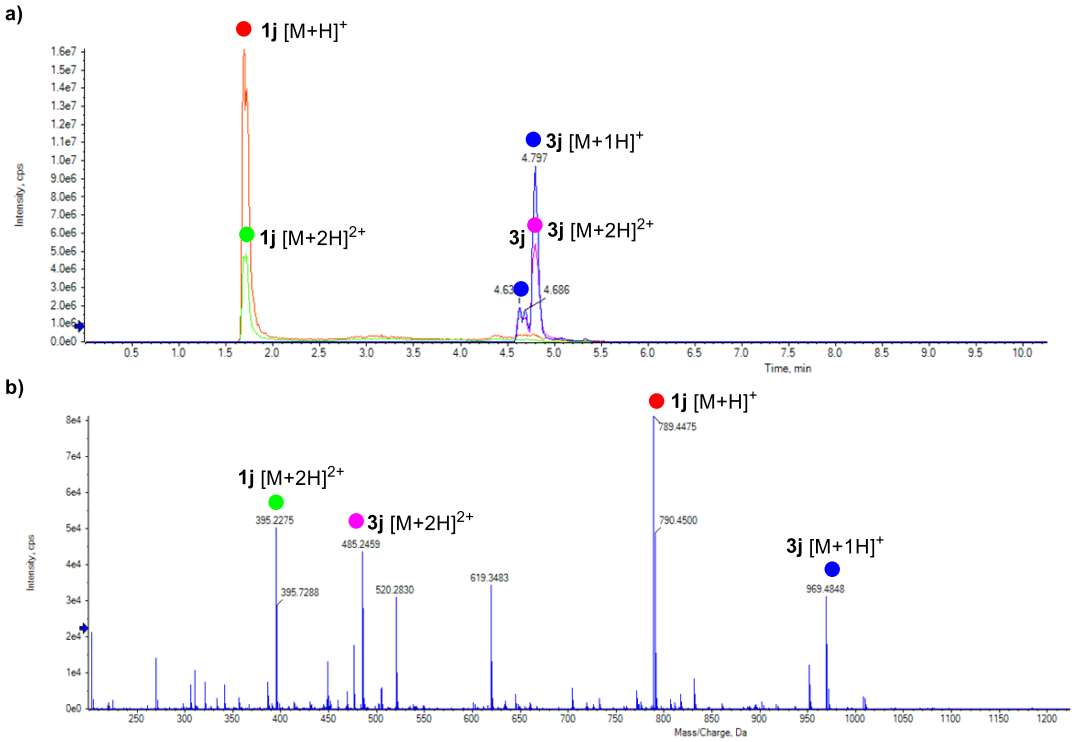


Figure S53. Trial 2: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (100 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

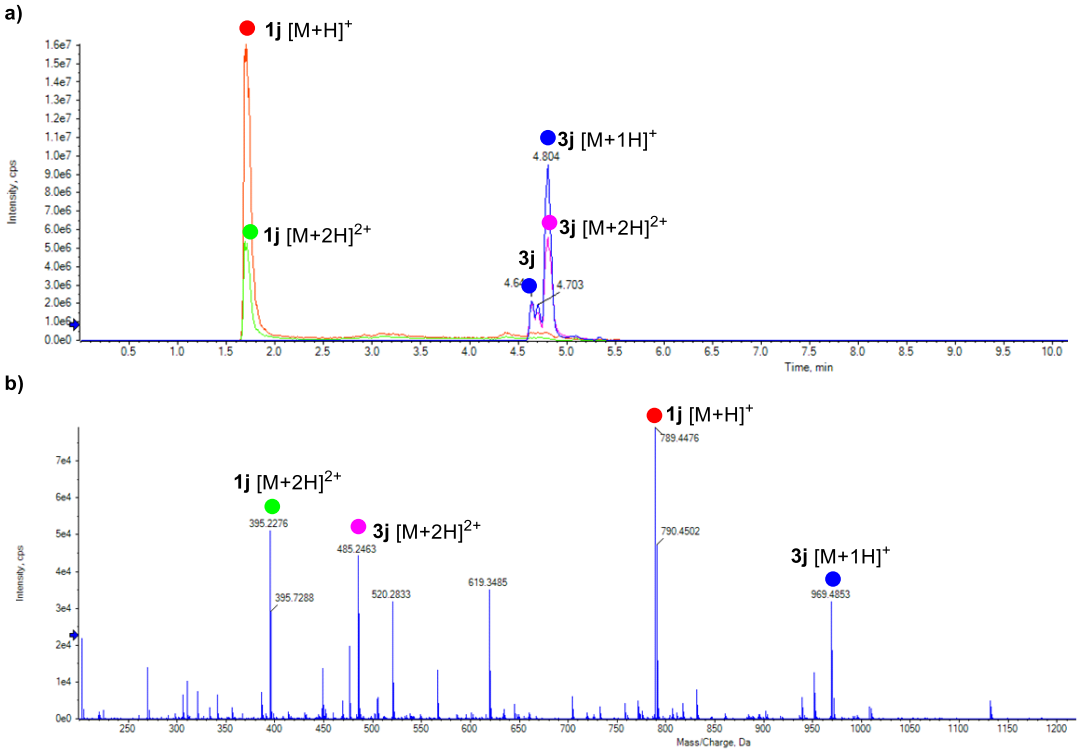


Figure S54. Trial 3: N-Gly-aminoalcohol (**3j**) formation with peptide (**1j**) using Gly-tag reagent **2** (100 equiv.). (a) XIC data for **1j** and **3j**. (b) ESI-MS data for **1j** and **3j**.

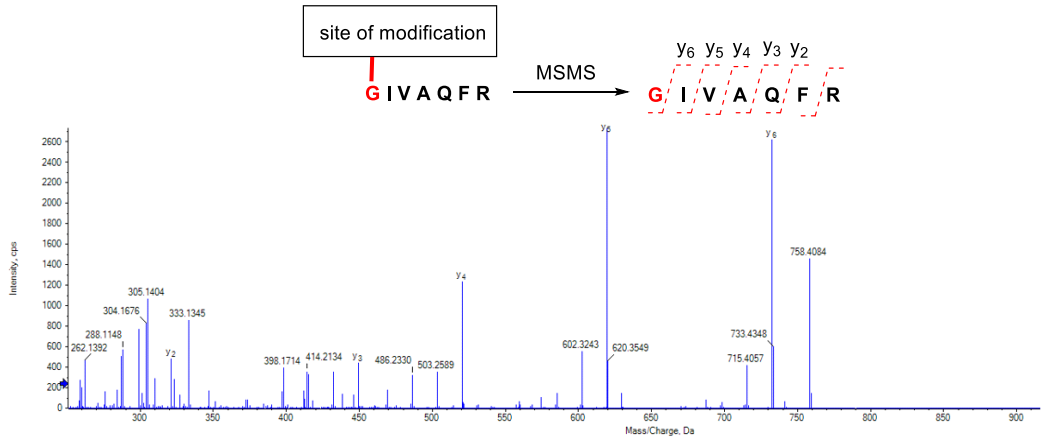


Figure S55. MS-MS spectrum of N-Gly-aminoalcohol (**3j**) peptide GIVAQFR (G1-R7, *m/z* 969.4853 $[M+H]^+$) confirms the N-Gly modification.

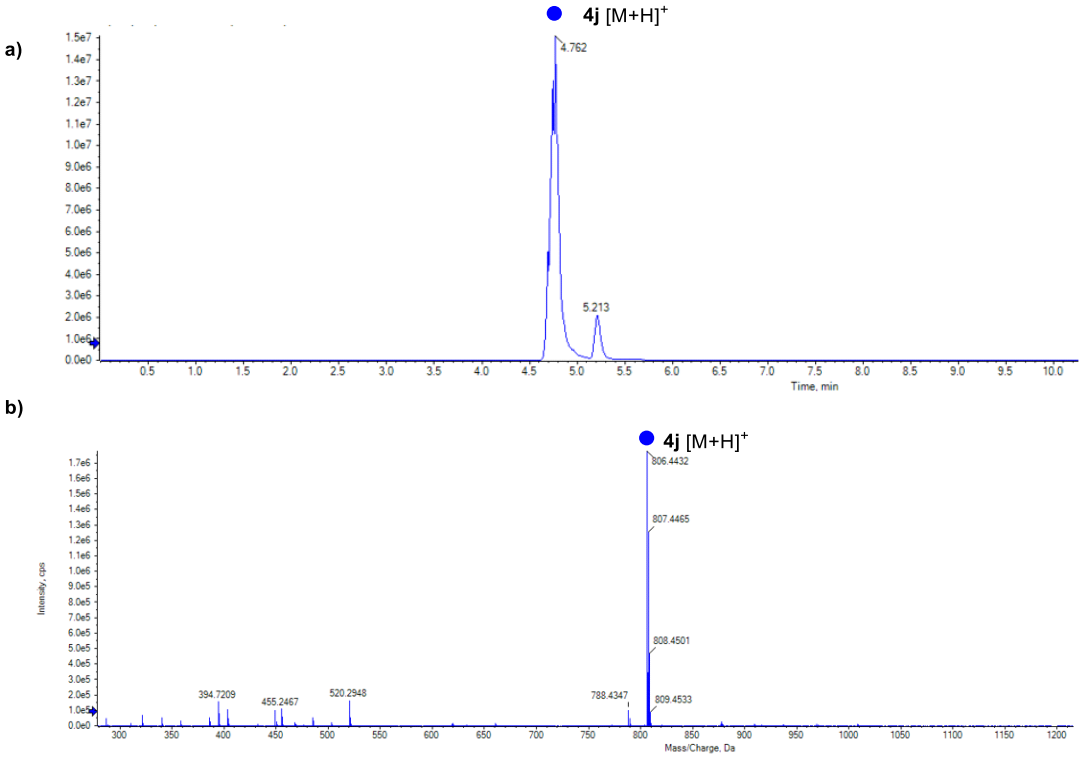


Figure S56. Periodate cleavage of N-Gly-aminoalcohol GIVEQFR(**3j)**. (a) XIC data for **4j**. (b) ESI-MS data for **4j**.

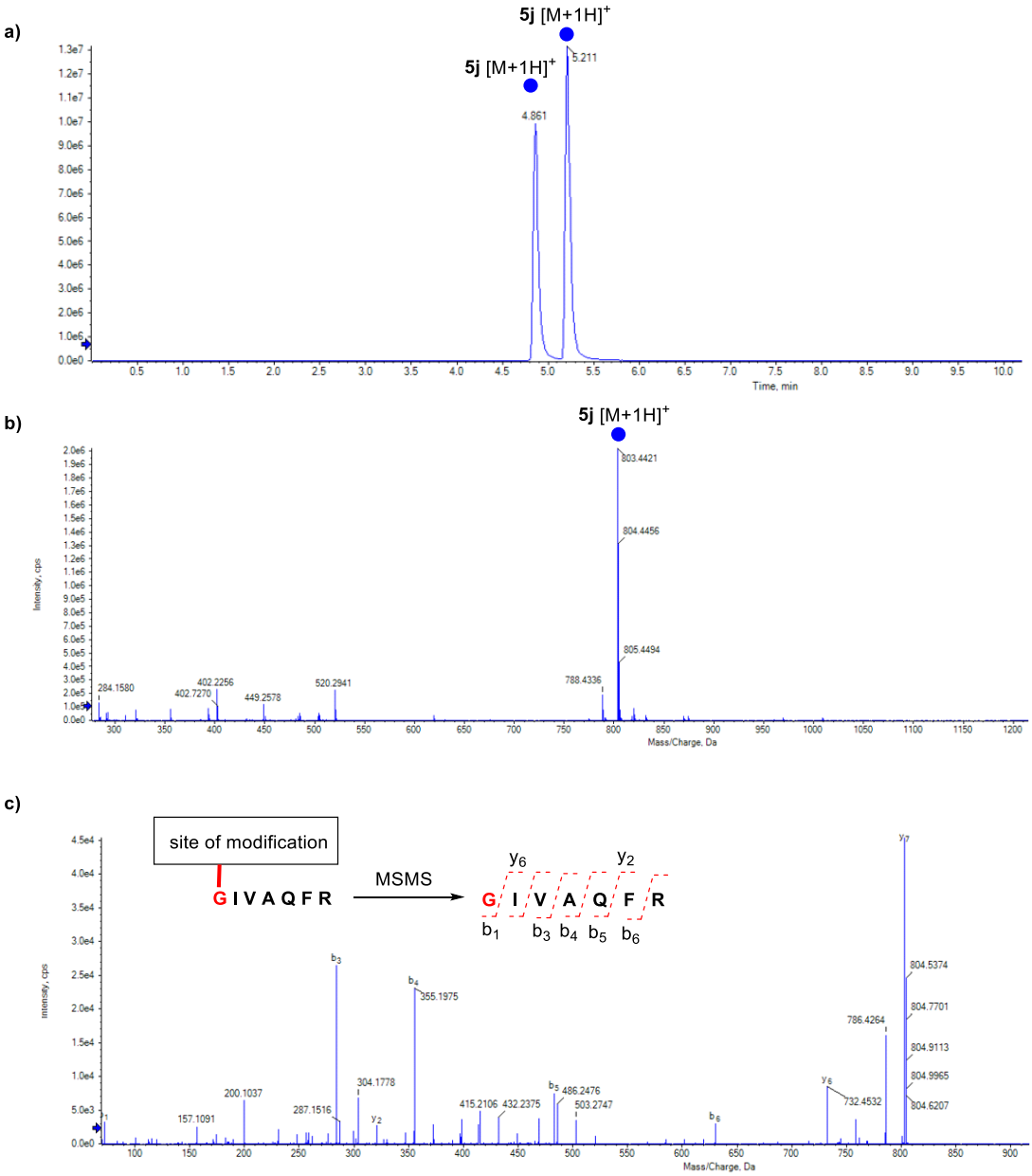


Figure S57. (a) XIC data for N-Gly-oxime GIVAQFR (**5j**). (b) ESI-MS data for **5j**. (c) MS-MS spectrum of N-Gly-oxime GIVAQFR (*m/z* 803.4421 $[M+H]^+$) confirms the N-Gly modification.

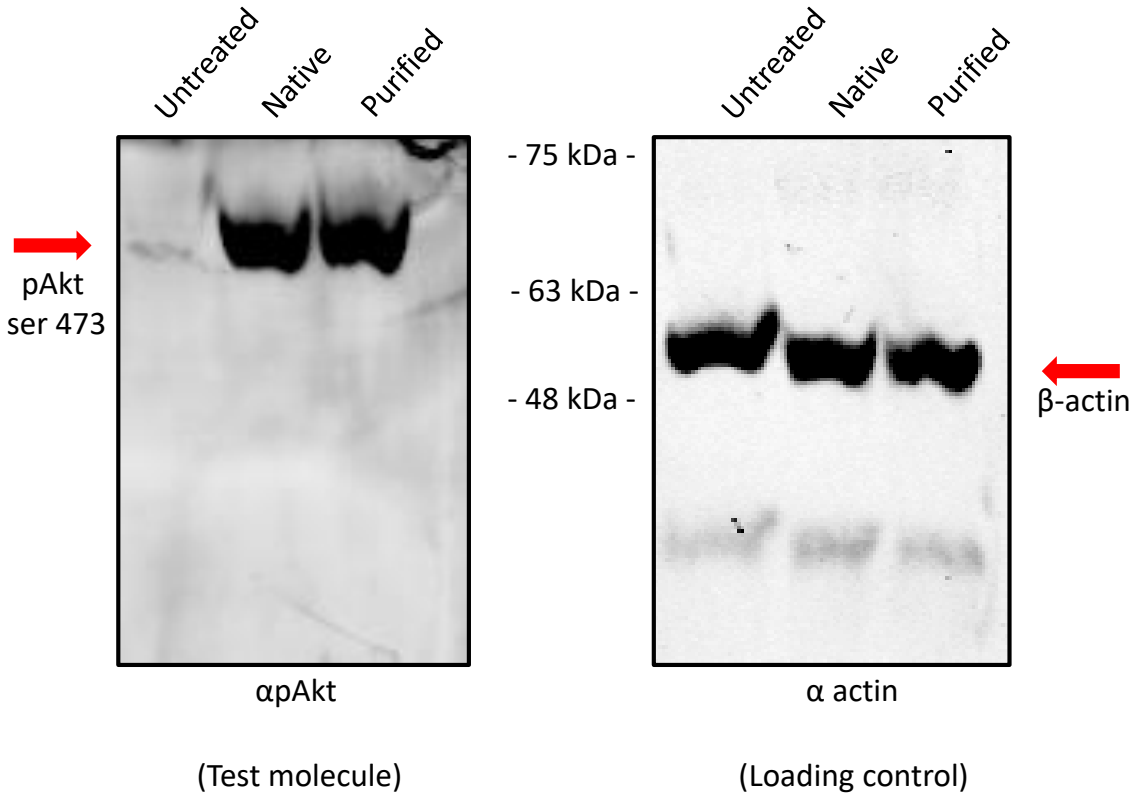


Figure S58. Raw images for Western blotting mediated detection of insulin receptor activation in HEK293T cell lysates by native (**1a**) and purified (**5a**) insulins

5. Protein sequence

A. Insulin
PDB ID: 1A7F
amino acid sequence:
Chain A: GIVEQCCTSICSLYQLENYCN
Chain B: FVNQHLCGSHLVEALYLVCGERGFFYTPKT

B. Melittin
PDB ID: 1BH1
amino acid sequence:
GIGAVLKVLTTGLPALISWIKRKRQQ

B. Myoglobin from equine skeletal muscle
PDB ID: 1WLA
amino acid sequence:
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAE
MKASEDLKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISD
AIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

6. NMR data

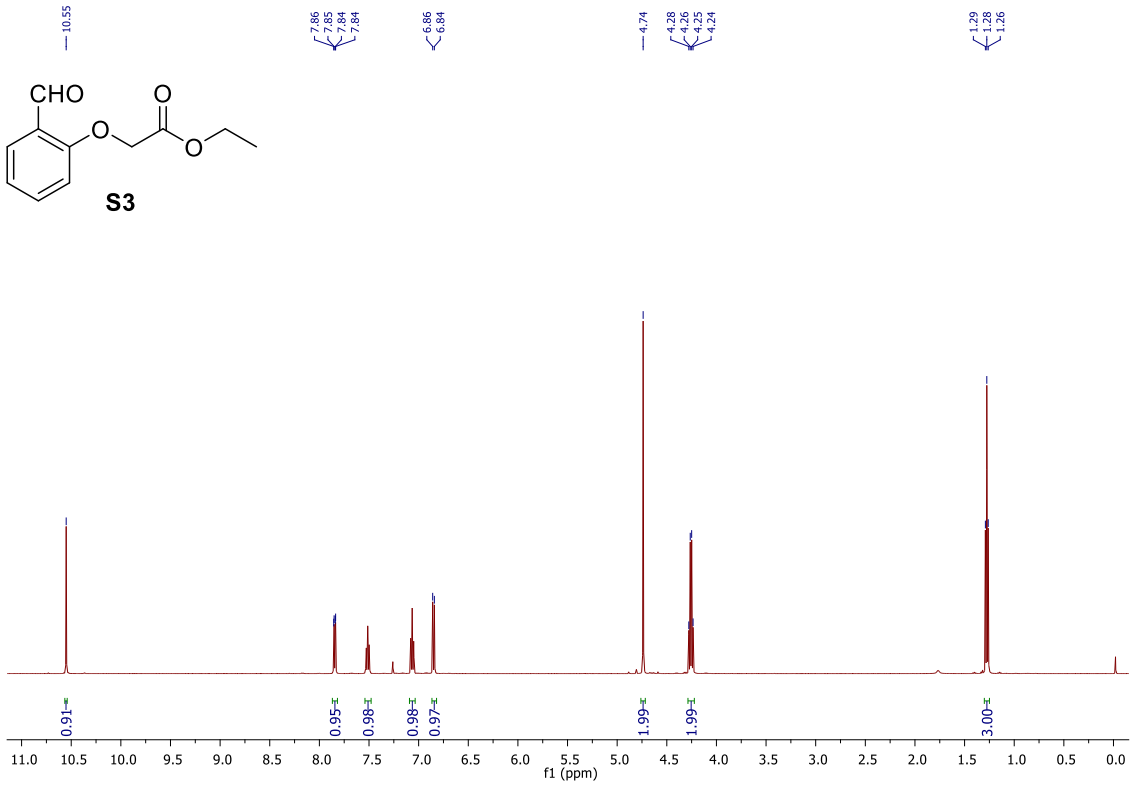


Figure S59. ^{1}H-NMR spectrum in $CDCl_3$ of compound **S2**.

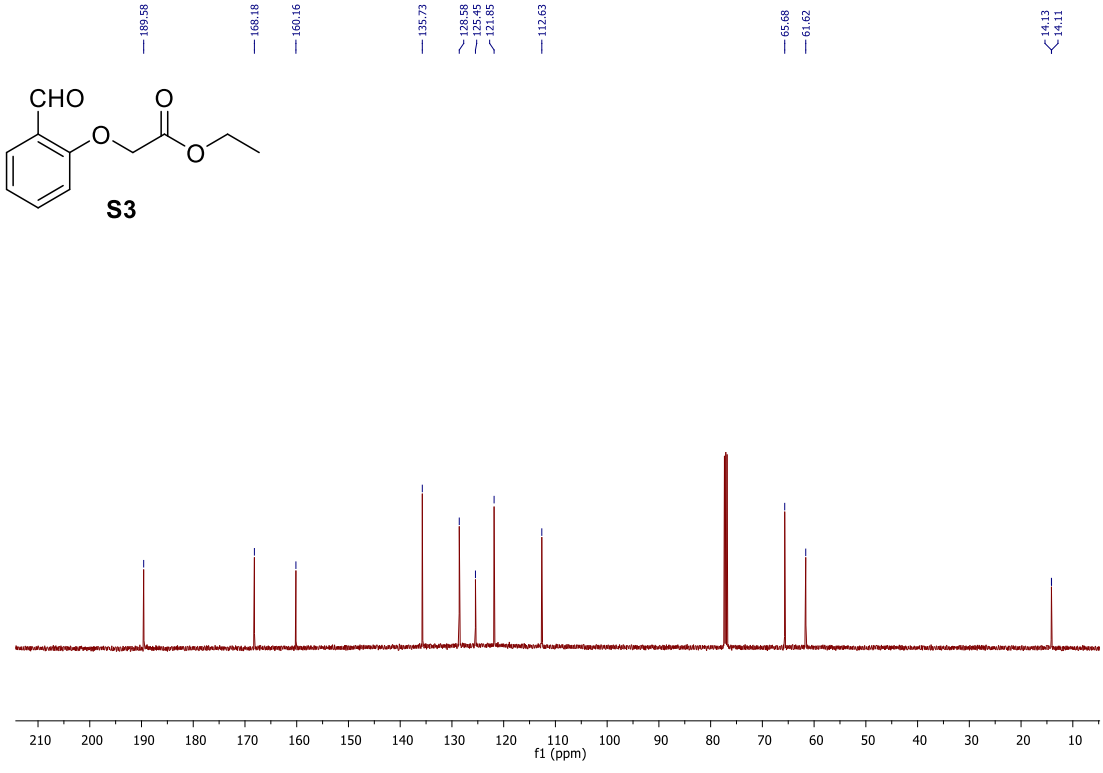


Figure S60. ^{13}C-NMR spectrum in $CDCl_3$ of compound **S2**.

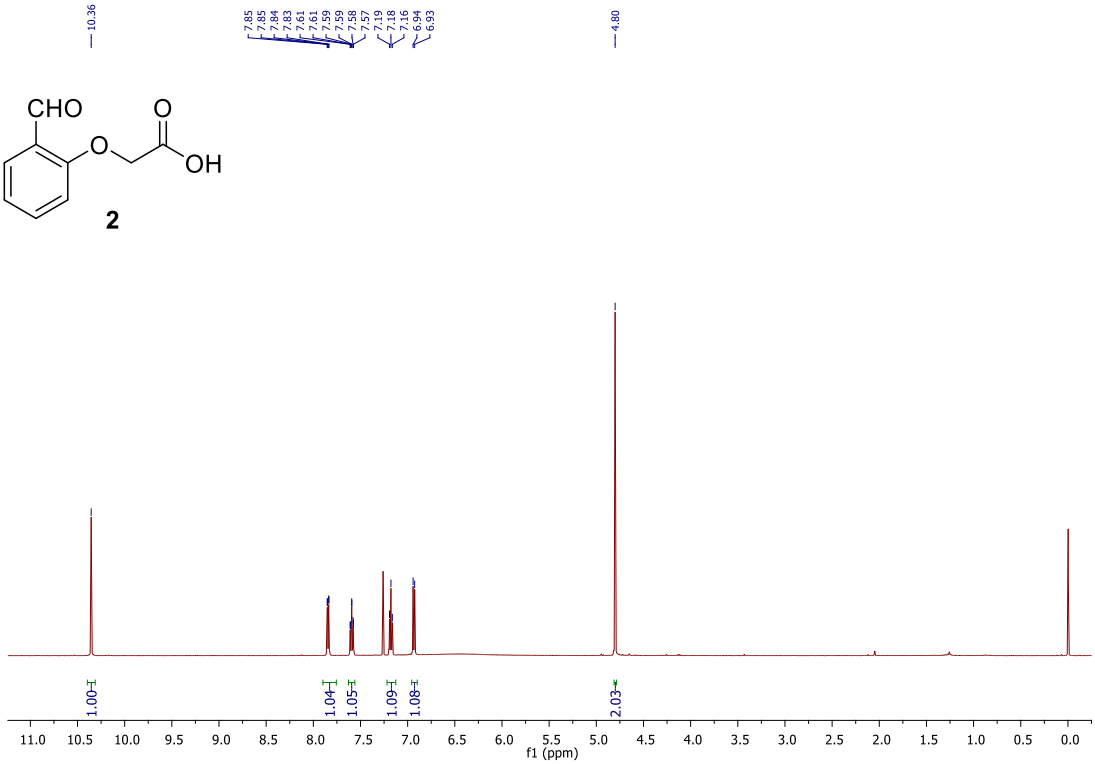


Figure S61. ^{1}H-NMR spectrum in $(CD_3)_2CO$ of compound **2**.

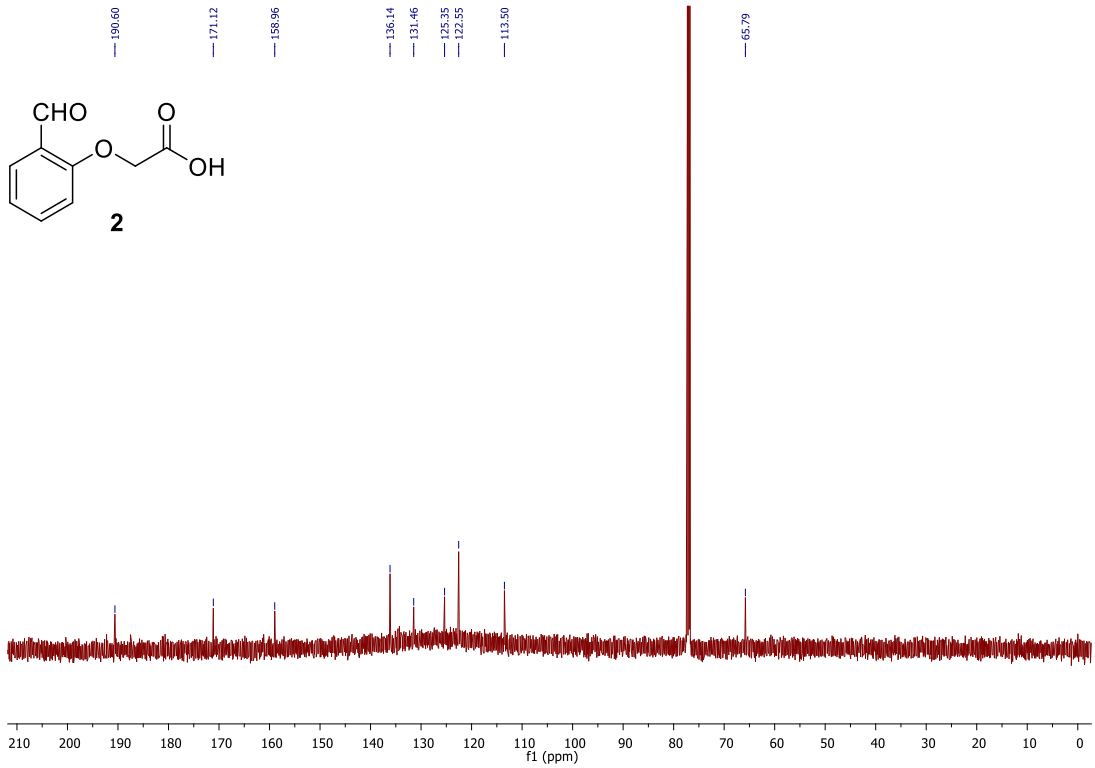


Figure S62. ^{13}C-NMR spectrum in $(CD_3)_2CO$ of compound **2**.

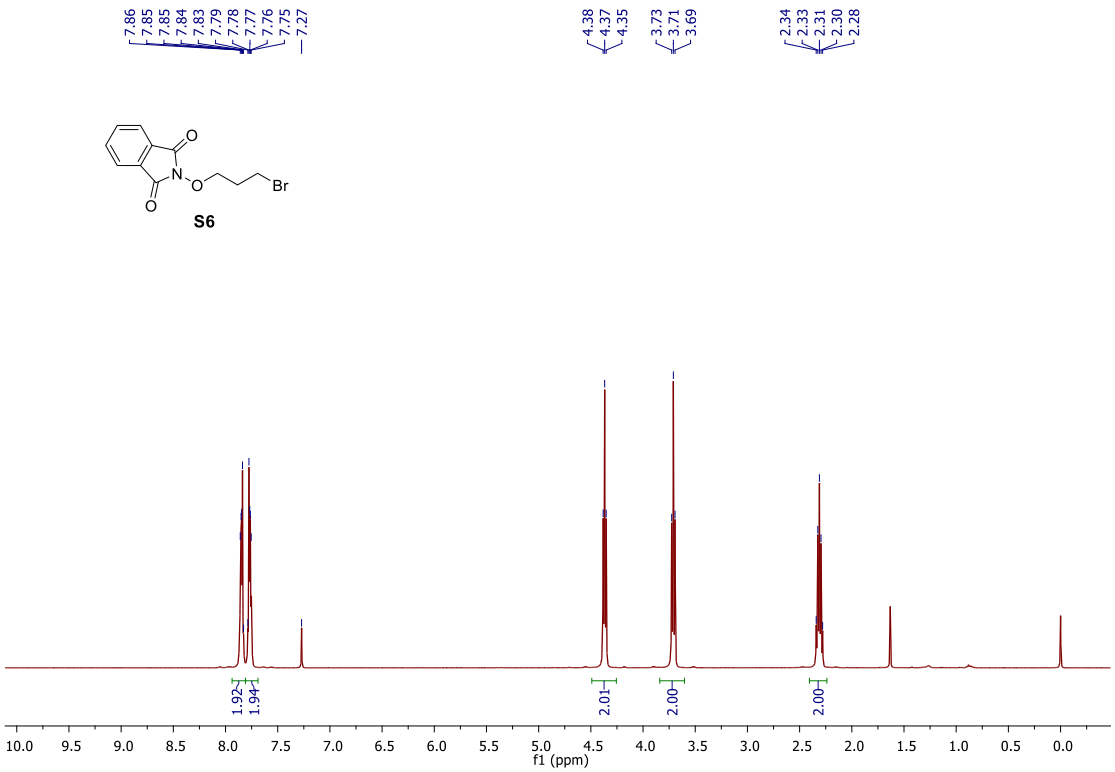


Figure S63. ^{1}H NMR spectrum in $CDCl_3$ of compound **S6**.

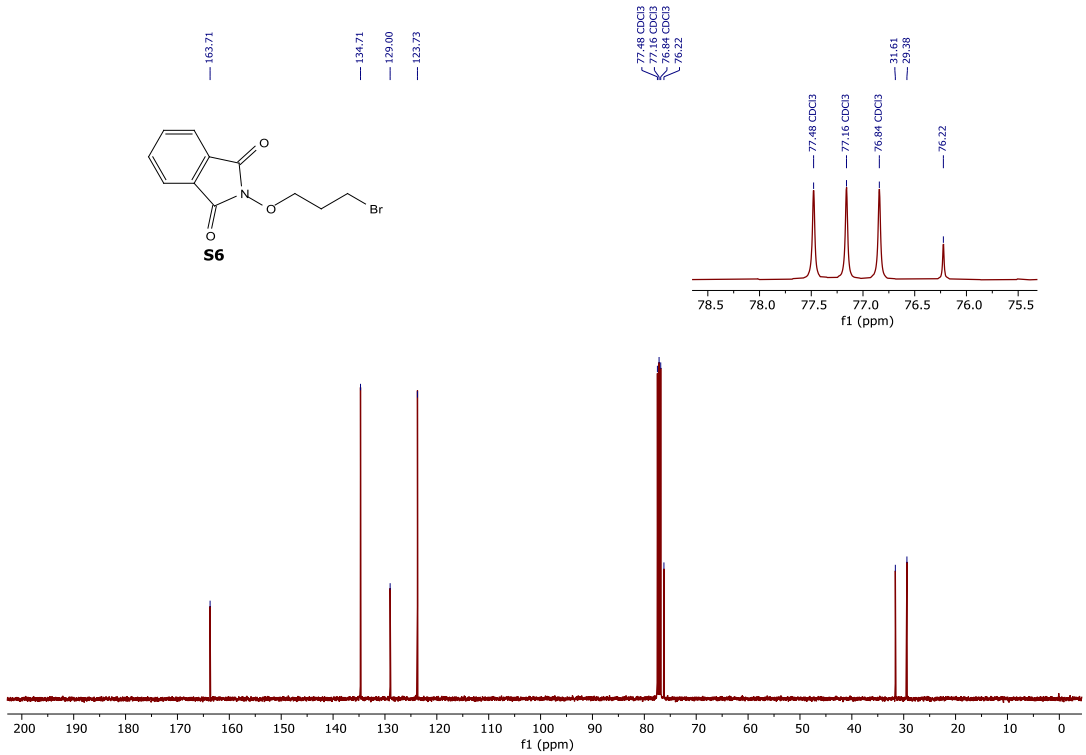


Figure S64. ^{13}C NMR spectrum in $CDCl_3$ of compound **S6**.

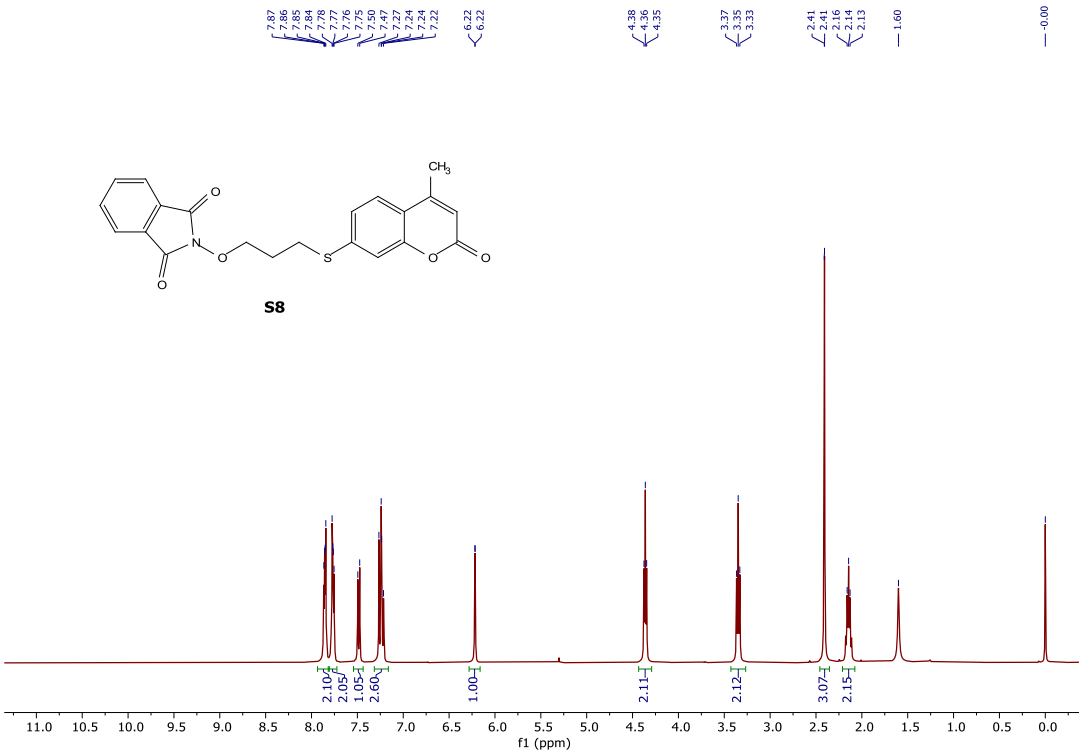


Figure S65. ^{1}H NMR spectrum in $CDCl_3$ of compound **S8**.

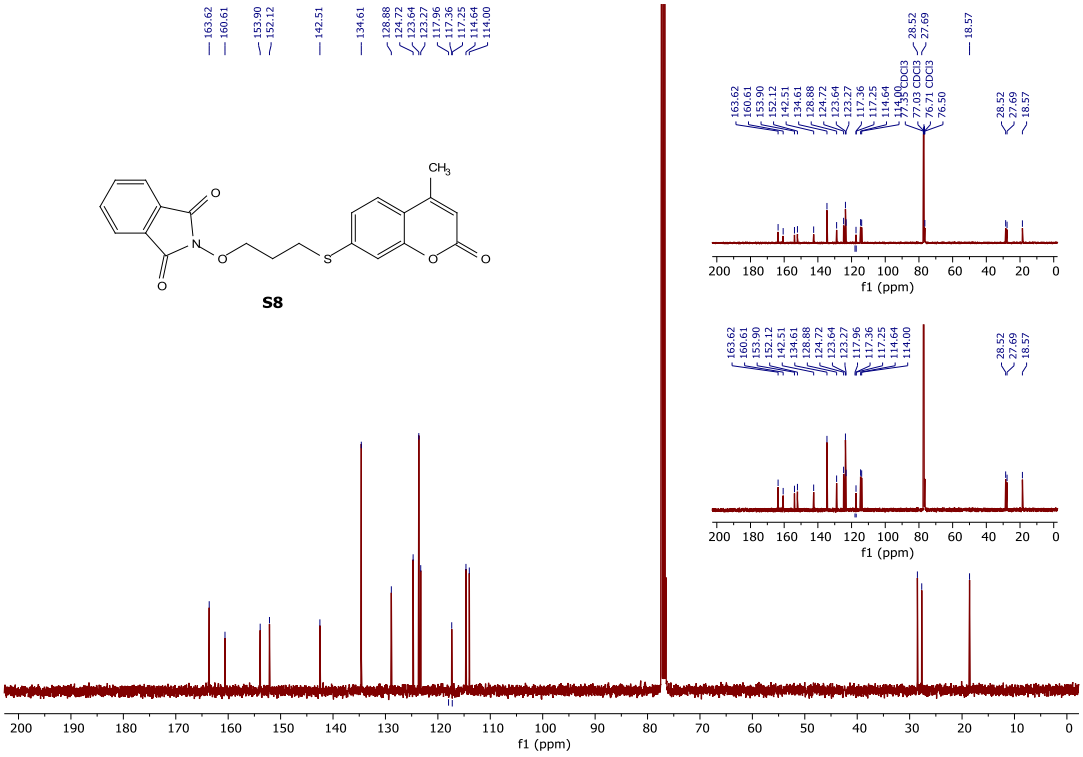


Figure S66. ^{13}C NMR spectrum in $CDCl_3$ of compound **S8**.

Figure S67. ^{1}H NMR spectrum in $CDCl_3$ of compound **6a**.

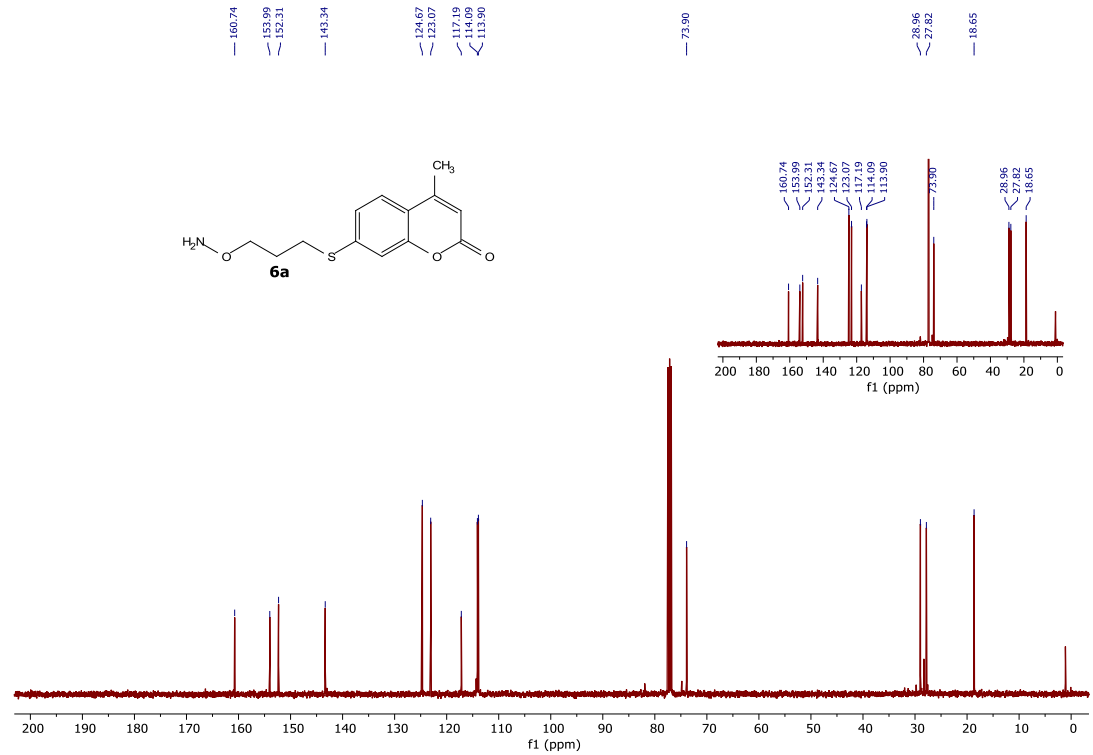


Figure S68. ^{13}C NMR spectrum in $CDCl_3$ of compound **6a**.

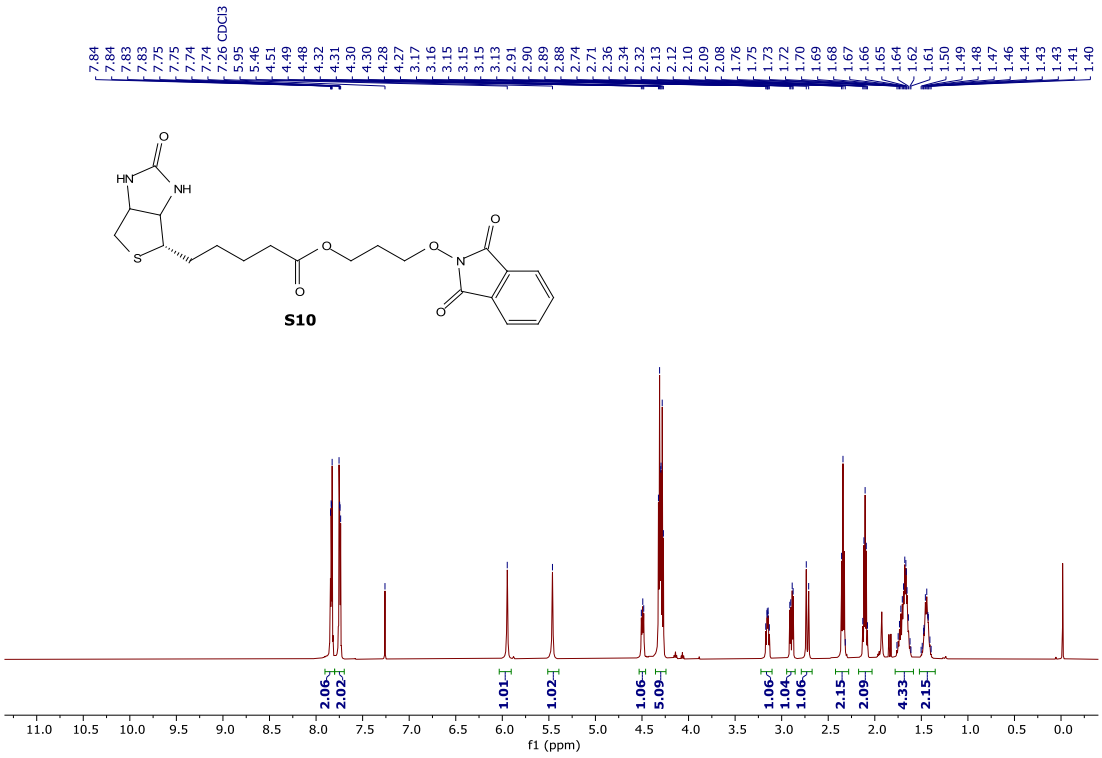


Figure S69. ^{1}H NMR spectrum in $CDCl_3$ of compound **S10**.

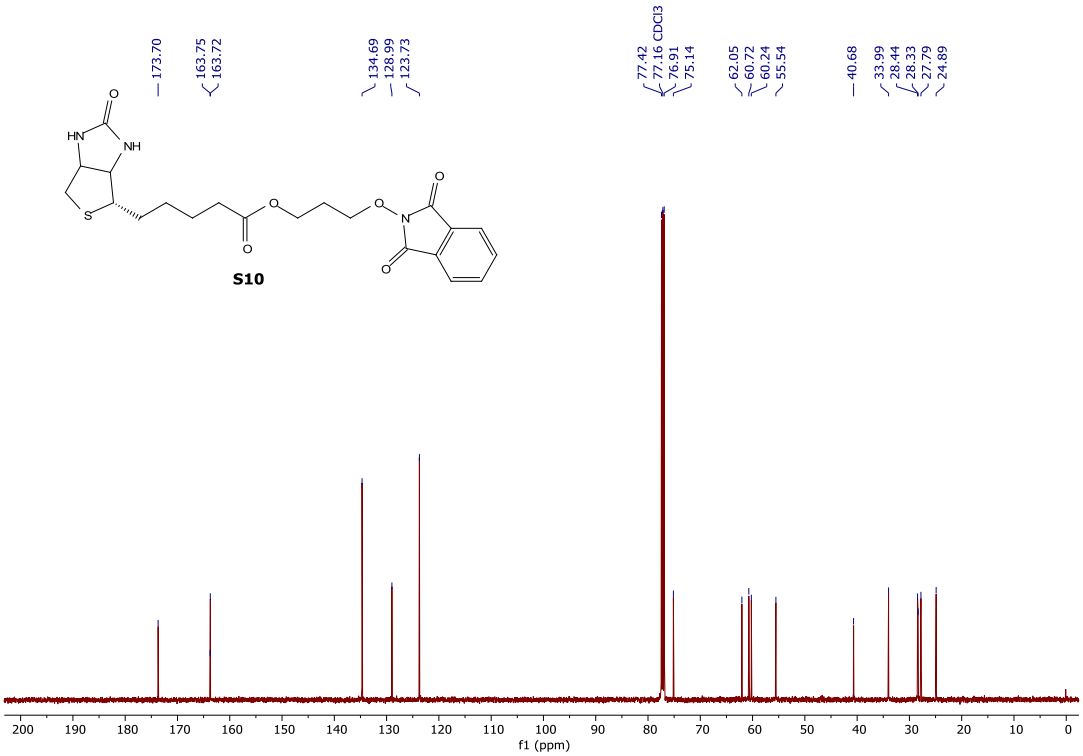


Figure S70. ^{13}C NMR spectrum in $CDCl_3$ of compound **S10**.

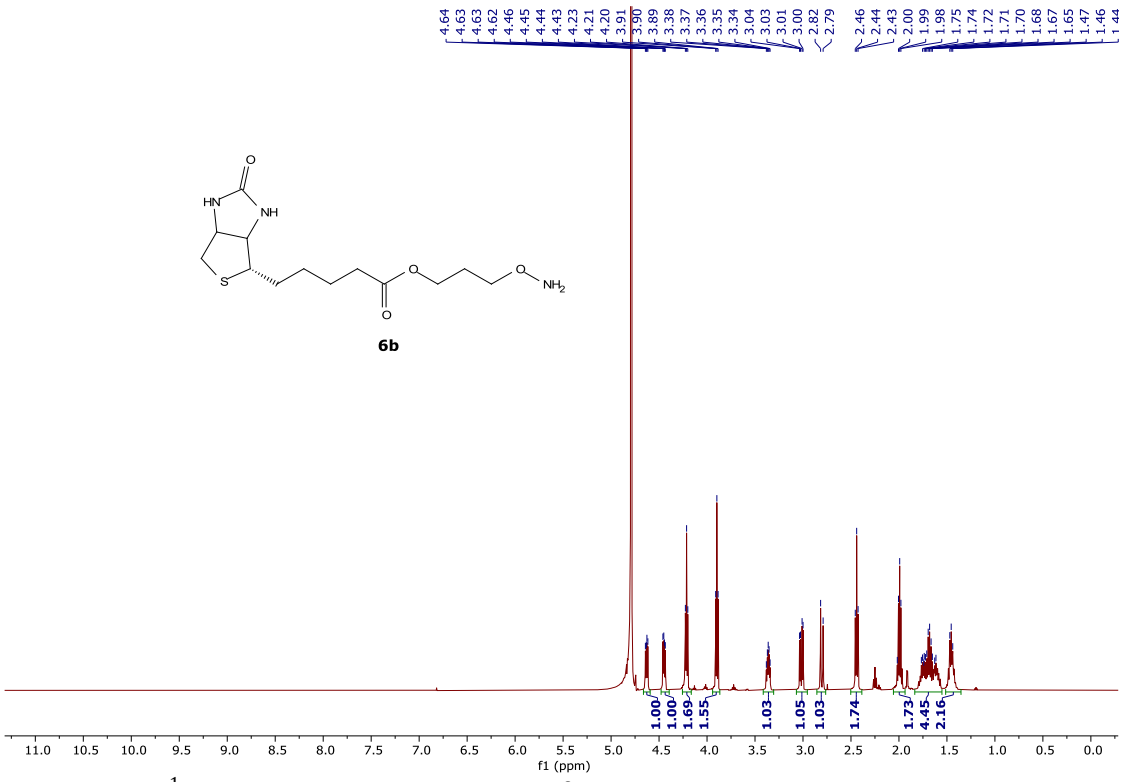


Figure S71. ^{1}H NMR spectrum in $CDCl_3$ of compound **6b**.

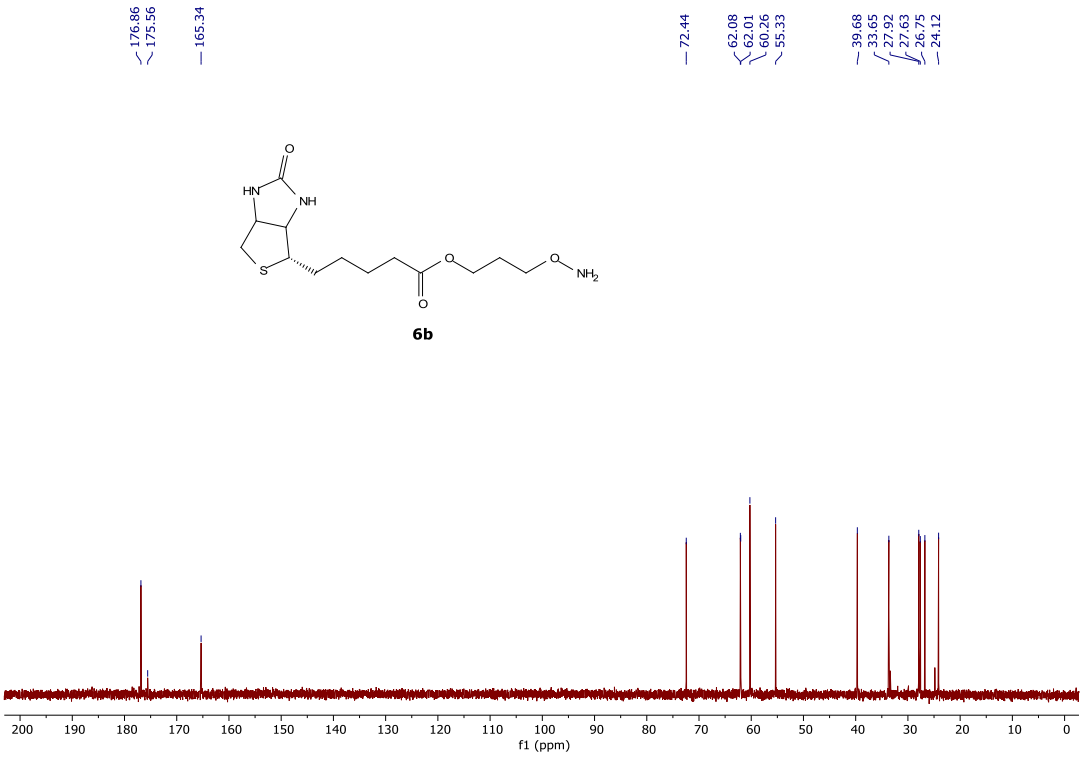


Figure S72. ^{13}C NMR spectrum in $CDCl_3$ of compound **6b**.

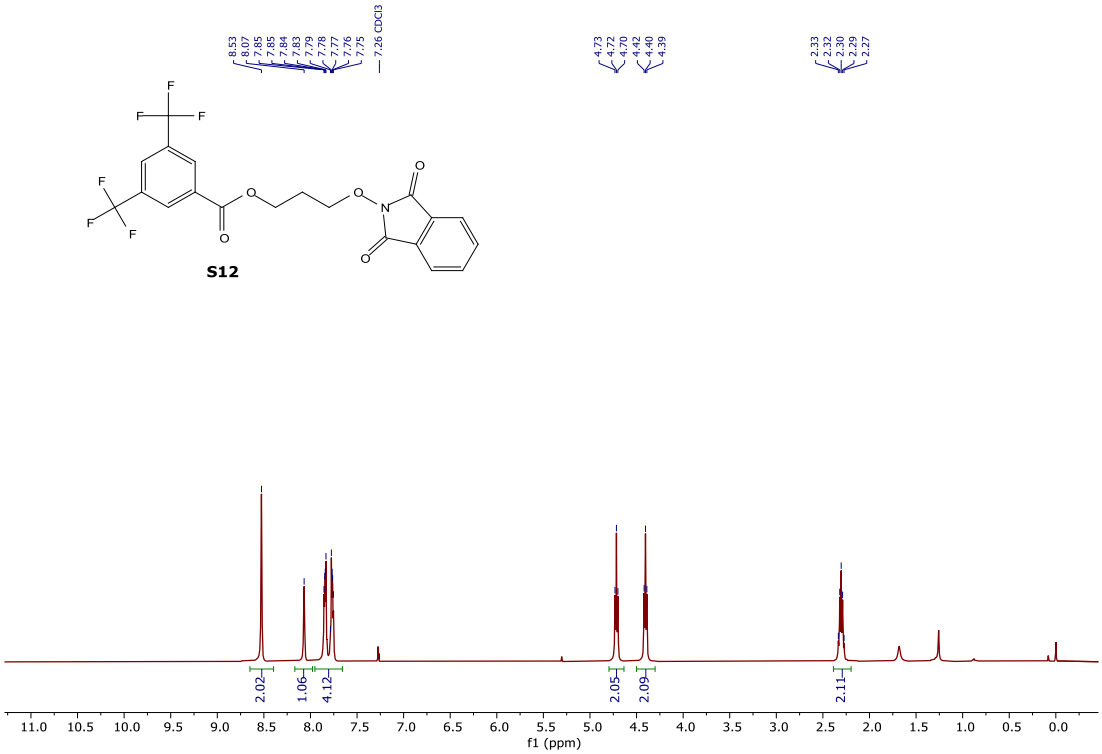


Figure S73. ^{1}H NMR spectrum in $CDCl_3$ of compound **S12**.

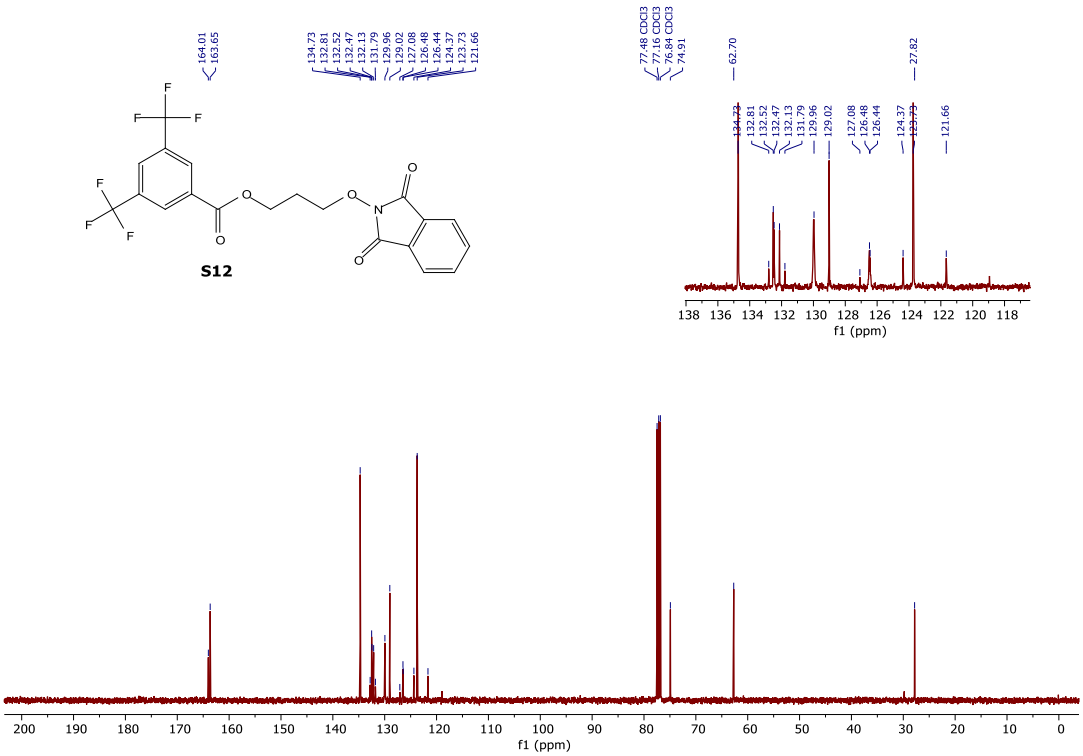


Figure S74. ^{13}C NMR spectrum in $CDCl_3$ of compound **S12**.

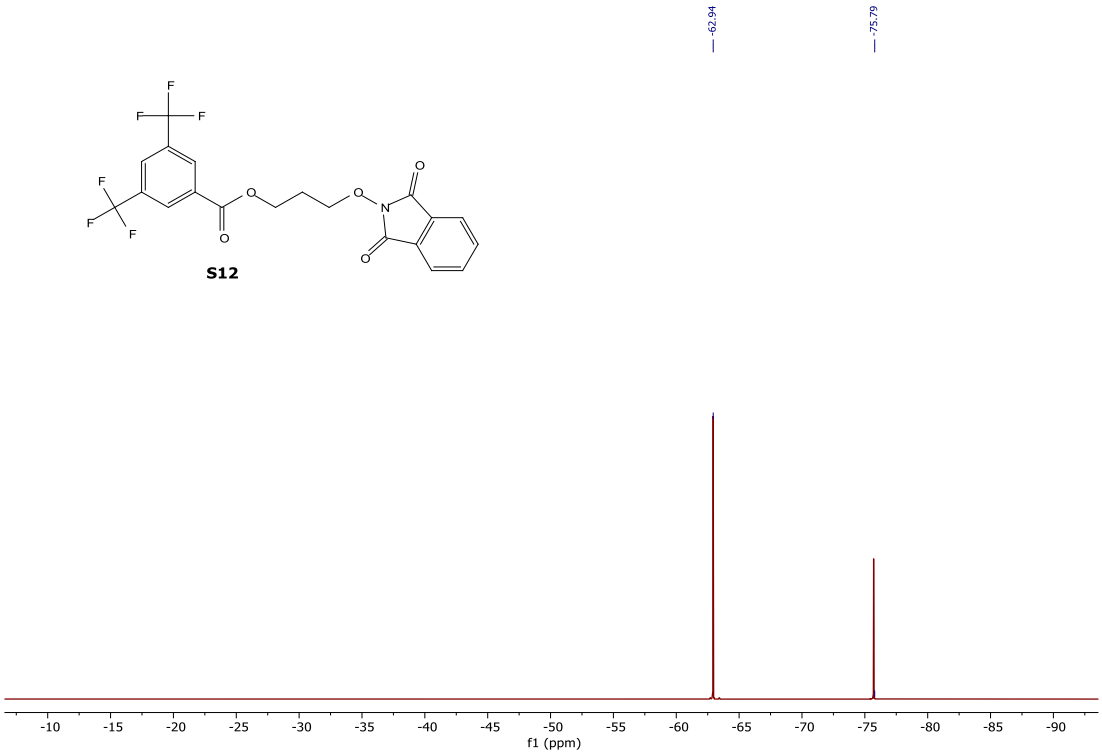


Figure S75. ^{19}F NMR spectrum in $CDCl_3$ of compound **S12**.

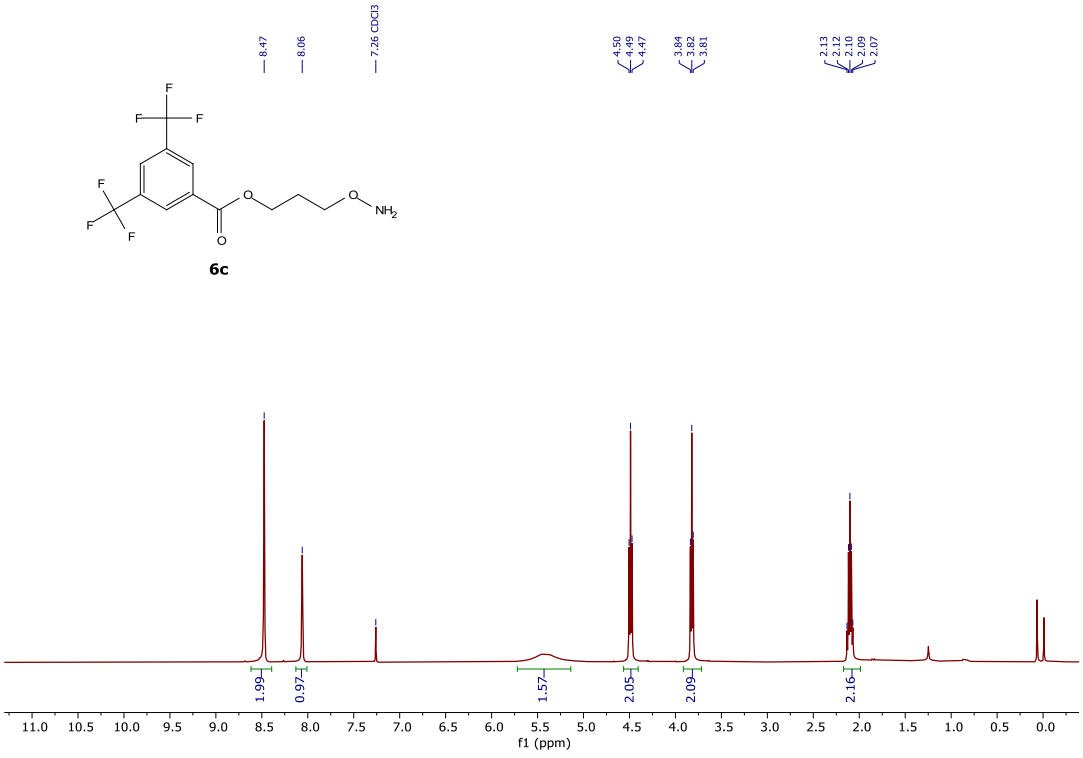


Figure S76. ^{1}H NMR spectrum in $CDCl_3$ of compound **6c**.

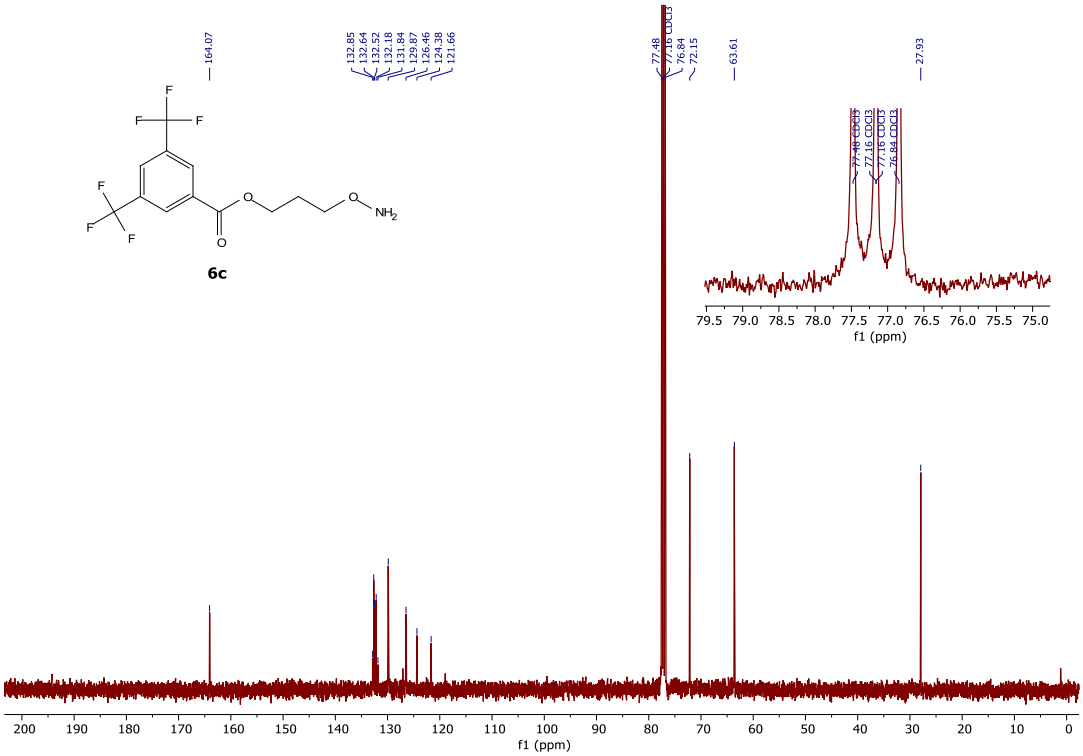


Figure S77. ^{13}C NMR spectrum in $CDCl_3$ of compound **6c**.

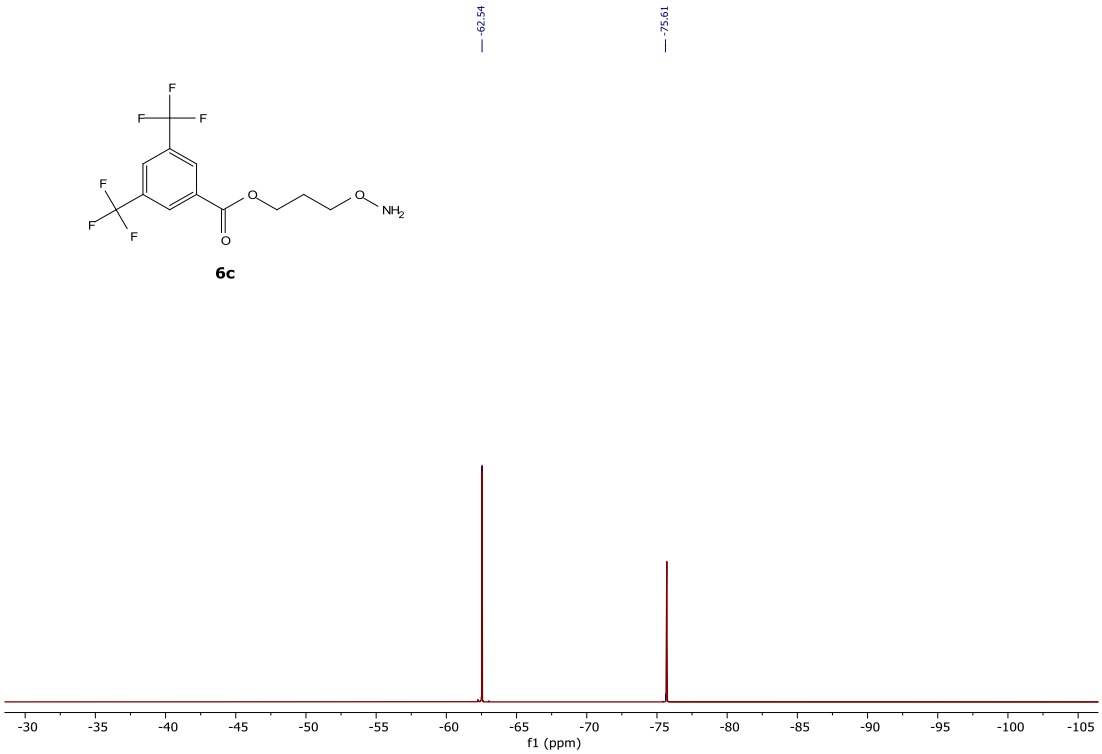


Figure S78. ^{19}F NMR spectrum in $CDCl_3$ of compound **6c**.

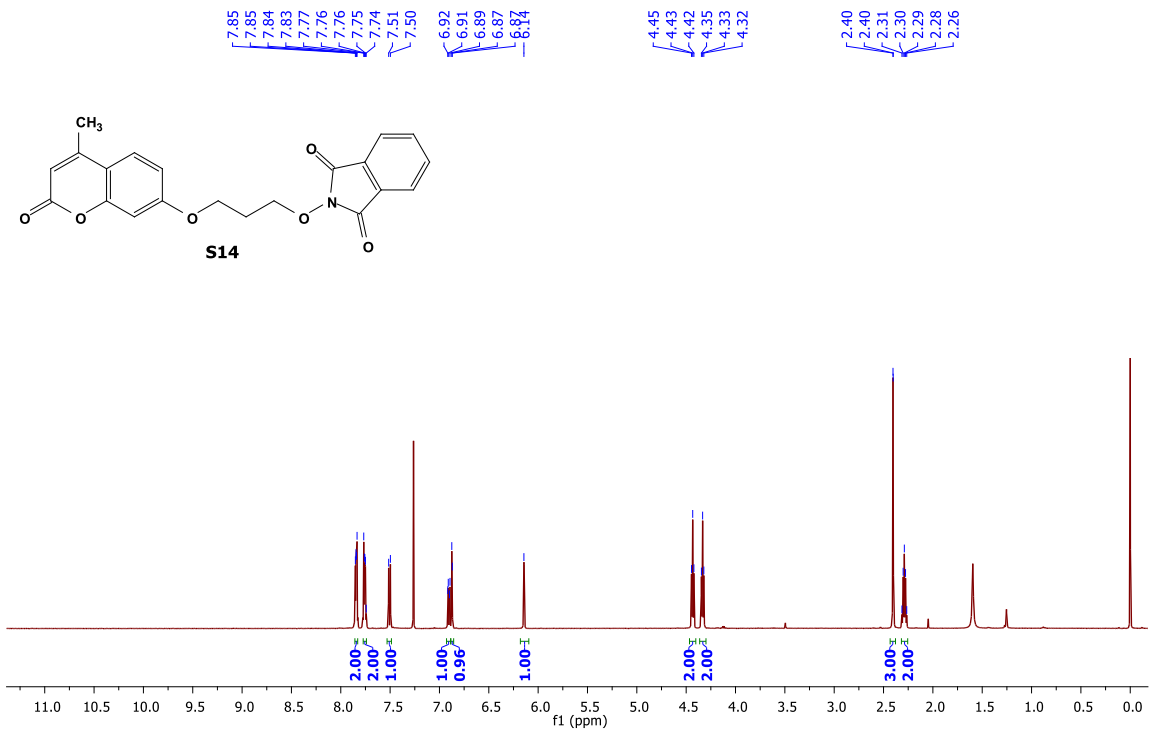


Figure S79. ^{1}H NMR spectrum in $CDCl_3$ of compound **S14.**

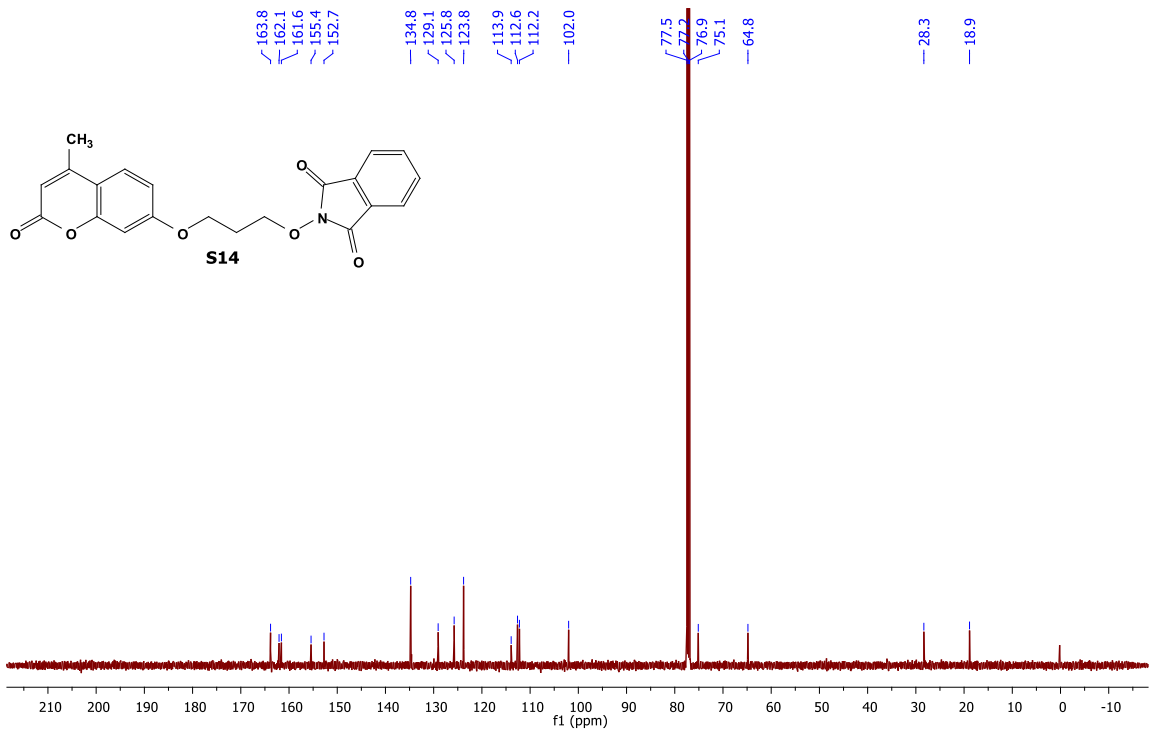


Figure S80. ^{13}C NMR spectrum in $CDCl_3$ of compound **S14**.

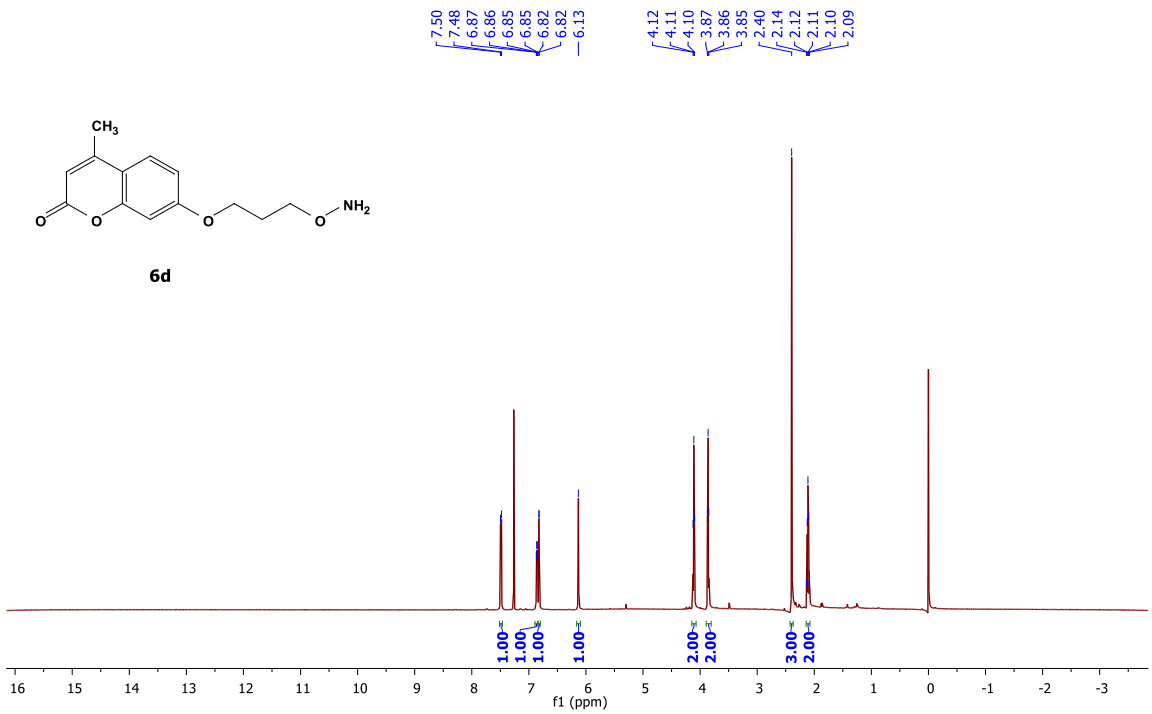


Figure S81. ^{1}H NMR spectrum in $CDCl_3$ of compound **6d**.

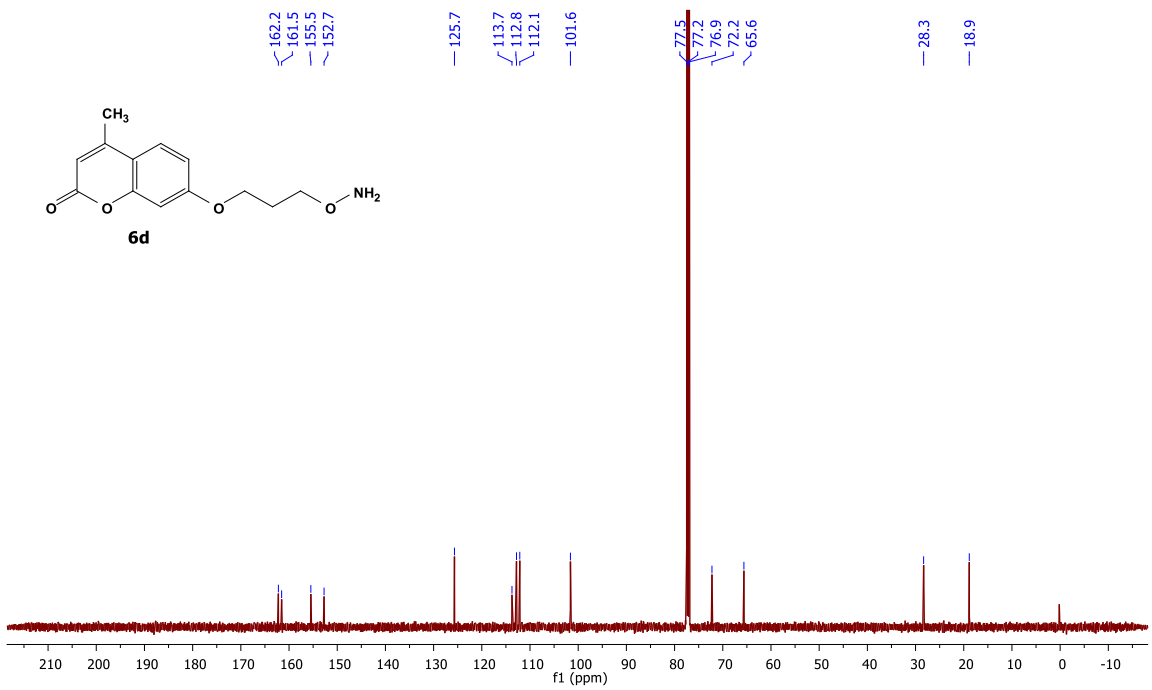


Figure S82. ^{13}C NMR spectrum in $CDCl_3$ of compound **6d**.

6. References

1. L. Purushottam, S. R. Adusumalli, U. Singh, V. B. Unnikrishnan, D. G. Rawale, M. Gujrati, R. K. Mishra and V. Rai, *Nat. Commun.*, 2019, **10**, 2539.
2. S. R. Adusumalli, D. G. Rawale, U. Singh, P. Tripathi, R. Paul, N. Kalra, R. K. Mishra, S. Shukla and V. Rai, *J. Am. Chem. Soc.*, 2018, **140**, 15114-15123.

6.1 Additional references

1. D. G. Rawale, K. Thakur, S. R. Adusumalli and V. Rai, *Eur. J. Org. Chem.*, 2019, 6749-6763.
2. A. Ramesh, K. Thakur and V. Rai, *Chem. Rec.*, 2021, **21**, 1941-1956.
3. I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nat. Chem. Biol.*, 2007, **3**, 321-322.
4. G. Mantovani, F. Lecolley, L. Tao, D. M. Haddleton, J. Clerx, J. J. L. M. Cornelissen and K. Velonia, *J. Am. Chem. Soc*., 2005, **127**, 2966-2973.
5. M. J. Matos, B. L. Oliveira, N. Martinez-Saez, A. Guerreiro, P. M. S. D. Cal, J. Bertoldo, M. Maneiro, E. Perkins, J. Howard, M. J. Deery, J. M. Chalker, F. Corzana, G. Jimenez-Oses and G. J. L. Bernardes, *J. Am. Chem. Soc.*, 2018, **140**, 4004-4017.
6. T. Sahu, M. Chilamari and V. Rai, *Chem. Commun.*, 2022, **58**, 1768-1771.
7. N. C. Reddy, M. Kumar, R. Molla and V. Rai, *Org. Biomol. Chem.*, 2020, **18**, 4669-4691.