Electronic Supplementary Material (ESI) for ChemComm.

Protein inspired chemically orthogonal imines for linchpin directed precise and modular labeling of lysine in proteins

Tularam Sahu,[‡] Maheshwerreddy Chilamari,[‡] Vishal Rai*
Department of Chemistry, 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 and Merck Novabiochem. 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) and the pH was adjusted using pH meter Mettler Toledo (FE20). The reaction mixture was either stirred (Heidolph, 500-600 rpm), or vortexed in incubator shaker Thermo Scientific MaxQ 8000 (350 rpm, 25-37 ºC). BUCHI rotavapor R-210/215 was used to remove organic solvents and CHRiST ALPHA 2-4 LD plus lyophilizer was used for lyophilization of aqueous samples. 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 2 cm path length cuvette at 25 ^{o}C. In fluorescence spectroscopy, emission measurements were 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. For reactions where chromatography was involved, flash column chromatography was carried out on Combiflash Rf 200 or gravity columns using 230-400 mesh silica gel from Merck.

Nuclear magnetic resonance spectra: ^{1}H, ^{13}C, and 19F 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 UHPLC and Agilent 6130 single quad coupled with Agilent 1200 series HPLC (ESI/APCI) were 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 UHPLC 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). Matrix assisted laser desorption/ionisation time of flight mass spectrometry was performed with Bruker Daltonics UltrafleXtreme Software-Flex control version 3.4, using sinapic acid and S3 α-cyano-4-hydroxycinnamic acid (HCCA) matrix. 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% formic acid and used as the mobile phase. Method A was used to record the LC-ESI-MS data for proteins and method B used for peptide mapping and MS/MS.

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

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.

2. Experimental methods

2.1 Procedure for protein labeling

In a 1.5 ml microcentrifuge tube, a protein (**1**, 7.3 nmol) in $NaHCO_3$ buffer (80 µl, pH 7.8, 0.1 M) was taken. LDM reagent (**4**, 73 nmol) in DMSO (6 µl) and triethylphosphite (4.3 µmol at 0 h and 12 h) in DMSO (7 µl × 2) were added from freshly prepared stock solutions. The overall concentration of protein and LDM reagent was 73 µM and 1.4 mM respectively. The reaction mixture was incubated at 37 °C. After 6-24 h, the reaction mixture was diluted with acetonitrile:water (10:90, 1500 µl). Unreacted LDM reagent and salts were removed by using Amicon® Ultra-0.5 mL 3-kDa or 10-kDa MWCO centrifugal filters spin concentrator. The protein mixture was further washed with Millipore Grade I water (5 × 0.4 ml). The sample was analyzed

by LC-ESI-MS or MALDI-ToF-MS. The sample was then buffer (80 µl, 0.1 M, pH 7.0) exchanged. To this solution, hydroxylamine **5a** (3.65 µmol) in DMSO (10 µl) from a freshly prepared stock solution was added for late-stage modification (oxime formation) and vortexed for 2 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 or MALDI-ToF-MS. The aqueous sample was concentrated by lyophilization before subjecting it to digestion, peptide mapping, and sequencing by MS-MS. Same procedure was followed for the reaction of protein with other nucleophiles and progress of bioconjugation was analyzed by LC-ESI-MS.

Procedure for late-stage orthogonal tagging of the protein

In a 1.5 ml microcentrifuge tube, cytochrome C (**1a**, 7.3 nmol) in $NaHCO_3$ buffer (80 µl, pH 7.8, 0.1 M) was taken. LDM reagent (**4**, 73 nmol) in DMSO (6 µl) and triethylphosphite (4.2 µmol at zero hours and 12 hours) in DMSO (7 µl × 2) were added from freshly prepared stock solutions. The overall concentration of protein and LDM reagent was 73 µM and 730 µM respectively. The reaction mixture was incubated at 37 ºC for 24 h. After 24 h, the reaction mixture was diluted with acetonitrile:water (10:90, 1500 µl). Unreacted LDM reagent and salts were removed by using Amicon® Ultra-0.5 mL 3-kDa or 10-kDa MWCO centrifugal filters spin concentrator. The protein mixture was further washed with Millipore Grade I water (5 × 0.4 ml). The sample was analyzed by ESI-MS or MALDI-ToF-MS. The sample was then buffer (80 µl, Phosphate buffer (0.1 M, pH 7.0)) exchanged. To the concentrated sample in phosphate buffer, hydroxylamine derivatives such as 3-(aminooxy)propyl 3,5-bis(trifluoromethyl)benzoate **5c** / 3-(aminooxy)propyl 5-((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoate **5d**/ 7-((3-(aminooxy)propyl)thio)-4-methyl-2H-chromen-2-one **5e** (3.6 µmol) in DMSO (20 µl) from a freshly prepared stock solution was added to convert mono-labeled cytochrome C to its oxime derivative. The excess of O-hydroxylamine derivative and salts were removed by the spin concentrator. The sample was analyzed by LC-ESI-MS. The salt free sample was concentrated by lyophilization before subjecting it to digestion, peptide mapping and sequencing by MS-MS.

2.2 Digestion protocol

Procedure for in-solution digestion of RNase A

All the solutions were prepared freshly before use in reactions.

Step 1. Protein (0.1 mg, 7.3 nmol) in 100 mM tris (10 µl, pH 7.8) with urea (6 M) was taken in 2 ml microcentrifuge tube and vortexed 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. **Sulfhydryl 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 alkylating reagent:** To quench the unreacted iodoacetamide, reducing agent (0.2 M, 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 [5 μg, based on ratio of chymotrypsin or trypsin/protein (1:20); α-chymotrypsin or trypsin in 1 mM HCl was dissolved in 0.4 M tris and 0.01 M $CaCl_2$] 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 used for peptide mapping by MS and sequencing by MS-MS investigations.

Procedure for in-solution digestion of cytochrome C and myoglobin

Protein (0.1 mg) in 100 mM tris (10 μl, pH 7.8) with urea (6 M) was taken in a 1.5 ml microcentrifuge tube. To this solution, tert-butanol (10 μl) was added and incubated for 1 h at 37 ^{o}C. The urea concentration of sample was reduced to 0.6 M with grade I water. To this solution, 10 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.

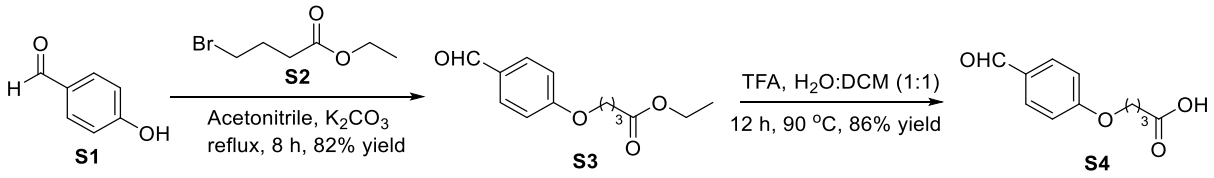
2.3 Enzymatic assay

Cytochrome C activity before and after the labeling was checked by ABTS (Diammonium 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonate) assay.[1] The oxidation of ABTS to $ABTS^{+\cdot}$ was monitored by UV-Vis absorption at 415 nm (A_{415}, quartz cuvette, path length 1 cm, 25 ^{o}C). All solutions were made using Millipore Grade I water. The freshly prepared cytochrome C (**1a**), labeled cytochrome C (**6a**), ABTS, and hydrogen peroxide (50%) solutions were used for the assay.
Conditions used for the assay in 1 ml reaction mixture:
1. ABTS (365 μM) and H_2O_2 (35 mM) in absence of cytochrome C (**1a**)
2. ABTS (365 μM), H_2O_2 (35 mM), and cytochrome C (**1a**, 2.4 μM)
3. ABTS (365 μM), H_2O_2 (35 mM), and labeled cytochrome C (**6a**, 2.4 μM)

3. Synthesis of reagents

Synthesis of precursors

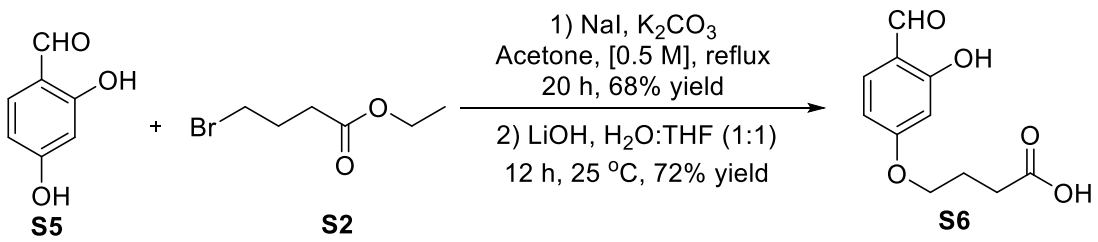
Synthesis of 4-(4-formylphenoxy)butanoate (S4)



Scheme S1. Synthesis of 4-(4-formylphenoxy)butanoate **S4**.[2]

In a 25 ml round bottom flask, *p*-hydroxybenzaldehyde (**S1**, 122.1 mg, 1 mmol) was dissolved in 10 ml acetonitrile. To this aldehyde solution, K_2CO_3 (276.4 mg, 2 mmol) and ethyl 4-bromobutanoate (**S2**, 0.17 ml, 1.2 mmol) were added and reaction mixture was allowed to reflux for 8 h. Reaction was monitored using thin layer chromatography and upon completion, the reaction mixtures was filtered to remove potassium carbonate. The solution was concentrated under vacuum and the product was purified using flash column chromatography (ethylacetate:*n*-hexane 2:98) to afford ethyl 4-(4-formylphenoxy)butanoate (**S3**) (82% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.88 (s, 1H), 7.83 (d, J = 8.6 Hz, 2H), 6.99 (d, J = 8.7 Hz, 2H), 4.15 (q, J = 7.2 Hz, 2H), 4.11 (t, J = 6.2 Hz, 2H), 2.53 (t, J = 7.2 Hz, 2H), 2.15 (p, J = 6.7 Hz, 2H), 1.26 (t, J = 7.1 Hz, 3H) ppm. ^{13}C NMR (101 MHz, $CDCl_3$) δ 190.9, 173.1, 164.0, 132.1, 130.0, 114.8, 67.2, 60.6, 30.7, 21.0, 14.3 ppm. MS (ESI) $[M+H]^+$ calcd. $C_{13}H_{16}O_4$ 237.1, found 237.1. The resulted product (**S3**) was dissolved in 10 ml water and dichloromethane mixture (1:1). To this solution trifluoroacetic acid (4 equiv.) was added and reaction mixture was transfer to 90 °C and allowed to stir for 12 h. The hydrolysis was monitored by using TLC and ESI-MS. After completion of the reaction, reaction mixture was allowed to filter and collected solid was subjected to the silica-gel flash column chromatography (ethylacetate:*n*-hexane 35:65) to afford the pure product **S4** (86% yield). ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.86 (s, 1H), 7.86 (d, J = 8.7 Hz, 2H), 7.12 (d, J = 8.7 Hz, 2H), 4.11 (t, J = 6.4 Hz, 2H), 2.39 (t, J = 7.3 Hz, 2H), 1.97 (m, 2H) ppm. ^{13}C NMR (126 MHz, DMSO-d_6) δ 191.7, 174.4, 163.9, 132.3, 130.0, 115.3, 67.6, 30.4, 21.0 ppm. LRMS (ESI) $[M+H]^+$ calcd. $C_{11}H_{12}O_4$ 209.1, found 209.3.

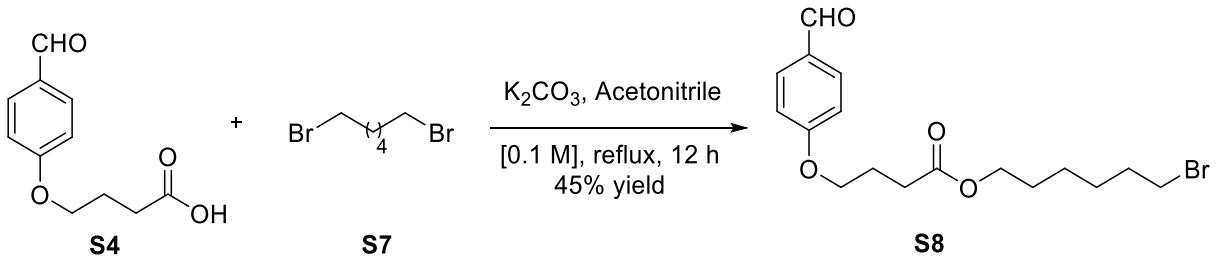
Synthesis of 4-(4-formylphenoxy)butanoate (S6)[3]



Scheme S2. Synthesis of 4-(4-formylphenoxy)butanoate **S6**.

In a 25 ml round bottom flask, 2,4-dihydroxybenzaldehyde (**S5**, 122.1 mg, 1 mmol) was dissolved in 10 ml acetonitrile. To this aldehyde solution, K_2CO_3 (276.4 mg, 2 mmol) and ethyl 4-bromobutanoate (**S2**, 0.17 ml, 1.2 mmol) were added and reaction mixture was allowed to reflux for 8 h. Reaction was monitored using thin layer chromatography and upon completion, the reaction mixtures was filtered to remove potassium carbonate. The solution was concentrated under vacuum and the product was purified using flash column chromatography (ethylacetate:*n*-hexane 2:98) to afford ethyl 4-(4-formylphenoxy)butanoate (82% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.46 (s, 1H), 9.71 (s, 1H), 7.42 (d, *J* = 8.7 Hz, 1H), 6.52 (dd, *J* = 8.7, 2.3 Hz, 1H), 6.41 (d, *J* = 2.3 Hz, 1H), 4.16 (q, *J* = 7.1 Hz, 2H), 4.07 (t, *J* = 6.2 Hz, 2H), 2.51 (t, *J* = 7.2 Hz, 2H), 2.19 – 2.09 (m, 2H), 1.27 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.3, 172.9, 166.0, 164.4, 135.2, 115.2, 108.5, 101.2, 67.3, 60.5, 30.5, 24.3, 14.2 ppm. LRMS (ESI) $[M+H]^+$ calcd. $C_{13}H_{16}O_4$ 237.1, found 237.1. In a 10 ml round bottom flask, ethyl 4-(4-formylphenoxy)butanoate (208 mg, 1 mmol) and trifluoroacetic acid (307 µl, 4 mmol) were dissolved in water (5 ml). The reaction mixture was stirred at 90 °C for 8 h. The reaction mixture turns into a clear solution while heating. The progress of the reaction was followed by TLC. The purification of crude reaction mixture was performed by silica gel flash column chromatography using ethyl acetate:n-hexane (30:70) to isolate gave **S6** (86% yield). ^{1}H NMR (500 MHz, Acetone-d_6) δ 11.51 (s, 1H), 9.83 (s, 1H), 7.71-7.64 (m, 1H), 6.66 (dt, *J* = 7.6, 3.8 Hz, 1H), 6.49 (t, *J* = 3.8 Hz, 1H), 4.20 (t, *J* = 6.4 Hz, 2H), 2.54 (t, *J* = 7.3 Hz, 2H), 2.09-2.05 (m, 2H) ppm. ^{13}C NMR (101 MHz, DMSO-d_6) δ191.1, 174.0, 165.2, 163.1, 132.3, 116.2, 107.7, 101.2, 67.2, 30.0, 24.0 ppm. LRMS (ESI) $[M+H]^+$ calcd. $C_{11}H_{13}O_5$ 225.1, found 225.1.

Synthesis of 6-bromohexyl 4-(4-formylphenoxy)butanoate (S8)

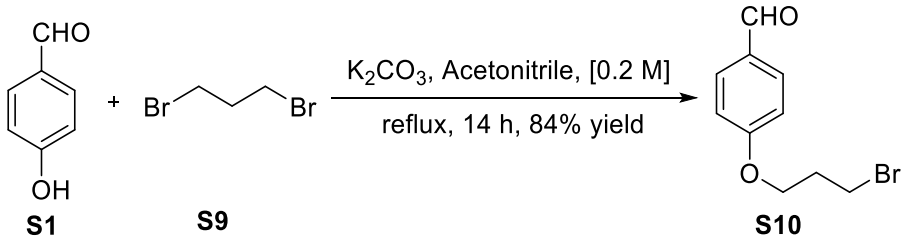


Scheme S3. Synthesis of 6-bromohexyl 4-(4-formylphenoxy)butanoate **S8**.

In a 25 ml round bottom flask, acid derivative (**S4**, 208 mg, 1 mmol) was dissolved in 10 ml acetonitrile. To this solution DBU (209 µl, 1.5 mmol) and 1,6-dibromohexane (**S7**, 308 µl, 2 mmol) were added at room temperature and the reaction mixture was allowed to reflux for 5 h. Reaction was monitored using thin layer chromatography and upon completion, the solution was concentrated under vacuum and the product was purified using flash column chromatography (ethylacetate:*n*-hexane 20:90) to afford pure compound **S8** (45% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.88 (s, 1H), 7.83 (d, *J* = 8.7 Hz, 2H), 6.99 (d, *J* = 8.7 Hz, 2H), 4.13-4.05 (m, 4H), 3.39 (t, *J* = 6.8 Hz, 2H), 2.53 (t, *J* = 7.2 Hz, 2H), 2.15 (m, 2H), 1.85 (m, 2H), 1.64 (m, 2H), 1.46 (m, 2H), 1.37 (m, *J* = 14.6, 7.3 Hz, 2H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 190.9, 173.1, 163.9,

132.1, 130.1, 114.8, 67.2, 64.5, 33.7, 32.6, 30.7, 28.5, 27.8, 25.2, 24.5 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{17}H_{23}BrO_4Na$ 393.0672, found 393.0684.

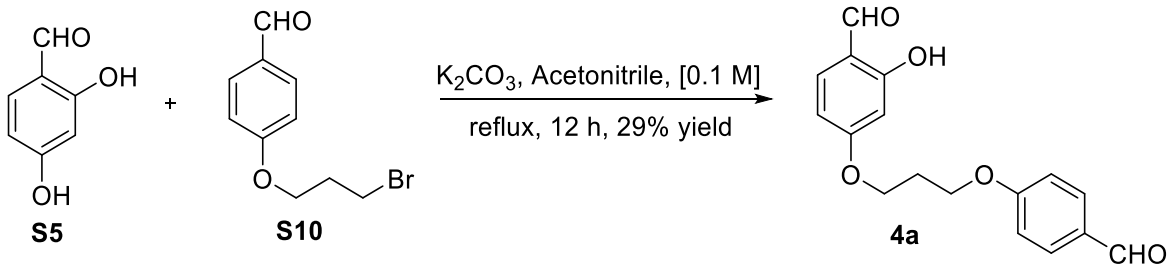
Synthesis of 4-(3-bromopropoxy)benzaldehyde (S10)



Scheme S4. Synthesis of 4-(3-bromopropoxy)benzaldehyde **S10**.

In a 25 ml round bottom flask, 4-hydroxybenzaldehyde (**S1**, 1.22 gm, 10 mmol) was dissolved in 50 ml acetonitrile. To this solution K_2CO_3 (2.76 mg, 20 mmol) and 1,3-dibromopropane (**S9**, 1.52 ml, 15 mmol) were added at room temperature and the reaction mixture was allowed to reflux for 14 h. Reaction was monitored using thin layer chromatography and upon completion, the solution was concentrated under vacuum and the product was purified using silica gel flash column chromatography (ethylacetate:*n*-hexane 2:98) to afford pure compound **S10** (84% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.89 (s, 1H), 7.84 (m, 2H), 7.01 (m, 2H), 4.20 (t, *J* = 5.8 Hz, 2H), 3.61 (t, *J* = 6.4 Hz, 2H), 2.36 (m, 2H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 190.7, 163.7, 132.0, 130.1, 114.7, 65.6, 32.0, 29.5 ppm. LRMS (ESI) $[M+H]^+$ calcd. $C_{10}H_{11}BrO_2$ 241.9, found 241.9.

Synthesis of 4-(3-(4-formylphenoxy)propoxy)-2-hydroxybenzaldehyde (4a)

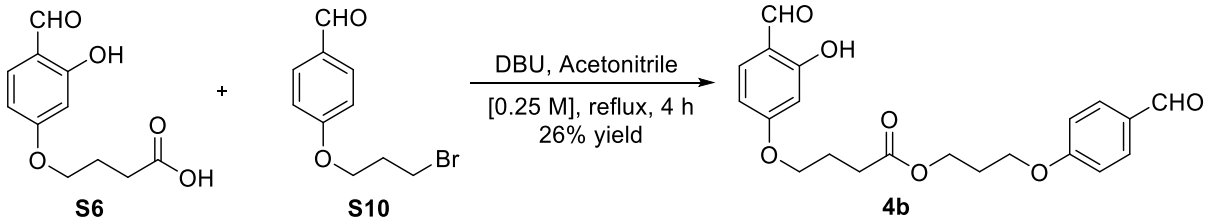


Scheme S5. Synthesis of 4-(3-(4-formylphenoxy)propoxy)-2-hydroxybenzaldehyde **4a**.

In a 25 ml round bottom flask, 2,4-dihydroxybenzaldehyde (**S5**, 138 mg, 1 mmol) was dissolved in 5 ml acetone. To this solution K_2CO_3 (1.38 mg, 1 mmol) and 4-(3-bromopropoxy)benzaldehyde (**S10**, 243 mg, 1 mmol) were added at room temperature and the reaction mixture was allowed to reflux for 24 h. Reaction was monitored using thin layer chromatography and upon completion, the solution was concentrated under vacuum and the product was purified using silica gel flash column chromatography (ethylacetate:*n*-hexane 10:90) to afford pure compound **4a** (29% yield).^{1}H NMR (500 MHz, $CDCl_3$) δ 11.46 (s, 1H), 9.88 (s, 1H), 9.71 (d, *J* = 9.8 Hz, 1H), 7.83 (d, *J* = 8.7 Hz, 2H), 7.42 (t, *J* = 9.8 Hz, 1H), 7.01 (d, *J* = 8.6 Hz, 2H), 6.54 (dd, *J* = 8.7, 2.2 Hz, 1H), 6.44 (d, *J* = 2.0 Hz, 1H), 4.23 (dd, *J* = 13.0, 6.2 Hz, 4H), 2.32 (p, *J* = 6.0 Hz, 2H) ppm. ^{13}C NMR

(126 MHz, $CDCl_3$) δ 194.4, 190.7, 165.9, 164.4, 163.7, 135.3, 132.0, 130.1, 115.3, 114.7, 108.5, 101.2, 64.6, 64.4, 28.8 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{17}H_{16}O_5Na$ 323.0890, found 323.0889.

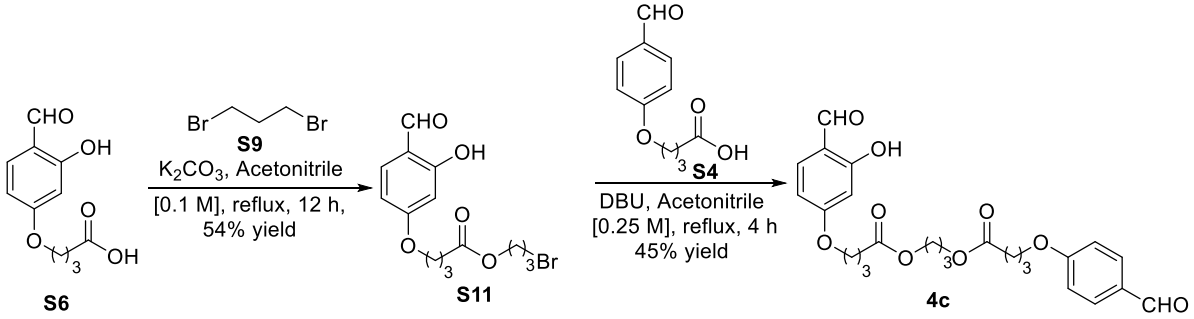
Synthesis of 3-(4-formylphenoxy)propyl 4-(4-formyl-3-hydroxyphenoxy)butanoate (4b)



Scheme S6. Synthesis of 3-(4-formylphenoxy)propyl 4-(4-formyl-3-hydroxyphenoxy) butanoate **4b**.

In a 25 ml round bottom flask, 4-(4-formyl-3-hydroxyphenoxy)butanoic acid (**S6**, 112 mg, 0.5 mmol) was dissolved in 2 ml acetonitrile. To this solution DBU (105 µl, 0.75 mmol) and 4-(3-bromopropoxy)benzaldehyde (**S10**, 121.5 mg, 0.5 mmol) were added at room temperature and the reaction mixture was allowed to reflux for 4 h. Reaction was monitored using thin layer chromatography and upon completion, the solution was concentrated under vacuum and the product was purified using silica gel flash column chromatography (ethylacetate:*n*-hexane 30:70) to afford pure compound **4b** (26% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.45 (s, 1H), 9.87 (s, 1H), 9.69 (s, 1H), 7.81 (m, 2H), 7.40 (d, *J* = 8.7 Hz, 1H), 6.97 (m, 2H), 6.50 (dd, *J* = 8.7, 2.3 Hz, 1H), 6.38 (d, *J* = 2.3 Hz, 1H), 4.31 (t, *J* = 6.3 Hz, 2H), 4.12 (t, *J* = 6.1 Hz, 2H), 4.05 (t, *J* = 6.1 Hz, 2H), 2.53 (t, *J* = 7.2 Hz, 2H), 2.18-2.10 (m, 4H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.3, 190.7, 172.8, 165.9, 164.4, 163.7, 135.3, 131.9, 130.0, 115.2, 114.7, 108.5, 101.1, 67.2, 64.7, 61.1, 30.5, 28.4, 24.2 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{21}H_{22}O_7Na$ 409.1258, found 409.1263.

Synthesis of 3-((4-(4-formyl-3-hydroxyphenoxy)butanoyl)oxy)propyl-4-(4-formylphenoxy) butanoate (4c)

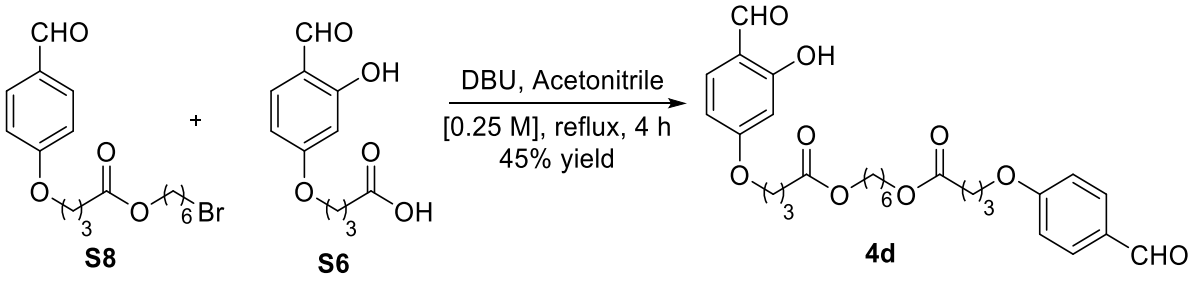


Scheme S7. Synthesis of 3-((4-(4-formyl-3-hydroxyphenoxy)butanoyl)oxy)propyl-4-(4-formylphenoxy)butanoate **4c**.

In a 25 ml round bottom flask, 4-(4-formyl-3-hydroxyphenoxy)butanoic acid (**S6**, 224 mg, 1 mmol) was dissolved in 10 ml acetonitrile. To this solution DBU (168 µl, 1.2 mmol) and 1,3-dibromopropane (**S9**, 112 µl, 1.1 mmol) were added at room temperature and the reaction mixture

was allowed to reflux for 14 h. Reaction was monitored using thin layer chromatography and upon completion, the solution was concentrated under vacuum and the product was purified using silica gel flash column chromatography (ethylacetate:*n*-hexane 15:85) to afford pure compound **S11** (54% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.45 (s, 1H), 9.71 (s, 1H), 7.42 (d, *J* = 8.7 Hz, 1H), 6.52 (dd, *J* = 8.7, 2.3 Hz, 1H), 6.41 (d, *J* = 2.3 Hz, 1H), 4.21 (t, *J* = 6.1 Hz, 2H), 4.07 (t, *J* = 6.1 Hz, 2H), 3.45 (t, *J* = 6.5 Hz, 2H), 2.53 (t, *J* = 7.2 Hz, 2H), 2.22-210 (m, 4H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.3, 172.7, 165.9, 164.4, 135.3, 115.2, 108.5, 101.2, 67.2, 62.3, 31.6, 30.4, 29.3, 24.2 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{14}H_{17}BrO_5Na$ 367.0152, found 367.0134. In a 25 ml round bottom flask, 4-(4-formyl-3-hydroxyphenoxy)butanoic acid (**S11**, 35 mg, 0.17 mmol) was dissolved in 2 ml acetonitrile. To this solution DBU (29 µl, 0.21 mmol) and 3-bromopropyl 4-(4-formyl-3-hydroxyphenoxy)butanoate (**S4**, 67 mg, 0.17 mmol) were added at room temperature and the reaction mixture was allowed to reflux for 4 h. Reaction was monitored using thin layer chromatography and upon completion, the solution was concentrated under vacuum and the product was purified using silica gel flash column chromatography (ethylacetate:*n*-hexane 20:80) to afford pure compound **4c** (45% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.45 (s, 1H), 9.87 (s, 1H), 9.70 (s, 1H), 7.82 (m, 2H), 7.41 (d, *J* = 8.7 Hz, 1H), 6.98 (m, 2H), 6.51 (dd, *J* = 8.7, 2.3 Hz, 1H), 6.40 (d, *J* = 2.3 Hz, 1H), 4.17 (t, *J* = 6.2 Hz, 4H), 4.09 (t, *J* = 6.1 Hz, 2H), 4.05 (t, *J* = 6.1 Hz, 2H), 2.55-2.50 (m, 4H), 2.16-2.09 (m, 4H), 1.97 (p, *J* = 6.3 Hz, 2H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.4, 190.8, 172.9, 172.8, 166.0, 164.5, 163.8, 135.3, 132.0, 130.0, 115.2, 114.7, 108.6, 101.2, 67.2, 67.1, 61.1, 61.1, 30.5, 30.5, 28.0, 24.4, 24.3 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{25}H_{28}O_9$ 473.1806, found 473.1807.

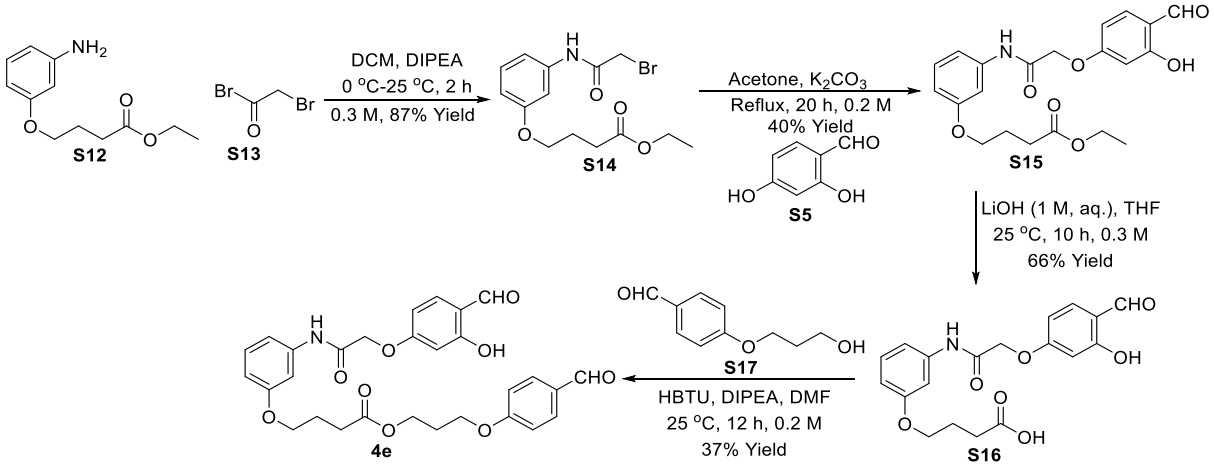
Synthesis of 6-((4-(4-formyl-3-hydroxyphenoxy)butanoyl)oxy)hexyl-4-(4-formylphenoxy) butanoate (4d)



Scheme S8. Synthesis of 6-((4-(4-formyl-3-hydroxyphenoxy)butanoyl)oxy)hexyl 4-(4-formylphenoxy)butanoate **4d**.

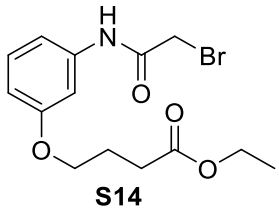
In a 10 ml round bottom flask, 4-(4-formyl-3-hydroxyphenoxy)butanoic acid (**S6**, 96 mg, 0.43 mmol) was dissolved in 3 ml acetonitrile. To this solution DBU (90 µl, 0.65 mmol) and 6-bromohexyl 4-(4-formylphenoxy)butanoate (**S8**, 160 mg, 0.27 mmol) were added at room temperature and the reaction mixture was allowed to reflux for 4 h. Reaction was monitored using thin layer chromatography and upon completion, the solution was concentrated under vacuum and the product was purified using silica gel flash column chromatography (ethylacetate:*n*-hexane

25:75) to afford pure compound **4d** (26% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.46 (s, 1H), 9.88 (s, 1H), 9.71 (s, 1H), 7.83 (m, 2H), 7.42 (d, *J* = 8.7 Hz, 1H), 6.99 (m,, 2H), 6.52 (dd, *J* = 8.7,2.3 Hz, 1H), 6.41 (d, *J* = 2.3 Hz, 1H), 4.11-4.05 (m, 8H), 2.52 (m, 4H), 2.17-2.10 (m, 4H), 1.65-1.60 (m, 4H), 1.38-1.35 (m, 4H) ppm. 13C NMR (126 MHz, $CDCl_3$) δ 194.34, 190.8, 173.0, 173.0, 166.0, 164.5, 163.9, 135.3, 132.0, 130.0, 115.2, 114.7, 108.6, 101.2, 67.3, 67.1, 64.5, 30.6, 30.6, 29.7, 28.5, 25.6, 24.4, 24.3 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{28}H_{34}O_9$ 537.2095, found 537.2097.



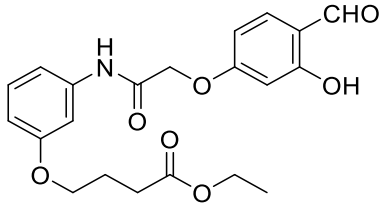
Scheme S9. 3-(4-formylphenoxy)propyl-4-(3-(2-(4-formyl-3-hydroxyphenoxy)acetamido) phenoxy)butanoate **4e**.[3]

Synthesis of ethyl 4-(3-(2-bromoacetamido)phenoxy)butanoate (S14)



In a 50 ml round bottom flask, ethyl 4-(3-aminophenoxy)butanoate **S12** (300 mg, 1.34 mmol) and DIPEA (494 µl, 2.68 mmol) in DCM (4 ml) were stirred at 0 °C for 5 minutes. To this solution, 2-bromoacetyl bromide **S13** (236 µl, 2 mmol) was added slowly over 15 minutes. The reaction mixture was brought to room temperature and stirred for 2 h. The progress of the reaction was followed by TLC. The reaction mixture was concentrated using rotary evaporator. Purification of crude mixture by silica gel column chromatography using ethyl acetate:n-hexane (10:90) gave ethyl 4-(3-(2-bromoacetamido)phenoxy)butanoate **S14** (87% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.10 (s, 1H), 7.29-7.18 (m, 2H), 7.00 (dd, *J* = 8.0, 1.4 Hz, 1H), 6.70 (dd, *J* = 8.3, 2.2 Hz, 1H), 4.15 (q, *J* = 7.1 Hz, 2H), 4.05-3.96 (m, 4H), 2.51 (t, *J* = 7.3 Hz, 2H), 2.15-2.05 (m, 2H), 1.26 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 173.2, 163.2, 159.4, 138.0, 129.8, 112.1, 111.4, 106.3, 66.8, 60.4, 30.7, 29.5, 24.6, 14.2 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{14}H_{19}BrNO_4$ 344.0497, found 344.0479.

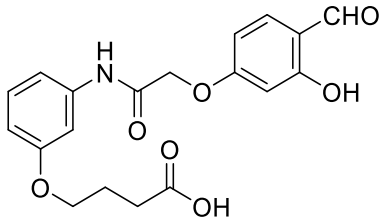
Synthesis of ethyl 4-(3-(2-(4-formyl-3-hydroxyphenoxy)acetamido)phenoxy)butanoate (S15)



S15

In a 25 ml round bottom flask, ethyl 4-(3-(2-bromoacetamido)phenoxy)butanoate **S14** (344 mg, 1 mmol), potassium carbonate (138 mg, 1 mmol) and 2,4-dihydroxybenzaldehyde (**S5**, 138 mg, 1 mmol) were dissolved in acetone (5 ml) to reflux for 20 h. The progress of the reaction was followed by TLC. The reaction mixture was concentrated using rotary evaporator. Purification of crude mixture by silica gel column chromatography using ethyl acetate:n-hexane (10:90) gave ethyl 4-(3-(2-(4-formyl-3-hydroxyphenoxy)acetamido) phenoxy)butanoate **S15** (40% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.43 (s, 1H), 9.79 (s, 1H), 8.08 (s, 1H), 7.54 (d, *J* = 8.6 Hz, 1H), 7.30 (t, *J* = 2.2 Hz, 1H), 7.24 (t, *J* = 8.2 Hz, 1H), 7.05 (dd, *J* = 7.9, 1.5 Hz, 1H), 6.70 (dd, *J* = 8.2, 2.3 Hz, 1H), 6.66 (dd, *J* = 8.6, 2.4 Hz, 1H), 6.54 (d, *J* = 2.3 Hz, 1H), 4.65 (s, 2H), 4.15 (q, *J* = 7.1 Hz, 2H), 4.02 (t, *J* = 6.1 Hz, 2H), 2.51 (t, *J* = 7.3 Hz, 2H), 2.15-2.07 (m, 2H), 1.26 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.6, 173.2, 164.7, 164.3, 163.4, 159.5, 137.7, 135.8, 129.8, 116.4, 112.3, 111.3, 107.8, 106.6, 102.4, 67.4, 66.8, 60.4, 30.7, 21.0, 14.2 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{21}H_{24}NO_7$ 402.1553, found 402.1530.

4-(3-(2-(4-formyl-3-hydroxyphenoxy)acetamido)phenoxy)butanoic acid (S16)

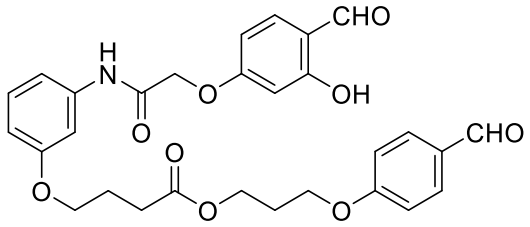


S16

In a 25 ml round bottom flask, ethyl 4-(3-(2-(4-formyl-3-hydroxyphenoxy)acetamido) phenoxy)butanoate **S15** (401 mg, 1 mmol) and 1 N LiOH (aq., 3 ml) were dissolved in THF (0.3 ml) and stirred at room temperature. The progress of the reaction was followed by thin-layer chromatography. After 8 h, the reaction mixture was quenched with 4 N HCl (aq.) and compounds were extracted with ethyl acetate. The collected organic fractions were dried over anhydrous sodium sulfate and concentrated in vacuo. The purification of crude reaction mixture was performed by silica gel flash column chromatography using ethyl acetate:n-hexane (70:30) to isolate 4-(4-(2-(4-formyl-3-hydroxyphenoxy)acetamido)phenoxy)butanoic acid **S16** (66% yield). ^{1}H NMR (500 MHz, DMSO-d_6) δ 12.11 (s, 1H), 11.01 (s, 1H), 10.12 (s, 1H), 7.64 (d, *J* = 8.7 Hz, 1H), 7.31 (t, *J* = 2.1 Hz, 1H), 7.21 (t, *J* = 8.1 Hz, 1H), 7.16 (d, *J* = 8.6 Hz, 1H), 6.66 (ddd, *J* = 8.1, 2.4, 0.7 Hz, 1H), 6.63 (dd, *J* = 8.7, 2.3 Hz, 1H), 6.52 (d, *J* = 2.3 Hz, 1H), 4.77 (s, 2H), 3.94 (t, *J* = 6.4 Hz, 2H), 2.37 (t, *J* = 7.3 Hz, 2H), 1.99-1.87 (m, 2H) ppm. ^{13}C NMR (126 MHz, DMSO-d_6) δ

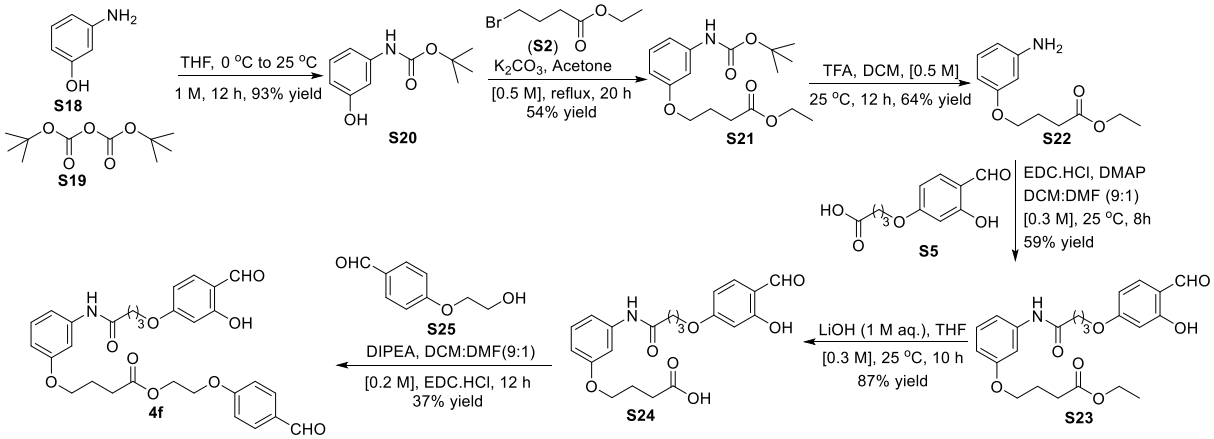
190.8, 174.1, 165.8, 164.4, 162.8, 158.8, 139.5, 131.9, 129.6, 116.7, 111.9, 109.8, 107.8, 105.9, 101.7, 67.1, 66.5, 30.1, 24.2 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{19}H_{20}NO_7$ 374.1240, found 374.1253.

Synthesis of 3-(4-formylphenoxy)propyl 4-(3-(2-(4-formyl-3-hydroxyphenoxy) acetamido)phenoxy)butanoate (4e)



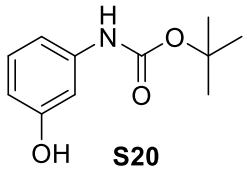
4e

In a 25 ml round bottom flask, 4-(4-(2-(4-formyl-3-hydroxyphenoxy)acetamido) phenoxy)butanoic acid **S16** (100 mg, 0.25 mmol), EDC.HCl (81.7 µl, 0.375 mmol) and DIPEA (62.5 µl, 0.375 mmol) were dissolved in dry DCM:DMF (9:1, 6.6 ml). To this solution, 4-(2-hydroxyethoxy)benzaldehyde **S17** (49.6 mg, 0.3 mmol) was added and stirred at 25 ^oC. The progress of the reaction was followed by thin-layer chromatography. After 8 h, the reaction mixture was quenched with 1N HCl (aq.) and crude was extracted with DCM. The collected organic fractions were dried over anhydrous sodium sulphate and concentrated in vacuo. Purification of crude mixture by flash chromatography (MeOH:DCM, 1:99) gave **4e** (37% yield). ^1H NMR (500 MHz, $CDCl_3$) δ 11.43 (s, 1H), 9.87 (s, 1H), 9.78 (s, 1H), 8.10 (s, 1H), 7.81 (m, 2H), 7.53 (d, *J* = 8.6 Hz, 1H), 7.33 (m, 1H), 7.22 (t, *J* = 8.1 Hz, 1H), 7.00-6.96 (m, 3H), 6.69-6.64 (m, 2H), 6.54 (d, *J* = 2.3 Hz, 1H), 4.64 (s, 2H), 4.30 (t, *J* = 6.2 Hz, 2H), 4.12 (t, *J* = 6.1, 2H), 4.01 (t, *J* = 6.0 Hz, 2H), 2.54 (t, *J* = 7.3 Hz, 2H), 2.18-2.08 (m, 4H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.7, 190.8, 173.1, 164.8, 164.3, 163.8, 163.5, 159.5, 137.8, 135.8, 132.0, 130.0, 129.9, 116.4, 114.7, 112.4, 111.2, 107.84, 106.7, 102.4, 67.4, 66.7, 64.8, 61.1, 30.7, 28.5, 24.6 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{29}H_{29}NO_9$ 536.1915, found 536.1942.



Scheme S10. **S**ynthesis of 2-(4-formylphenoxy)ethyl-4-(3-(4-(4-formyl-3-hydroxyphenoxy) butanamido)phenoxy)butanoate **4f**.[3]

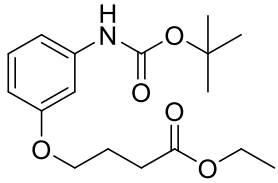
Synthesis of tert-butyl (3-hydroxyphenyl)carbamate (S20)



S20

In a dry 100 ml round bottom flask, 3-aminophenol **S18** (2.182 g, 20 mmol) was dissolved in THF (10 ml). To this solution, di-tert-butyl dicarbonate **S19** (5.5 ml, 24 mmol) in THF (10 ml) was added slowly by using dropping funnel at 0 ºC. After 2 h, the reaction mixture was brought to room temperature and stirred for 10 h. The reaction mixture was concentrated in vacuo and washing of the crude solid by hexane gave analytically pure product tert-butyl (3-hydroxyphenyl)carbamate **S20** (93% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.12 (t, *J* = 8.1 Hz, 1H), 6.72 (dd, *J* = 8.0, 1.2 Hz, 1H), 6.53 (dd, *J* = 8.1, 1.9 Hz, 1H), 6.49 (s, 1H), 5.49 (s, 1H), 1.51 (s, 9H) ppm. ^{13}C NMR (101 MHz, $CDCl_3$) δ 156.4, 152.8, 139.4, 129.9, 110.6, 110.1, 105.8, 80.8, 28.3 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{11}H_{15}NNaO_3$ 232.0950, found 232.0953.

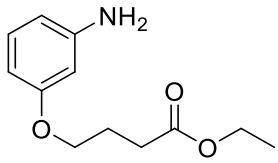
Synthesis of ethyl 4-(3-((tert-butoxycarbonyl)amino)phenoxy)butanoate (S21)



S21

In a 100 ml round bottom flask, tert-butyl (3-hydroxyphenyl)carbamate **S20** (2 g, 9.6 mmol), ethyl 4-bromobutanoate **S2** (1.24 ml, 8.7 mmol) and K_2CO_3 (1.987 g, 11 mmol) were dissolved in acetone (20 ml) and refluxed for 20 h. The reaction mixture was concentrated in vacuo and portioned between ethyl acetate and water. The organic layer was separated, dried over anhydrous sodium sulphate and then filtered. The filtrate was concentrated under reduced pressure. The crude compound was purified by silica gel column chromatography using ethyl acetate:n-hexane (4:96) to afford ethyl 4-(3-((tert-butoxycarbonyl) amino)phenoxy)butanoate **S21** (54% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.15 (t, *J* = 8.1 Hz, 1H), 7.11 (s, 1H), 6.80 (dd, *J* = 8.0, 1.4 Hz, 1H), 6.56 (dd, *J* = 8.1, 2.2 Hz, 1H), 6.46 (s, 1H), 4.14 (q, *J* = 7.1 Hz, 2H), 3.99 (t, *J* = 6.1 Hz, 2H), 2.50 (t, *J* = 7.4 Hz, 2H), 2.13-2.05 (m, 2H), 1.51 (s, 9H), 1.26 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ: 173.2, 159.5, 152.6, 139.6, 129.6, 110.7, 109.2, 104.8, 80.5, 66.7, 60.4, 30.8, 28.3, 24.6, 14.2 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{17}H_{25}NNaO_5$ 346.1630, found 346.1653.

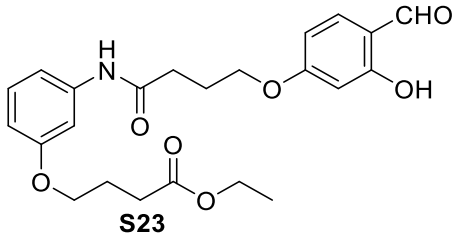
Synthesis of ethyl 4-(3-aminophenoxy)butanoate (S22)



S22

In a 100 ml round bottom flask, ethyl 4-(3-((tert-butoxycarbonyl)amino)phenoxy)butanoate **S21** (1.617 g, 5 mmol) was dissolved in CH_2Cl_2 (10 ml). To this solution, trifluoroacetic acid (1.530 ml, 20 mmol) was added slowly over 15 minutes and stirred at 25°C for 12 h. The reaction mixture was quenched with sodium bicarbonate (slow addition), added CH_2Cl_2 and then separated the organic layer. The organic layer was dried with anhydrous sodium sulphate, filtered and concentrated in vacuo to give ethyl 4-(3-aminophenoxy)butanoate **S22** (64% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.04 (t, *J* = 8.0 Hz, 1H), 6.33-6.26 (m, 2H), 6.23 (t, *J* = 2.2 Hz, 1H), 4.14 (q, *J* = 7.1 Hz, 2H), 3.96 (t, *J* = 6.1 Hz, 2H), 3.63 (bs, 2H), 2.50 (t, *J* = 7.3 Hz, 2H), 2.13-2.02 (m, 2H), 1.26 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 173.3, 160.0, 147.7, 130.1, 107.9, 101.0, 101.6, 66.5, 60.4, 30.8, 24.6, 14.2 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{12}H_{18}NO_3$ 224.1287, found 224.1287.

Synthesis of ethyl 4-(3-(4-(4-formyl-3-hydroxyphenoxy)butanamido)phenoxy) butanoate (S23)

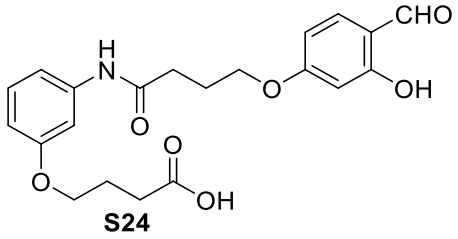


S23

In a 25 ml round bottom flask, 4-(4-formyl-3-hydroxyphenoxy)butanoic acid **S5** (448 mg, 2 mmol), 3-(((ethylimino)methylene)amino)-N,N-dimethylpropan-1-amine (465 mg, 3 mmol) and N,N-Dimethyl 4-amino pyridine (244 mg, 2 mmol) were dissolved in dry DCM:DMF (9:1, 6.6 ml). To this solution, ethyl 4-(3-aminophenoxy)butanoate **S22** (446 mg, 2 mmol) was added and stirred at 25 °C. The progress of the reaction was followed by thin-layer chromatography. After 8 h, the reaction mixture was quenched with 1N HCl (aq.) and crude was extracted with ethyl acetate. The collected organic fractions were dried over anhydrous sodium sulphate and concentrated in vacuo. Purification of crude mixture by flash chromatography (ethyl acetate:n-hexane, 30:70) gave ethyl 4-(3-(4-(4-formyl-3-hydroxyphenoxy)butanamido)phenoxy) butanoate **S23** (59 % yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 11.44 (s, 1H), 9.71 (s, 1H), 7.42 (d, *J* = 8.7, Hz, 1H), 7.32 (bs, 1H), 7.28 (bs, 1H), 7.22-7.15 (m, 1H), 6.97-6.89 (m, 1H), 6.64 (dd, *J* = 8.2, 1.7 Hz, 1H), 6.53 (dd, *J* = 8.6, 2.3 Hz, 1H), 6.43 (d, *J* = 2.2 Hz, 1H), 4.17-4.09 (m, 4H), 4.00 (t, *J* = 6.1 Hz, 2H), 2.56 (t, *J* = 7.1 Hz, 2H), 2.50 (t, *J* = 7.3 Hz, 2H), 2.28-2.19 (m, 2H), 2.14-2.05 (m, 2H), 1.26 (t, *J* = 7.1, 3.2 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ: 191, 173.4, 170.3, 166.1, 164.6, 159.6, 139.1, 135.5,

129 .8, 115.4, 111.9, 110.8, 108.5, 106.3, 101.5, 67.5, 66.9, 60.6, 38.8, 33.8, 30.9, 24.8, 14.4 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{23}H_{27}NNaO_7$ 452.1685, found 452.1707.

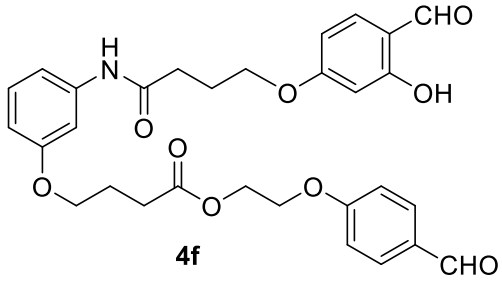
Synthesis of 4-(3-(4-(4-formyl-3-hydroxyphenoxy)butanamido)phenoxy)butanoicacid (S25)



S24

In 25 ml round bottom flask, ethyl 4-(3-(4-(4-formyl-3-hydroxyphenoxy)butanamido)phenoxy) butanoate **S23** (207 mg, 0.48 mmol) and lithium hydroxide (316 mg, 1.44 mmol) were dissolved in THF (3.3 ml) and stirred at room temperature. The progress of the reaction was followed by thin-layer chromatography. After 10 h, reaction mixture was quenched with 1N HCl (aq.) and compounds were extracted with ethyl acetate. The collected organic fractions were dried over anhydrous sodium sulphate, concentrated in vacuo. Purification of crude mixture by flash chromatography (MeOH:DCM, 1:99) gave 4-(3-(4-(4-formyl-3-hydroxyphenoxy)butanamido) phenoxy)butanoic acid **S24** (87% yield). ^{1}H NMR (500 MHz, Acetone-d_6) δ 11.45 (s, 1H), 9.80 (s, 1H), 9.14 (s, 1H), 7.64 (d, J = 8.7 Hz, 1H), 7.44 (s, 1H), 7.15 (m, 2H), 6.71-6.55 (m, 2H), 6.47 (d, J = 2.2 Hz, 1H), 4.20 (t, J = 6.3 Hz, 2H), 4.02 (t, J = 6.3 Hz, 2H), 2.58 (t, J = 7.2 Hz, 2H), 2.50 (t, J = 7.3 Hz, 2H), 2.24-2.11 (m, 2H), 2.11-1.99 (m, 2H) ppm. ^{13}C NMR (126 MHz, Acetone-d_6) δ 196.1, 174.4, 171.4, 167.3, 165.2, 160.4, 141.6, 136.6, 130.3, 116.3, 112.4, 110.2, 109.3, 106.7, 102.0, 68.8, 67.6, 33.8, 33.7 30.7, 25.5 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{21}H_{24}NO_7$ 402.1553, found 402.1556.

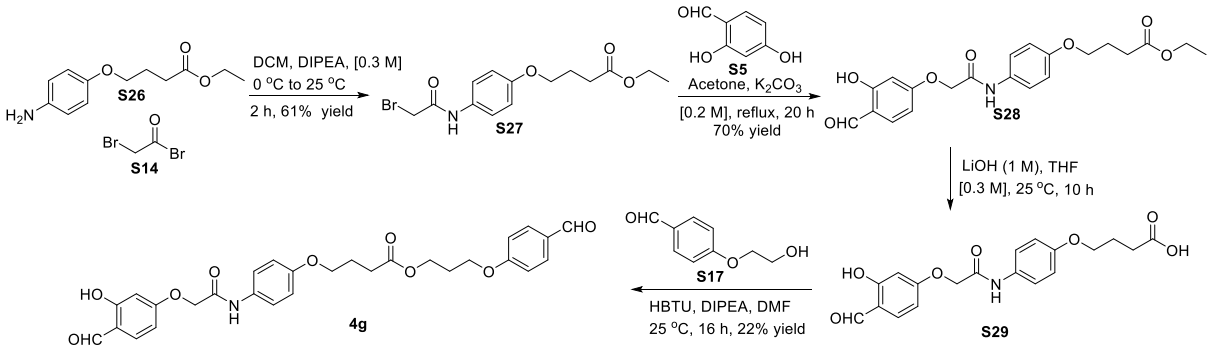
Synthesis of 2-(4-formylphenoxy)ethyl-4-(3-(4-(4-formyl-3-hydroxyphenoxy)butanamido) phenoxy)butanoate (4f)



4f

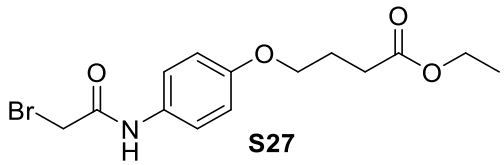
In a 25 ml round bottom flask, 4-(3-(4-(4-formyl-3-hydroxyphenoxy)butanamido)phenoxy)butanoic acid **S24** (100 mg, 0.25 mmol), EDC.HCl (81.7 µl, 0.375 mmol) and DIPEA (62.5 µl, 0.375 mmol) were dissolved in dry DCM:DMF (9:1, 6.6 ml). To this solution, 4-(3-hydroxypropoxy)benzaldehyde **S25** (49.6 mg, 0.3 mmol) was added and stirred at 25 ^{o}C. The progress of the reaction was followed by thin-layer chromatography. After 8 h, the reaction mixture was quenched with 1N HCl (aq.) and crude was extracted with DCM. The collected organic fractions were dried over anhydrous sodium sulphate and concentrated in vacuo. Purification of crude mixture by flash chromatography (MeOH:DCM,

1:99)) gave **4f** (37%) yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.44 (s, 1H), 9.88 (s, 1H), 9.70 (s, 1H), 7.82 (d, *J* = 8.7 Hz, 2H), 7.41 (d, *J* = 8.6 Hz, 1H), 7.30 (d, *J* = 12.0 Hz, 2H), 7.16 (t, *J* = 8.1 Hz, 1H), 6.99 (d, *J* = 8.7 Hz, 2H), 6.90 (d, *J* = 7.9 Hz, 1H), 6.61 (d, *J* = 8.2 Hz, 1H), 6.52 (dd, *J* = 8.6, 2.1 Hz, 1H), 6.41 (d, *J* = 2.0 Hz, 1H), 4.47 (m, 2H), 4.25 (m, 2H), 4.11 (t, *J* = 6.8 Hz, 2H), 3.98 (t, *J* = 6.0 Hz, 4H), 2.60-2.52 (m, 2H), 2.25-2.20 (m, 2H), 2.13-2.07 (m, 2H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.3, 190.8, 173.0, 170.1, 165.9, 164.4, 163.4, 159.3, 138.9, 138.4, 135.3, 132.0, 130.2, 129.7, 115.3, 114.8, 111.8, 110.5, 108.4, 106.1, 101.3, 67.3, 66.5, 66.1, 62.4, 33.6, 30.6, 24.6, 24.5 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{30}H_{31}NO_9$ 550.2072, found 550.2088.



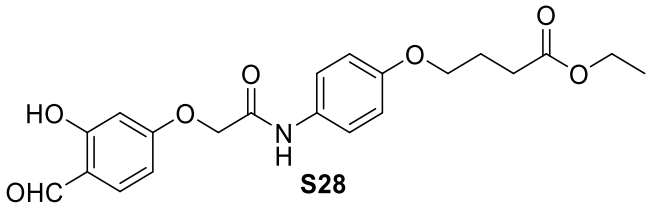
Scheme S11. 3-(4-formylphenoxy)propyl-4-(4-(2-(4-formyl-3-hydroxyphenoxy)acetamido) phenoxy)butanoate **4g**.[3]

Ethyl 4-(4-(2-bromoacetamido)phenoxy)butanoate (S27)



In a 50 ml round bottom flask, ethyl 4-(4-aminophenoxy)butanoate **S26** (300 mg, 1.34 mmol) and DIPEA (494 µl, 2.68 mmol) in DCM (4 ml) were stirred at 0 °C for 5 minutes. To this solution, 2-bromoacetyl bromide **S13** (236 µl, 2 mmol) was added slowly over 15 minutes. The reaction mixture was brought to room temperature and stirred for 2 h. The progress of the reaction was followed by TLC. The reaction mixture was concentrated using rotary evaporator. Purification of crude mixture by silica gel column chromatography using ethyl acetate:n-hexane (10:90) gave ethyl 4-(4-(2-bromoacetamido)phenoxy)butanoate **S27** (61% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.03 (s, 1H), 7.46-7.37 (m, 2H), 6.92-6.82 (m, 2H), 4.15 (q, *J* = 7.1 Hz, 2H), 4.02 (s, 2H), 3.99 (t, *J* = 6.1 Hz, 2H), 2.51 (t, *J* = 7.3 Hz, 2H), 2.16-2.05 (m, 2H), 1.26 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 173.4, 165.3, 156.3, 130.6, 122.1, 114.9, 67.1 (2C), 60.6, 30.9, 24.7, 14.4 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{14}H_{19}BrNO_4$ 344.0497, found 344.0482.

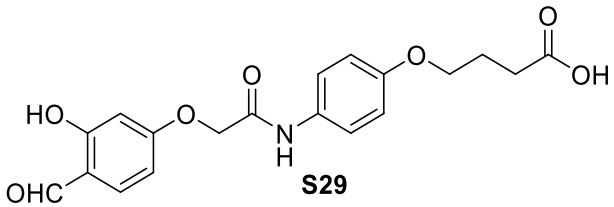
Ethyl 4-(4-(2-(4-formyl-3-hydroxyphenoxy)acetamido)phenoxy)butanoate (S28)



S28

In a 25 ml round bottom flask, ethyl 4-(4-(2-bromoacetamido)phenoxy)butanoate **S27** (344 mg, 1 mmol), potassium carbonate (138 mg, 1 mmol) and 2,4-dihydroxybenzaldehyde **S5** (138 mg, 1 mmol) were dissolved in acetone (5 ml) to reflux for 20 h. The progress of the reaction was followed by TLC. The reaction mixture was concentrated using rotary evaporator. Purification of crude mixture by silica gel column chromatography using ethyl acetate:n-hexane (10:90) gave ethyl 4-(4-(2-(4-formyl-3-hydroxyphenoxy)acetamido) phenoxy)butanoate **S28** (40% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.43 (s, 1H), 9.78 (s, 1H), 8.02 (s, 1H), 7.53 (d, *J* = 8.6 Hz, 1H), 7.49-7.43 (m, 2H), 6.91-6.85 (m, 2H), 6.65 (dd, *J* = 8.6, 2.4 Hz, 1H), 6.54 (d, *J* = 2.4 Hz, 1H), 4.65 (s, 2H), 4.15 (q, *J* = 7.1 Hz, 2H), 4.00 (t, *J* = 6.1 Hz, 2H), 2.51 (t, *J* = 7.3 Hz, 2H), 2.16-2.06 (m, 2H), 1.26 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.8, 173.3, 164.8, 161.0, 163.7, 156.4, 136.0, 129.7, 122.3, 116.5, 115.1, 108.0, 102.6, 67.5, 67.2, 60.6, 30.9, 24.8, 14.4 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{21}H_{24}NO_7$ 402.1553, found 402.1558.

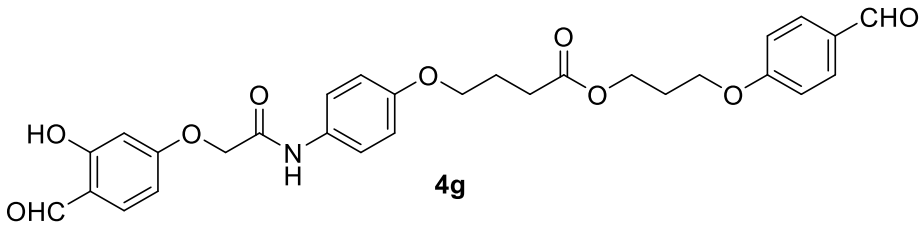
Synthesis of 4-(4-(2-(4-formyl-3-hydroxyphenoxy)acetamido)phenoxy)butanoic acid (S29)



S29

In a 25 ml round bottom flask, ethyl 4-(4-(2-(4-formyl-3-hydroxyphenoxy)acetamido) phenoxy) butanoate **S28** (401 mg, 1 mmol) and 1 N LiOH (aq., 3 ml) were dissolved in THF (0.3 ml) and stirred at room temperature. The progress of the reaction was followed by thin-layer chromatography. After 8 h, the reaction mixture was quenched with 4 N HCl (aq.) and compounds were extracted with ethyl acetate. The collected organic fractions were dried over anh. sodium sulfate and concentrated in vacuo. The purification of crude reaction mixture was performed by silica gel flash column chromatography using ethyl acetate:n-hexane (70:30) to isolate 4-(4-(2-(4-formyl-3-hydroxyphenoxy)acetamido)phenoxy)butanoic acid **S29** (48% yield). ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.04 (s, 1H), 10.00 (s, 1H), 7.64 (d, *J* = 8.7 Hz, 1H), 7.51 (d, *J* = 9.0 Hz, 2H), 6.89 (d, *J* = 9.0 Hz, 2H), 6.63 (dd, *J* = 8.7, 2.2 Hz, 1H), 6.52 (d, *J* = 2.2 Hz, 1H), 4.74 (s, 2H), 3.94 (t, *J* = 6.4 Hz, 2H), 2.37 (t, *J* = 7.3 Hz, 2H), 1.97-1.86 (m, 2H) ppm. ^{13}C NMR (126 MHz, DMSO-d_6) δ 190.8, 165.2, 164.4, 162.8, 154.8, 131.9, 131.4, 121.2, 116.7, 116.6, 114.4, 107.7, 101.7, 67.1, 66.7, 30.1, 24.2 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{19}H_{19}NO_7$ 396.1059, found 396.1057.

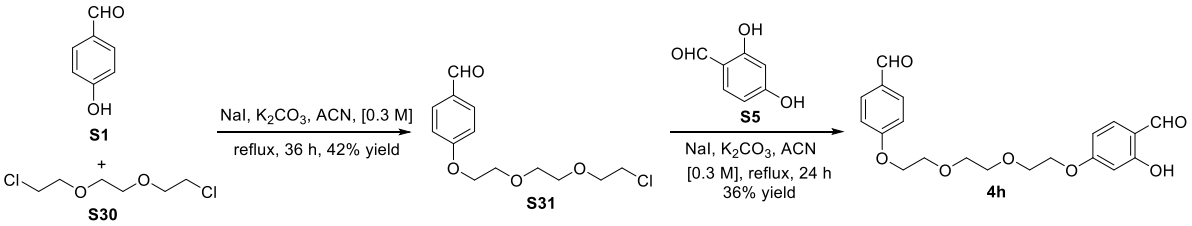
Synthesis of 3-(4-formylphenoxy)propyl-4-(4-(2-(4-formyl-3-ydroxyphenoxy)acetamido) phenoxy)butanoate (4g)



4g

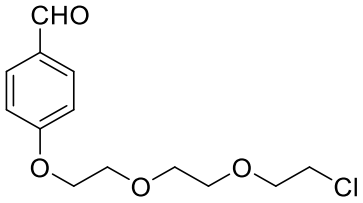
The compound **4g** was synthesized same as **4f**.

(22% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.44 (s, 1H), 9.88 (s, 1H), 9.71 (s, 1H), 7.81 (d, *J* = 8.7 Hz, 2H), 7.42 (d, *J* = 8.7 Hz, 1H), 7.36 (d, *J* = 8.9 Hz, 2H), 7.14 (s, 1H), 6.94 (d, *J* = 8.7 Hz, 2H), 6.83 (d, *J* = 8.9 Hz, 2H), 6.52 (dd, *J* = 8.6, 2.2 Hz, 1H), 6.42 (d, *J* = 2.1 Hz, 1H), 4.62 (s, 2H), 4.41 (t, *J* = 6.2 Hz, 2H), 4.12 (t, *J* = 5.9 Hz, 2H), 4.03 (t, *J* = 6.0 Hz, 2H), 2.55 (t, *J* = 7.1 Hz, 2H), 2.23 (m, 2H), 2.17 (m, 2H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.4, 190.9, 170.0, 169.0, 166.0, 164.5, 163.7, 154.5, 135.4, 132.0, 130.1, 121.6, 115.3, 115.0, 114.8, 108.4, 101.4, 67.4, 65.7, 64.4, 61.8, 33.4, 28.3, 24.7 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{29}H_{29}NO_9$ 536.1915, found 536.1927.



Scheme 12. Synthesis of 4-(2-(2-(2-(4-formylphenoxy)ethoxy)ethoxy)ethoxy)-2-hydroxybenzaldehyde 4h.

Synthesis of 4-(2-(2-(2-chloroethoxy)ethoxy)benzaldehyde (S31).

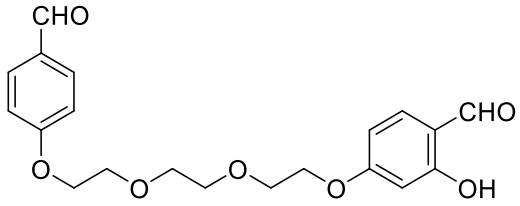


S31

In a 50 ml round bottom flask, *p*-hydroxy benzaldehyde **S1** (1220 mg, 10 mmol), 1,2-bis(2-chloroethoxy)ethane **S30** (4715 mg, 30 mmol), sodium iodide (4470 mg, 30 mmol) and K_2CO_3 (2760 mg, 20 mmol) were taken and dissolved in 30 ml acetonitrile. The reaction mixture was refluxed and the progress of the reaction was monitored by thin layer chromatography. After 36 h, the reaction mixture was concentrated, the compound was extracted in DCM and purified by silica gel flash column chromatography using ethyl acetate:n-hexane (20:80) to obtain 4-(2-(2-(2-chloroethoxy)ethoxy)benzaldehyd **S31** (42 % yield). ^{1}H NMR (500MHz, $CDCl_3$) δ 9.87 (s, 1H), 7.81 (d, *J* = 8.6 Hz, 2H), 7.01 (d, J = 8.6 Hz, 2H), 4.23–4.15 (m, 2H), 3.91–3.86 (m,

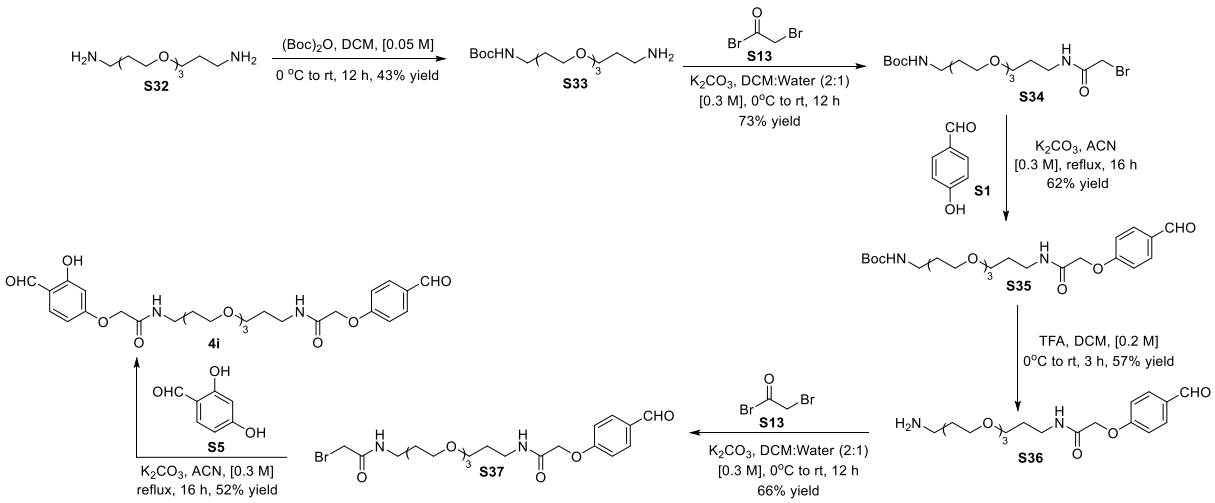
2H), 3.78–3.66 (m, 6H), 3.62 (t, J = 5.8 Hz, 2H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 190.8, 163.8, 131.9, 130.0, 114.9, 71.4, 70.9, 70.7, 69.6, 67.8, 42.7 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{13}H_{18}{}^{35}ClO_4$ 273.0894, found 273.0895.

Synthesis of 4-(2-(2-(2-(4-formylphenoxy)ethoxy)ethoxy)ethoxy)-2 hydroxybenzaldehyde (4h).



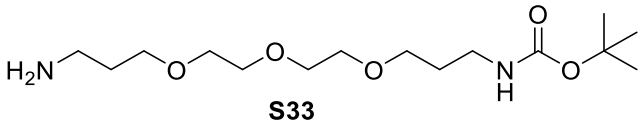
4h

In a 10 ml round bottom flask 4-(2-(2-(2-chloroethoxy)ethoxy)benzaldehyde **S31** (374 mg, 1 mmol**),** 2,4-dihydroxy benzaldehyde **S5** (276 mg, 2 mmol), sodium iodide (447 mg, 3 mmol) and K_2CO_3 (276 mg, 2 mmol) were taken and dissolved in 3 ml acetonitrile. The reaction mixture was refluxed, and the progress of the reaction was monitored by thin layer chromatography. After 24 h, the reaction mixture was concentrated, the compound was extracted in DCM and purified by silica gel flash column chromatography using ethyl acetate:n-hexane (30:70) to obtain 4-(2-(2-(2-(4-formylphenoxy)ethoxy)ethoxy)ethoxy)-2 hydroxybenzaldehyde **4h** (36% yield). ^{1}H NMR (500MHz, $CDCl_3$) δ 11.45 (s, 1H), 9.87 (s, 1H), 9.70 (s, 1H), 7.81 (d, *J* = 8.8 Hz 2H), 7.41 (d, *J* = 8.7 Hz, 1H), 7.00 (d, *J* = 8.7 Hz, 2H), 6.54 (dd, *J* = 8.7 Hz, 2.3 Hz, 1H),), 6.42 (d, *J* = 2.3 Hz, 1H), 4.20 (t, *J* = 4.7 Hz, 2H), 4.16 (t, *J* = 4.7 Hz, 2H), 3.90–3.86 (m, 4H), 3.75 (s, 4H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.5, 190.9, 166.1, 164.5, 163.9, 135.4, 132.1, 130.2, 115.2, 108.8, 101.4, 71.1, 69.9, 69.5, 68.0, 67.8 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{20}H_{23}O_7$ 375.1444 found 375.1404.



Scheme 13. Synthesis of 2-(4-formyl-3-hydroxyphenoxy)-N-(3-(3-(2-(4-formylphenoxy) acetamido)propyl)acetamide 4i.

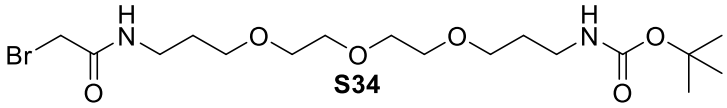
Tert-butyl (3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)carbamate (S33).[4]



S33

In a 250 ml round bottom flask, 4,7,10-trioxa-1,13-tridecanediamine **S32** (26.13 mmol, 5.7 g) was taken and dissolved in 116 ml of DCM. To this solution, Boc anhydride (8.71 mmol, 1.9 g) in 58 ml DCM was added dropwise via dropping funnel at 0 °C. After the addition was complete, reaction mixture was stirred at room temperature for overnight. The solvent was removed under reduced pressure and the residue was purified by silica gel chromatography (MeOH:$CHCl_3$ 3:97) to afford *t*ert-butyl (1-bromo-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S33** (43% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 3.63-3.61 (m, 4H), 3.60-3.56 (m, 4H), 3.56-3.52 (m, 4H), 3.22 (d, *J* = 6.0 Hz, 2H), 2.79 (t, *J* = 6.7 Hz, 2H), 1.79-1.69 (m, 4H), 1.45 (s, 9H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 156.2, 78.7, 70.5, 70.4, 70.2, 70.1, 69.5, 69.4, 39.5, 38.4, 33.3, 29.5, 28.4 ppm. LRMS (ESI) $[M+H]^+$ calcd. $C_{15}H_{33}N_2O_5$ 321.2, found 321.1.

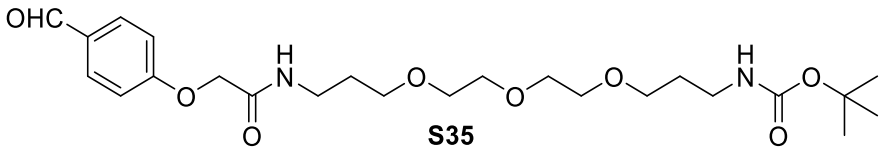
Tert-butyl (1-bromo-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate (S34).[4]



S34

In 25 ml round bottom flask, tert-butyl (3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)carbamate **S33** (4.7 mmol, 1.5g) was dissolved in 5 ml DCM, and K_2CO_3 (9.4 mmol, 1.3 g) in 6 ml of H_2O was added to it. Bromoacetyl bromide **S13** (9.4 mmol, 821 mg), dissolved in DCM (5 ml), was added drop wise to the mixture at 0 °C. The reaction mixture was stirred for 12 h and the progress of the reaction was analyzed by using thin layer chromatography. After completion, the reaction mixture was extracted in DCM and the solution was concentrated and purified using silica gel flash column chromatography (MeOH:$CHCl_3$ 3:97) to afford tert-butyl (1-bromo-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S34** (73% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 3.78 (s, 2H), 3.62-3.50 (m, 10H), 3.46 (t, *J* = 6.0 Hz, 2H), 3.33 (dd, *J* = 12.1, 5.8 Hz, 2H), 3.14 (d, *J* = 6.0 Hz, 2H), 1.77-1.72 (m, 2H), 1.68 (p, J = 6.2 Hz, 2H), 1.36 (s, 9H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 165.6, 156.0, 78.7, 70.4, 70.2, 70.1, 70.0, 69.4, 69.4, 38.7, 38.4, 29.6, 29.1, 28.5, 28.4 ppm. LRMS (ESI) $[M+H]^+$ calcd. $C_{17}H_{34}BrN_2O_6$ 441.1, found 441.1.

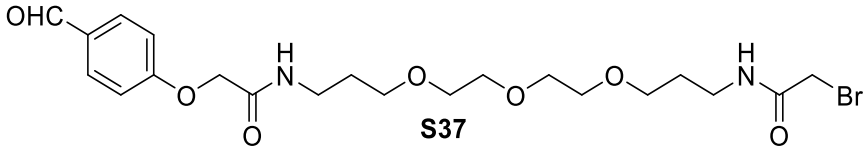
Tert-butyl (1-(4-formylphenoxy)-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate (S35).



S35

In a 10 ml round bottom flask, tert-butyl (1-bromo-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S34** (1.1 mmol, 500 mg) was dissolved in 4 ml acetonitrile. To this solution, K_2CO_3 (311 mg, 2.3 mmol) and *p*-hydroxybenzaldehyde **S1** (2.3 mmol, 276 mg) were added, and the reaction mixture was allowed to reflux and monitored using thin layer chromatography. After 16 h, the reaction mixture was filtered, concentrated and the product was purified using silica gel column chromatography (MeOH:DCM 5:95) to afford tert-butyl (1-(4-formylphenoxy)-2-oxo-7,10,13-trioxa-3-az3ahexadecan-16-yl)carbamate **S35** (62% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.90 (s, 1H), 7.86 (d, *J* = 8.8 Hz, 2H), 7..08 (s, 1H), 7.05 (d, *J* = 8.7 Hz, 2H)), 4.55 (s, 2H), 3.64-3.53 (m, 10H), 3.51-3.45 (m, 4H), 3.20 (m, 2H), 1.81 (p, 2H), 1.73 (p, 2H), 1.42 (s, 9H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 190.5, 167.0, 162.0, 155.9, 131.9, 130.8, 115.0, 78.6, 70.3, 70.1, 70.0, 69.7, 69.3, 67.3, 38.3, 37.4, 29.6, 28.8, 28.3 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{24}H_{39}N_2O_8$ 483.2706, found 483.2705.

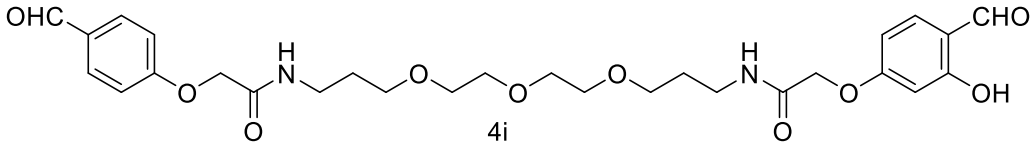
2-bromo-N-(1-(4-formylphenoxy)-2-oxo-7,10,13-tioxa-3-azahexadecan-16-yl)acetamide (S37).



S37

In a 10 ml round bottom flask, tert-butyl (1-(4-formylphenoxy)-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S3** (482 mg, 1.0 mmol), was mixed with 5 ml dichloromethane. To this solution trifluoroacetic acid (0.4 ml, 5 mmol) was added dropwise at 0 ºC. After addition, reaction mixture was allowed to stir at room temperature for 3 h. The reaction was monitored using thin layer chromatography and upon completion of the reaction, the solution was concentrated under vacuum to afford N-(3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)-2-(4-formylphenoxy)acetamide **S36** (54% yield). HRMS (ESI) $[M+H]^+$ calcd. $C_{19}H_{30}N_2O_6$ 383.2182, found 383.2192. The compound **S37** is synthesized according to the synthesis of compound **S34** (66% yield).^{1}H NMR (500 MHz, $CDCl_3$) δ 9.91 (s, 1H), 7.87(d, J = 8.7 Hz, 2H), 7.16 (bs, 1H), 7.05 (d, J = 8.7 Hz, 2H), 4.56 (s, 2H), 3.84 (s, 2Hz), 3.64-3.54(m, 12H), 3.46 (dd, J = 12.4, 6.2 Hz, 2H), 3.39 (dd, *J* = 12.1, 5.9 Hz, 2H), 1.85-1.77 (m, 4H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 190.6, 167.2, 165.5, 162.1, 132.1, 131.1, 115.1, 70.5, 70.3(d), 70.3 69.9, 67.5, 39.0, 37.6, 29.3, 29.0, 28.6 ppm. HRMS (ESI) $[M+H]^+$ calcd. $C_{21}H_{32}BrN_2O_7$ 503.1393, found 503.1369.

2-(4-formyl-3-hydroxyphenoxy)-N-(1-(4-formylphenoxy)-2-oxo-7, 10, 13-trioxa-3-azahexadecan-16-yl)acetamide (4i).



In a 10 ml round bottom flask, 4-hydroxybenzaldehyde **S1** (276 mg, 2 mmol) was dissolved in acetonitrile (3 ml). To this solution, K_2CO_3 (276 mg, 2 mmol) and 2-bromo-N-(1-(4-formylphenoxy)-2-oxo-7, 10, 13-tioxa-3-azahexadecan-16-yl)acetamide **S37** (1 mmol, 502 mg) were added and the reaction mixture was allowed to reflux for 16 h. The reaction was monitored using thin layer chromatography. After completion, the reaction mixture was filtered, concentrated under vacuum and the product was purified using silica gel column chromatography (MeOH:DCM 2:98) to isolate 2-(4-formyl-3-hydroxyphenoxy)-N-(1-(4-formylphenoxy)-2-oxo-7, 10, 13-trioxa-3-azahexadecan-16-yl)acetamide **4i (**52% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.38 (s, 1H), 9.89 (s, 1H), 9.73 (s, 1H), 7.84 (d, J = 8.7 Hz, 2H), 7.47 (d, J = 8.6 Hz), 7.07 (s, 1 H), 7.03 (d, J = 8.7 Hz, 2H), 6.56 (dd, 8.6, 2.3 Hz, 1H), 6.44 (d, 2.2 Hz, 1H), 4.54 (s, 2H), 4.50 (s, 2H), 3.59-3.51 (m, 12H), 3.47-3.41 (m, 4H), 1.79 (m, 4H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 194.6, 190.6, 167.1, 166.8, 164.2, 164.1, 162.0, 135.7, 132.0, 131.0, 116.1, 115.0, 107.7, 102.3, 70.4, 70.2, 69.8, 67.4, 67.3, 37.5, 29.0 ppm. HRMS (ESI) $[M+H]^+$ calcd. 561.2448 $C_{28}H_{37}N_2O_{10}$, found 561.2412.

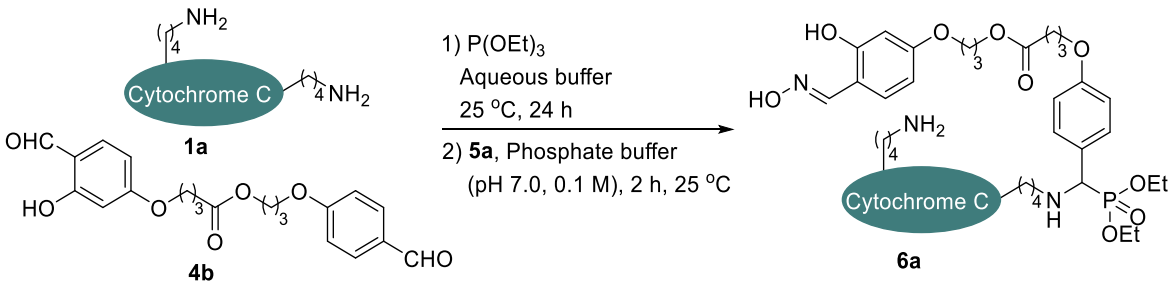
4. Additional results and discussion

Single-site modification beyond reactivity hotspots:

We and others have targeted reactivity hotspots in the past to establish precision engineering of proteins. In these cases, it is possible to achieve good to excellent yields without compromising the site-selectivity. However, the reactivity landscape is entirely different as soon as we go beyond the reactivity hotspot. It requires a combination of parameters to work cooperatively to deliver the covalent single-site modification beyond the reactivity order defined by the proteins. In the $LDM_{K\text{-}K}$ method outlined in this manuscript, the sequence starts with the linchpin formation at most of the solvent-accessible lysine residues (step 1). However, the extent varies from one Lys to other. A fraction of the selected intermediate would find a suitable proximal Lys to form the second imine (step 2). We believe that the adaptive conformational preferences of spacer along with its interaction with the surface of protein determines the effective placement of the two terminal functional groups of LDM reagent. Besides, the addition of external nucleophile (step 3) in the final step must target a relatively less accessible imine as the more accessible neighboring Lys forms the linchpin. It explains the observed conversions (up to 72%) that emerge from over three steps. At the same time, this multi-step process is essential to deliver modularity and allows one to go beyond the inherent reactivity order of the proteinogenic residues.

4.1 Reaction parameter optimization

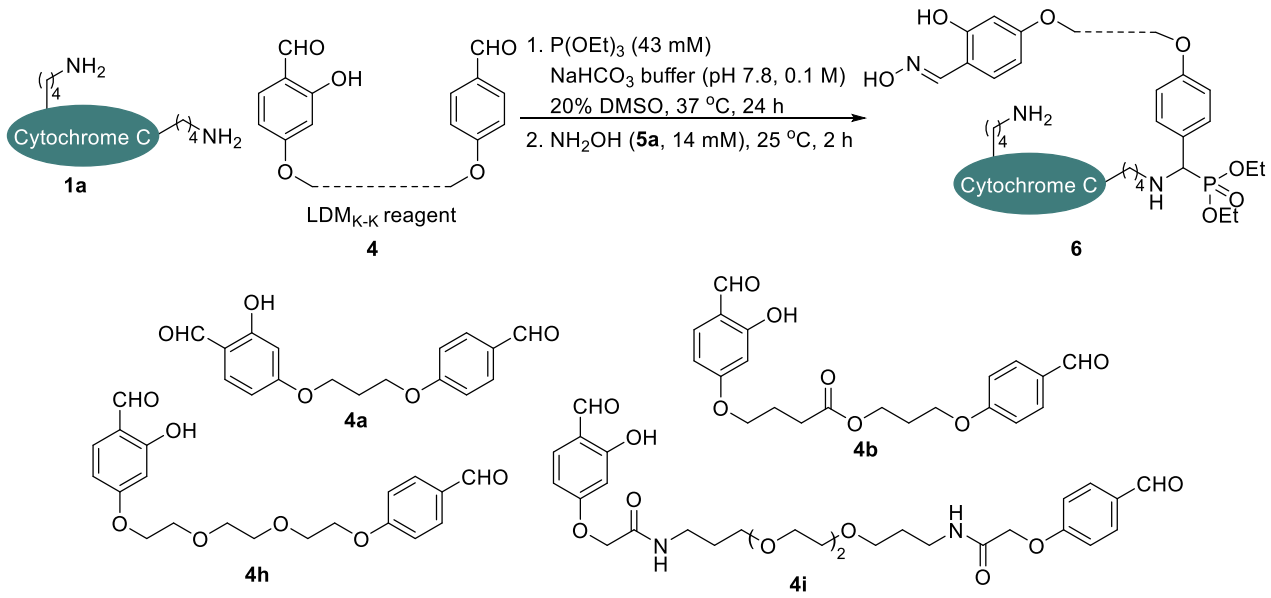
Table S1. Optimization of reagent stoichiometry and buffer



Entries	Aldehyde (4b) (equiv.)	$P(OEt)_3$ (equiv.)	Aqueous buffer (pH 7.8, 0.1 M)	6a (% Conversion)[a]
1	5	25	Phosphate	--
2	5	50	Phosphate	No reaction
3	5	100	Phosphate	<5%
4	10	25	Phosphate	No reaction
5	10	50	Phosphate	<10%
6	10	100	Phosphate	15%
7	10	300	Phosphate	18%
8	10	600	Phosphate	24%
9	10	600	Phosphate	29%[b]
10	10	600	MOPS	No reaction
11	10	600	HEPES	No reaction
12	10	600	Tris	14%
13	10	600	$NaHCO_3$	35%[b]

[a] % Conversions are estimated based on LC-ESI-MS. [b] Reactions performed at 37 °C.

Table S2. Reaction of cytochrome C (**1a**) with diverse relative stoichiometry of selected LDM_{K-K} reagents (**4a**, **4b**, **4h**, and **4i**) in 24 h.



Entry	LDM_{K-K} reagent	Equivalent(s)	Product (6)	% Conversion[a]
1	**4a**	1	**6g**	10 m
2		5		30 m
3		10[b]		27 m, 15 b
4		20		46 m, 20 b, 5t
5	**4b**	1	**6a**	3 m
6		5		18 m
7		10		35 m
8		20		32 m, 7 b
9	**4h**	1	**6b**	10 m
10		5		41 m, 6 b
11		10		42 m, 19 b
12		20		27 m
13	**4i**	1	**6m**	9 m
14		5		31 m
15		10[c]		30 m, 5 b
16		20		37 m, 11 b

[a] % Conversions are estimated based on LC-ESI-MS. m = mono-labeled, b = bis-labeled, t = tris-labeled. The data for entries 1-16 are given below in Figures S1-S18. [b]At 16 h, mono-labeled cytochrome C is formed with 18% conversion (data: Figure S4). [c] At 16 h, mono-labeled cytochrome C is formed with 27% conversion (data: Figure S17).

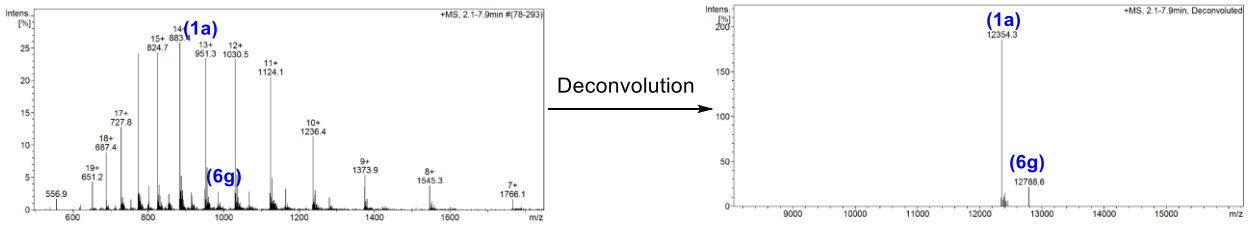


Figure S1. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4a** in 24 h (entry 1, Table S1).

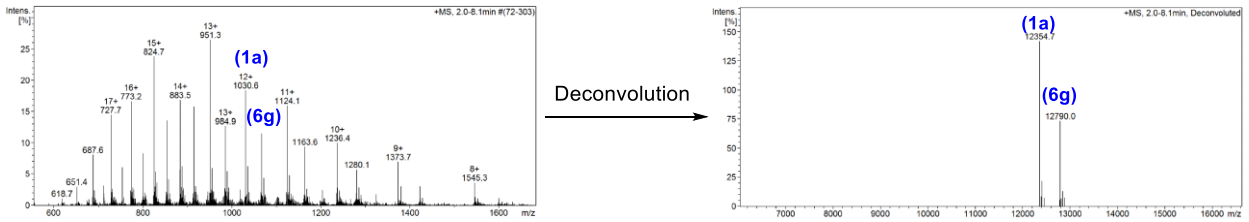


Figure S2. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4a** in 24 h (entry 2, Table S1).

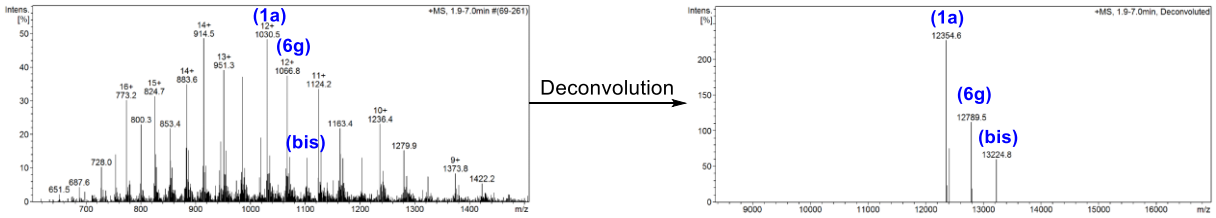


Figure S3. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4a** in 24 h (entry 3, Table S1).

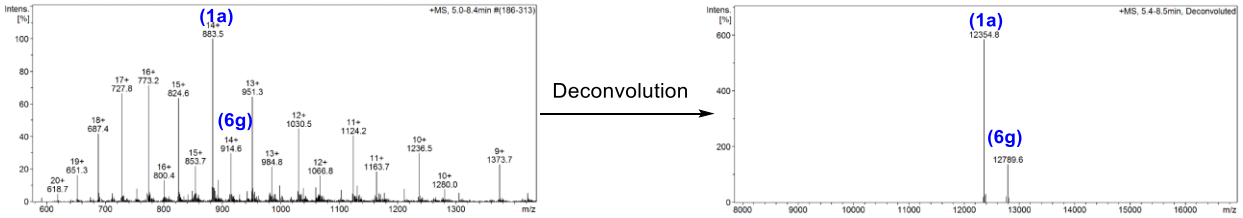


Figure S4. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4a** in 16 h (entry 3, Table S1).

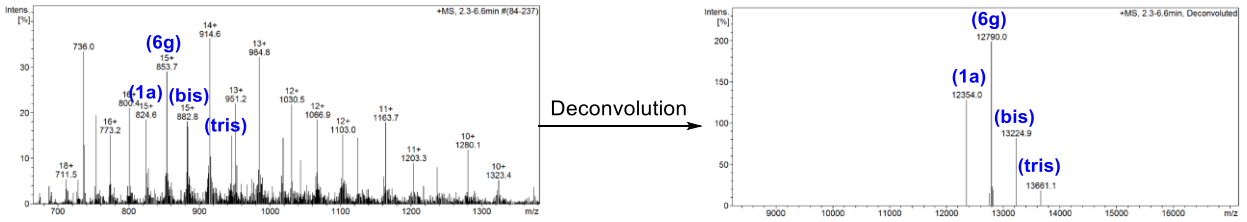


Figure S5. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4a** in 24 h (entry 4, Table S1).

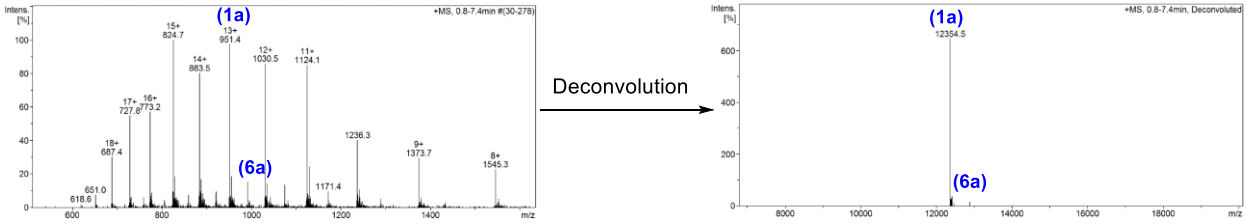


Figure S6. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4b** in 24 h (entry 5, Table S1).

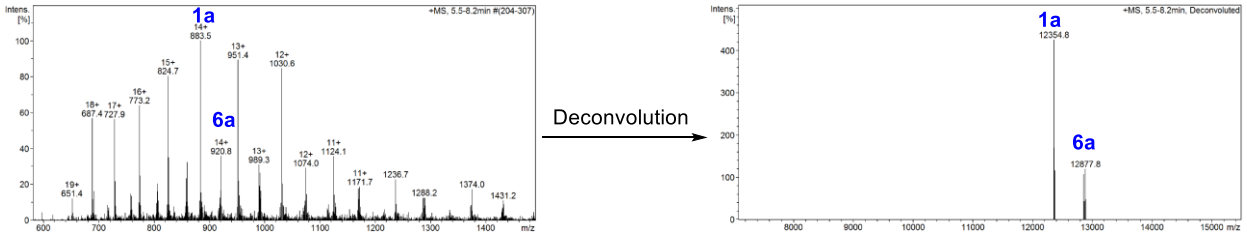


Figure S7. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4b** in 24 h (entry 6, Table S1).

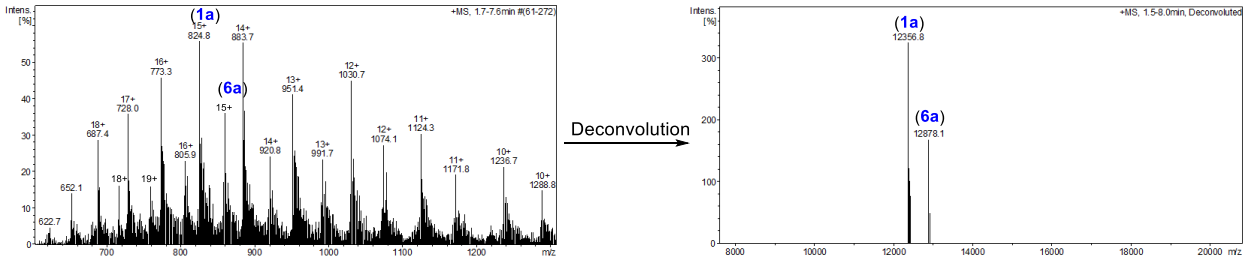


Figure S8. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4b** in 24 h (entry 7, Table S1).

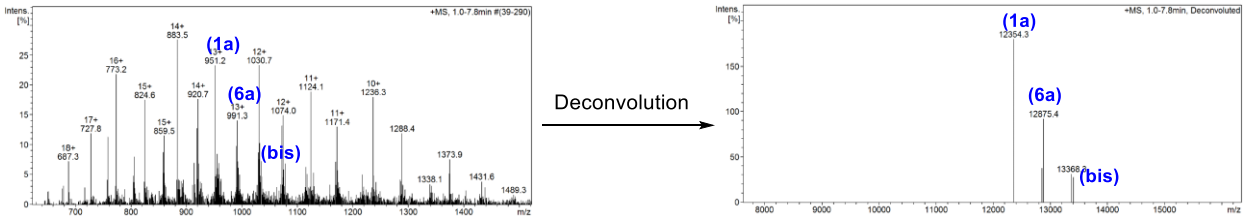


Figure S9. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4b** in 24 h (entry 8, Table S1).

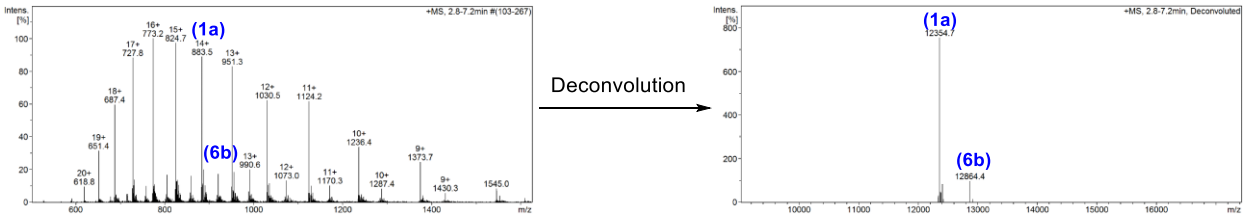


Figure S10. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4h** in 24 h (entry 9, Table S1).

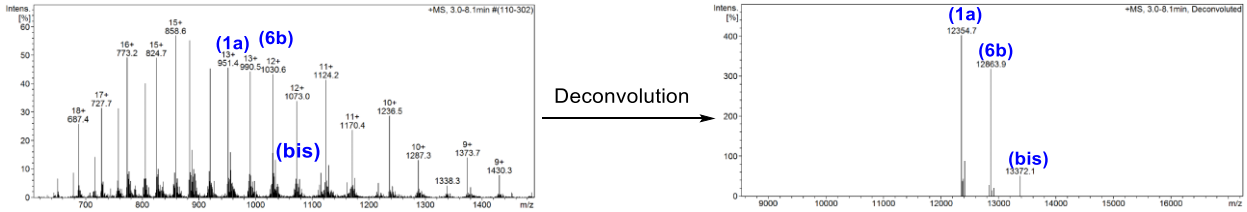


Figure S11. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4h** in 24 h (entry 10, Table S1).

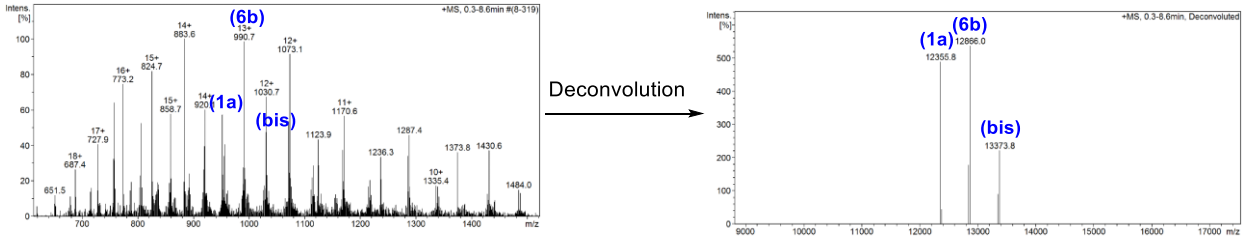


Figure S12. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4h** in 24 h (entry 11, Table S1).

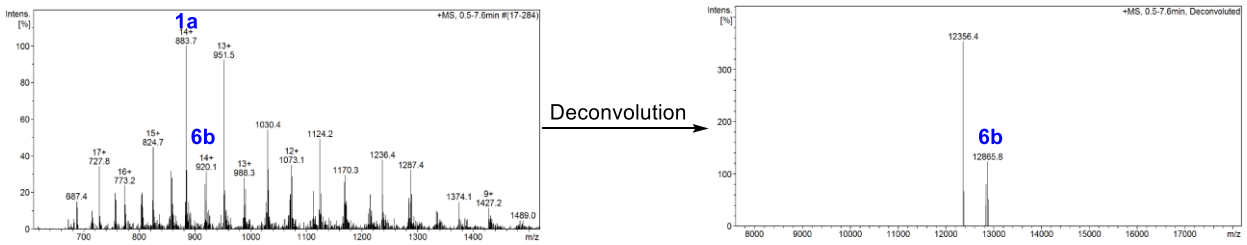


Figure S13. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4h** in 24 h (entry 12, Table S1).

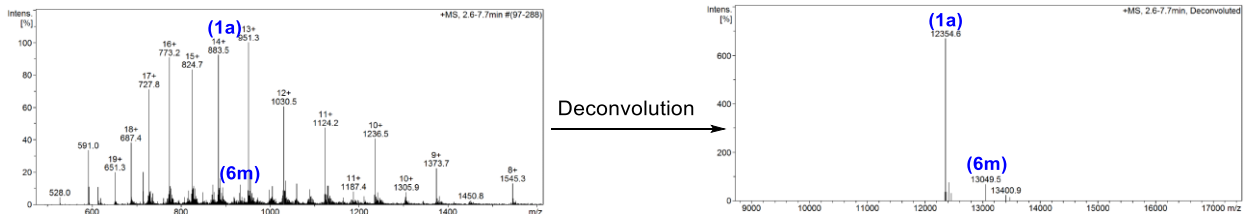


Figure S14. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 13, Table S1).

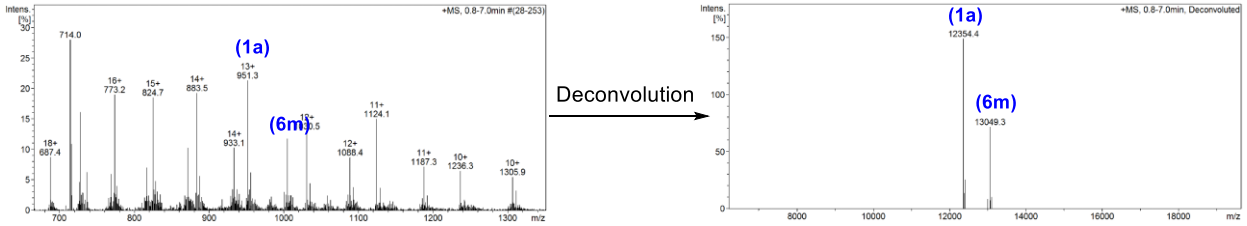


Figure S15. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 14, Table S1).

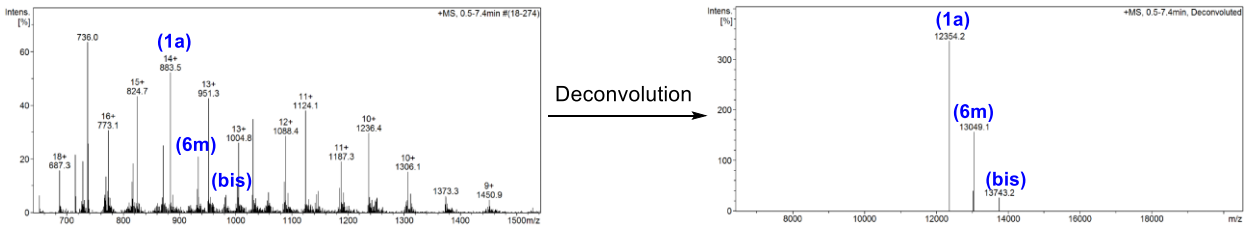


Figure S16. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 15, Table S1).

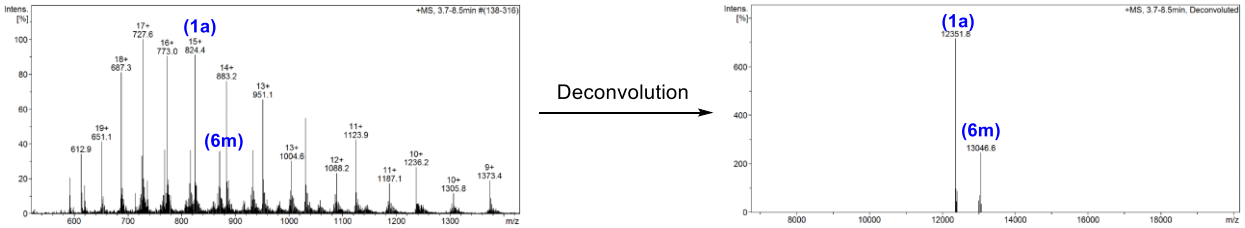


Figure S17. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 16 h (entry 15, Table S1).

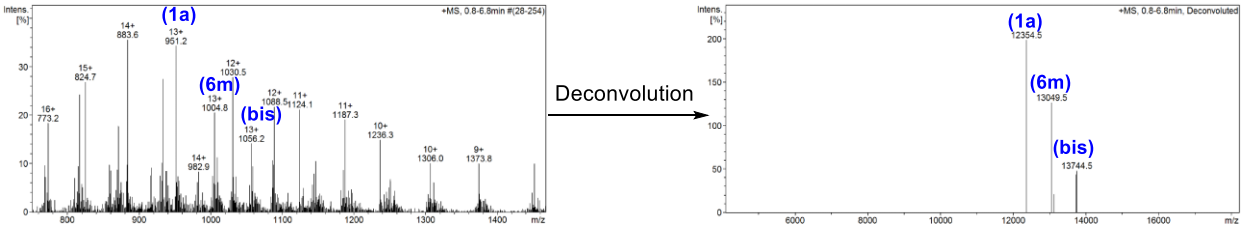


Figure S18. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 16, Table S1).

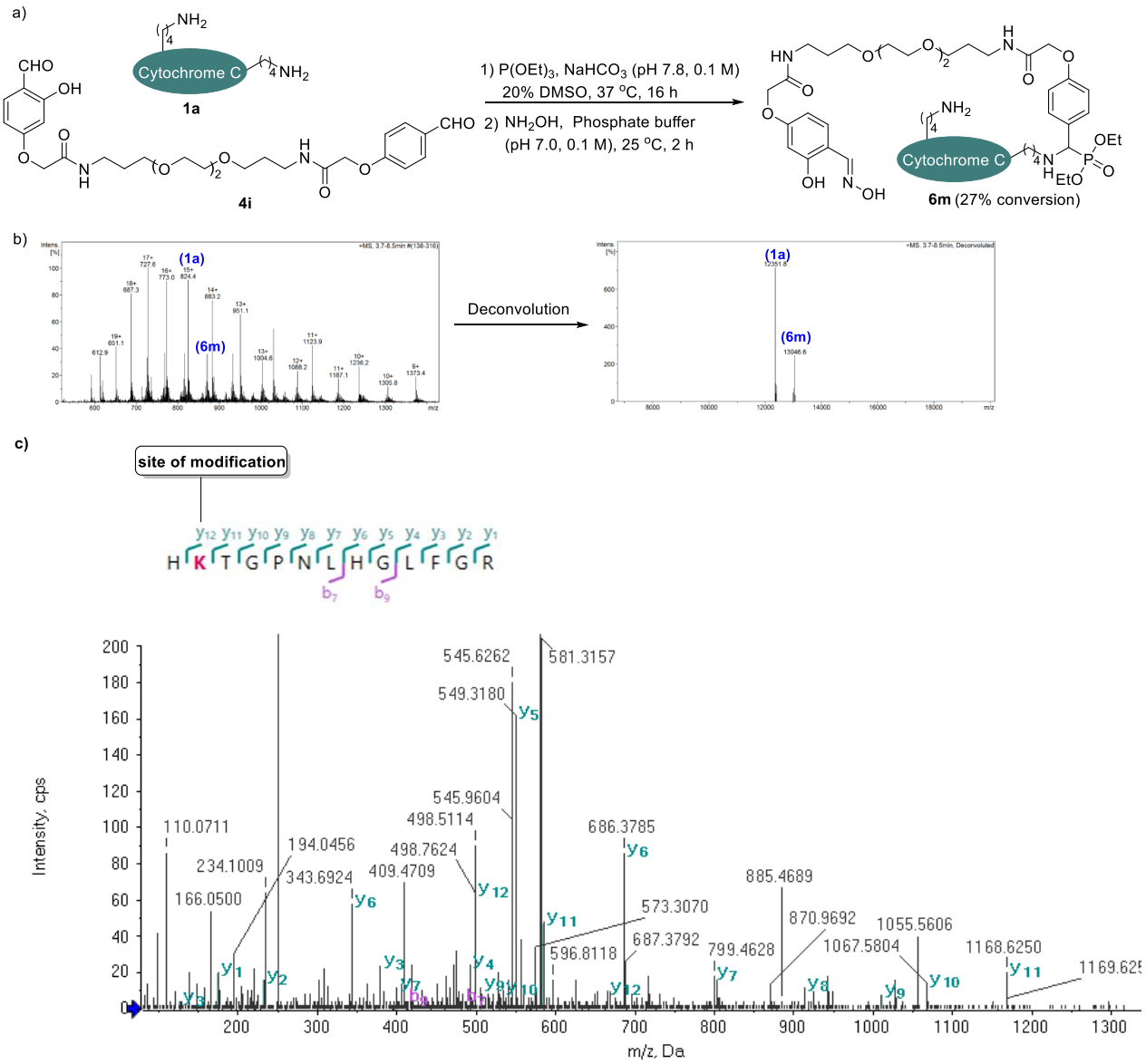


Figure S19. (a) Single site labeling of cytochrome C enabled by $LDM_{K\text{-}K}$ reagent **4i**. (b) LC-ESI-MS spectra of reaction mixture containing cytochrome C (**1a**) and mono-labeled cytochrome C (**6m**). (c) MS-MS spectrum of labeled HKTGPNLHGLFGR (H26-R38, *m/z* 533.02 $[\{(M+4)/4\}+H]^+$) confirms the site of modification (K27).

Additional examples of single-site modification of cytochrome C with the remaining LDM reagents (4c, 4d, 4e, 4f, and 4g)

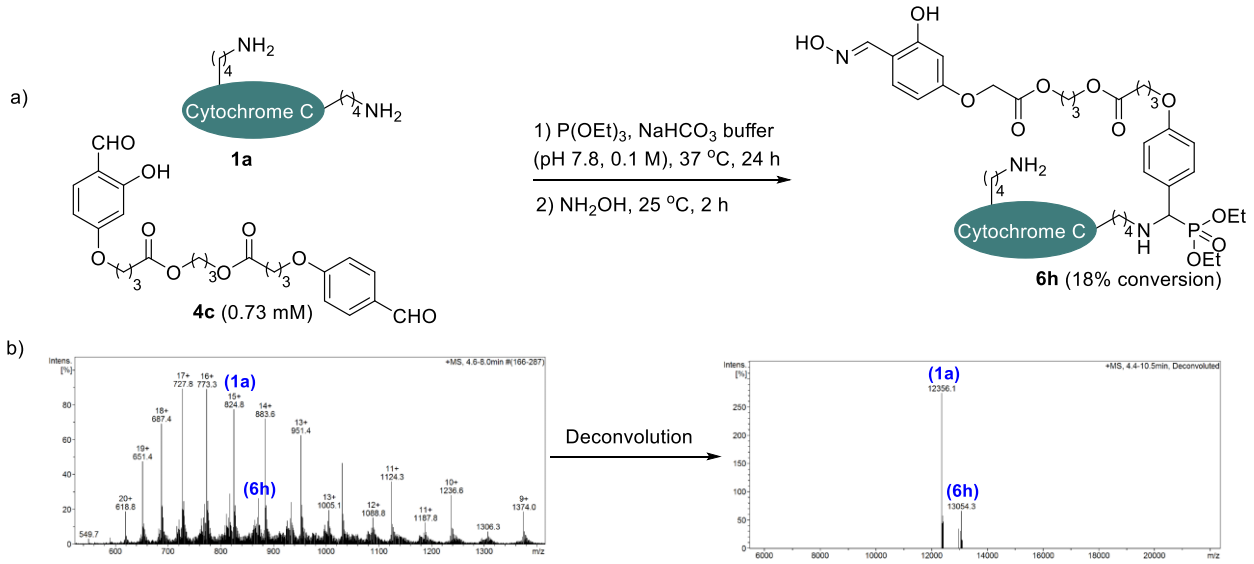


Figure S20. (a) Single site labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4c**. (b) LC-ESI-MS spectra of the reaction mixture containing cytochrome C (**1a**) and mono-labeled cytochrome C (**6h**).

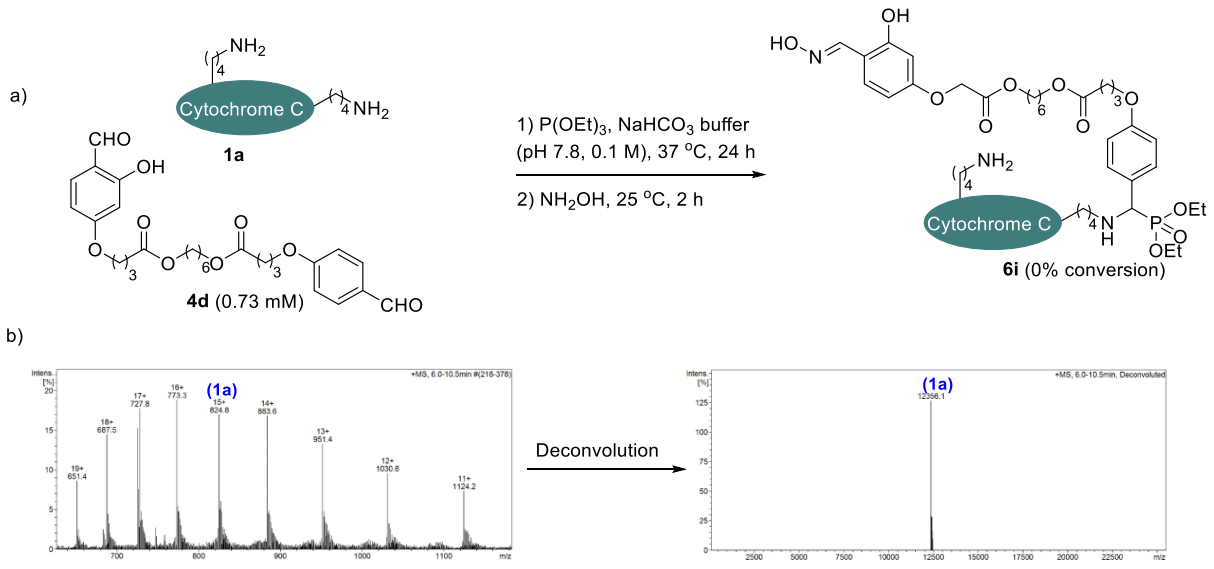


Figure S21. (a) Reaction of cytochrome C (**1a**) with LDM$_{K\text{-}K}$ reagent **4d**. (b) LC-ESI-MS spectra of the reaction mixture.

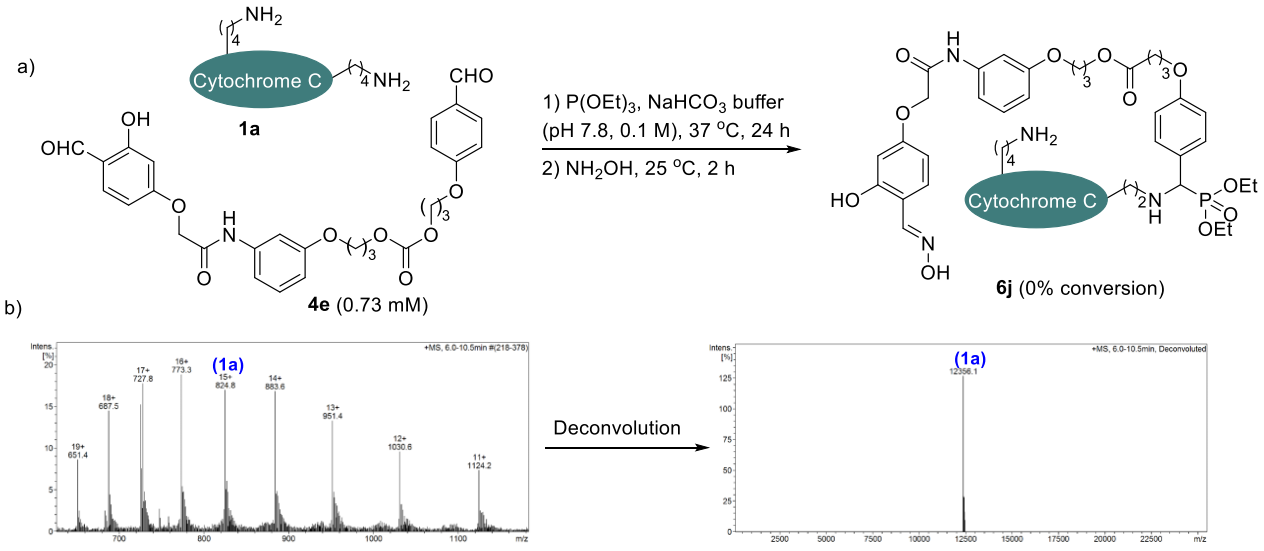


Figure S22. (a) Reaction of cytochrome C (**1a**) with LDM$_{K\text{-}K}$ reagent **4e**. (b) LC-ESI-MS spectra of the reaction mixture.

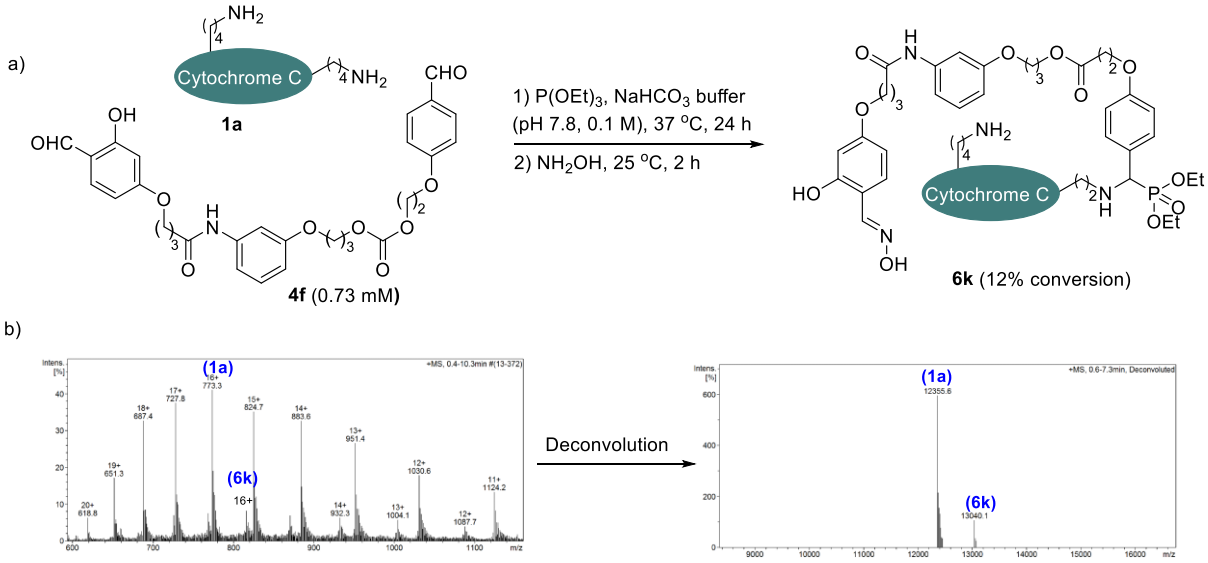


Figure S23. (a) Site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4f**. (b) LC-ESI-MS spectra of the reaction mixture containing cytochrome C (**1a**) and mono-labeled cytochrome C (**6k**).

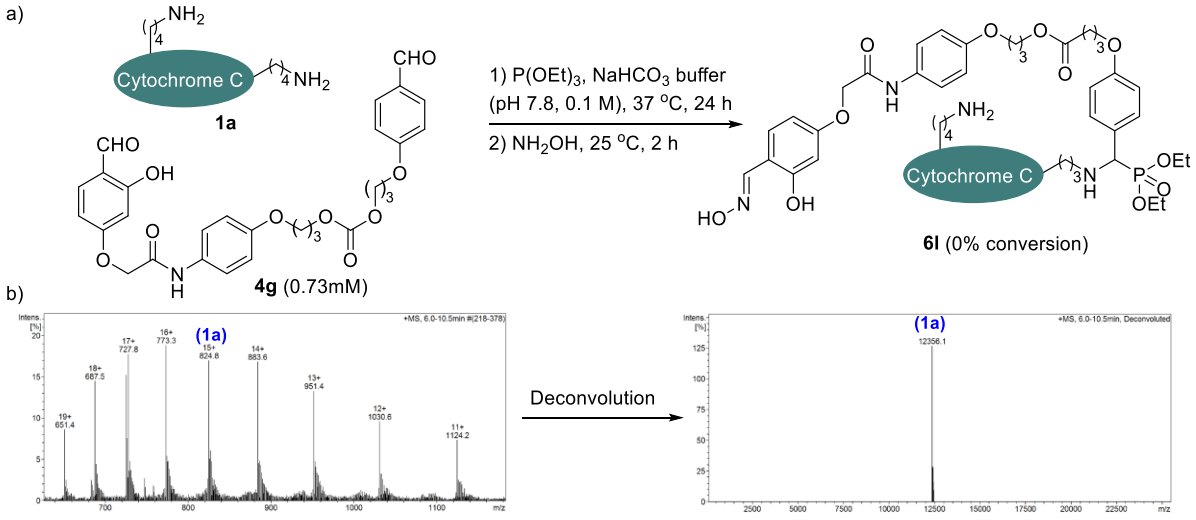
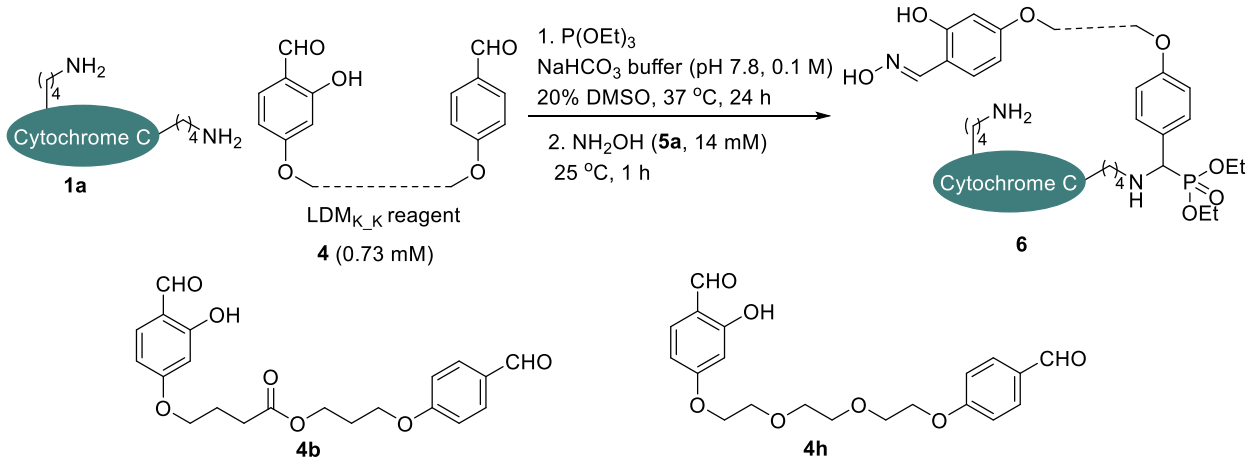


Figure S24. (a) Reaction of cytochrome C with $LDM_{K\text{-}K}$ reagent **4g**. (b) LC-ESI-MS spectra of the reaction mixture.

Table S3. Reaction of cytochrome C with diverse relative stoichiometry of triethylphosphite.



Entry	LDM reagent	$P(OEt)_3$ (equiv.)	Product (6)	% Conversion[a]
1	**4b**	200	**6a**	13 m
2		400		27 m
3		600		35 m
4		800		65 m, 12 b
5		1000		29 m, 12 b
6	**4h**	200	**6b**	9 m
7		400		33 m, 4 b
8		600		42 m, 19 b
9		800		41 m, 19 b
10		1000		37 m, 29 b

[a] %Conversions are estimated based on LC-ESI-MS. m = mono-labeled, b = bis-labeled. The data for entries 1-10 are given below in Figures S25-S34.

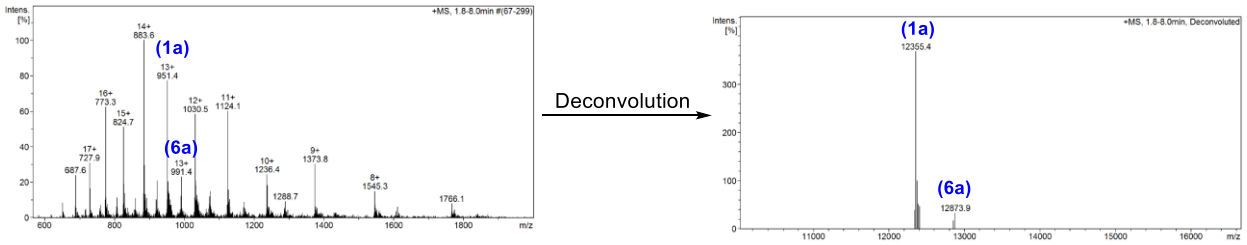


Figure S25. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4b** in 24 h (entry 1, Table S3).

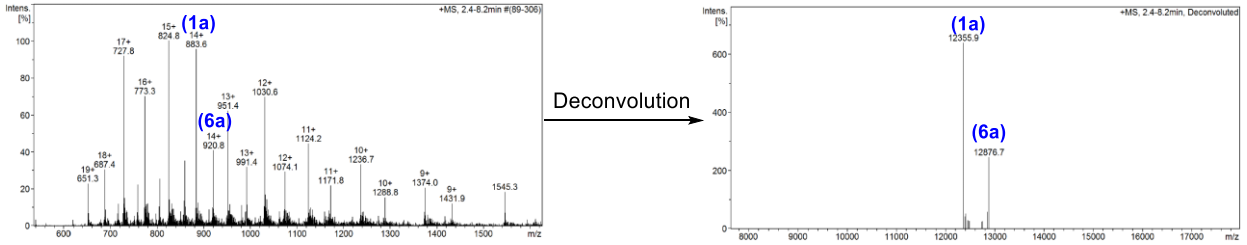


Figure S26. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4b** in 24 h (entry 2, Table S3).

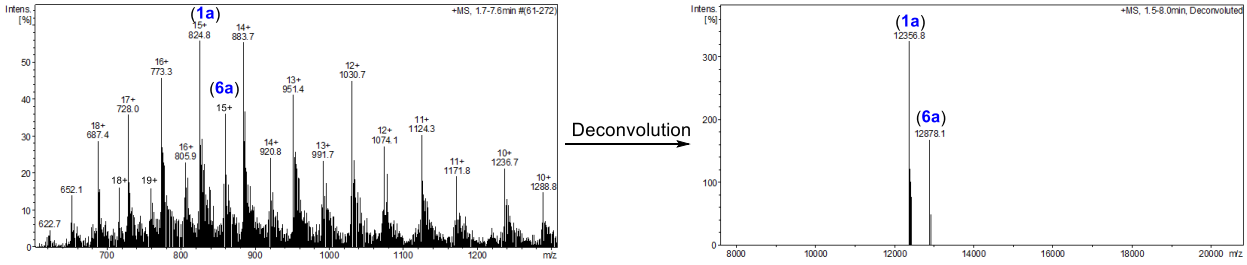


Figure S27. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4b** in 24 h (entry 3, Table S3).

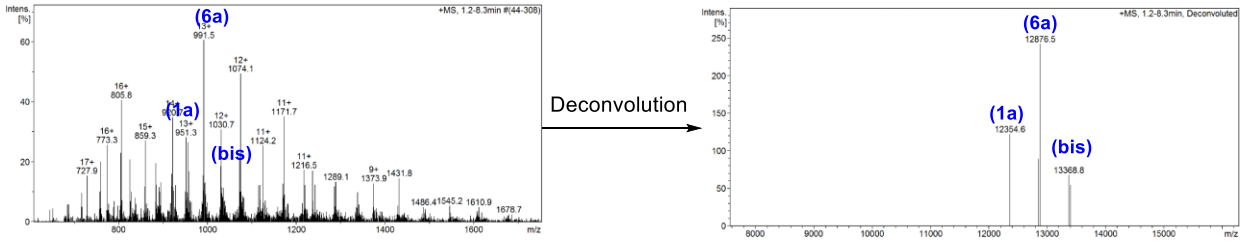


Figure S28. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4b** in 24 h (entry 4, Table S3).

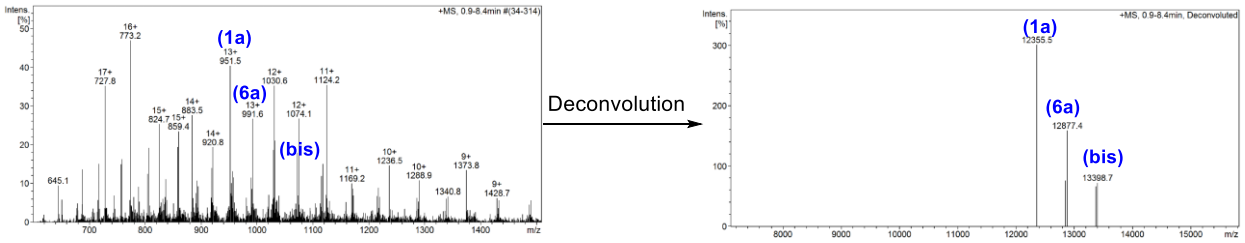


Figure S29. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4b** in 24 h (entry 5, Table S3).

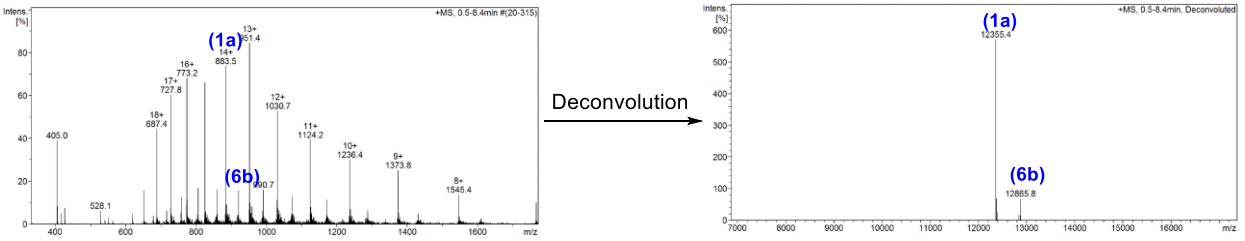


Figure S30. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4h** in 24 h (entry 6, Table S3).

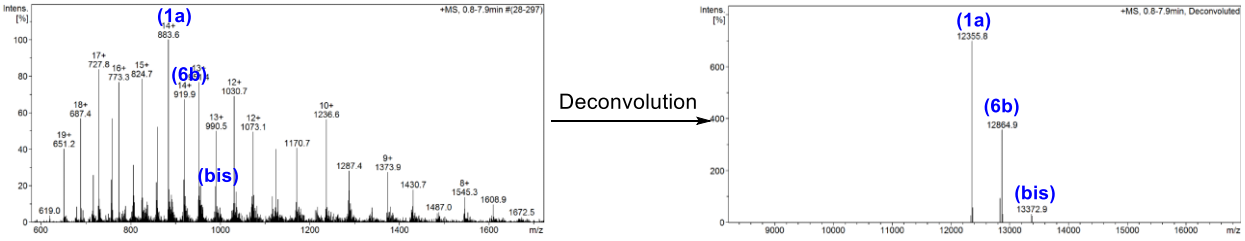


Figure S31. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4h** in 24 h (entry 7, Table S3).

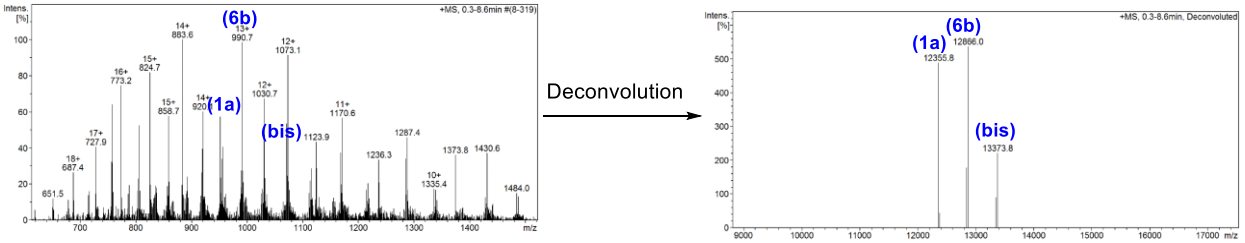


Figure S32. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4h** in 24 h (entry 8, Table S3).

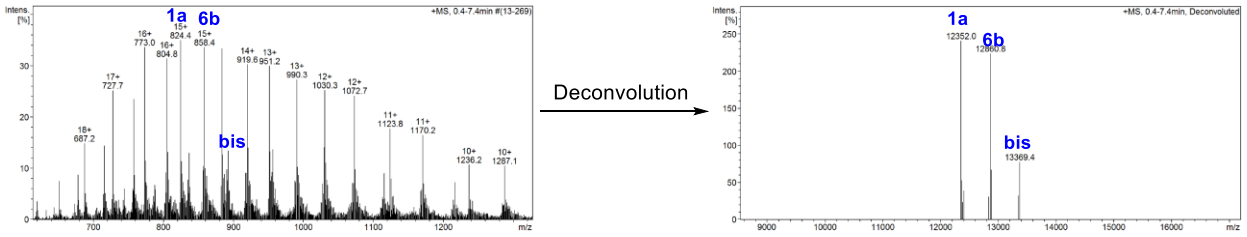


Figure S33. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4h** in 24 h (entry 9, Table S3).

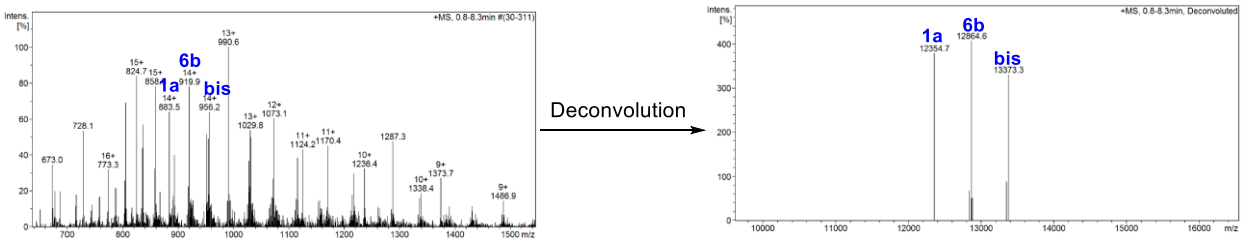
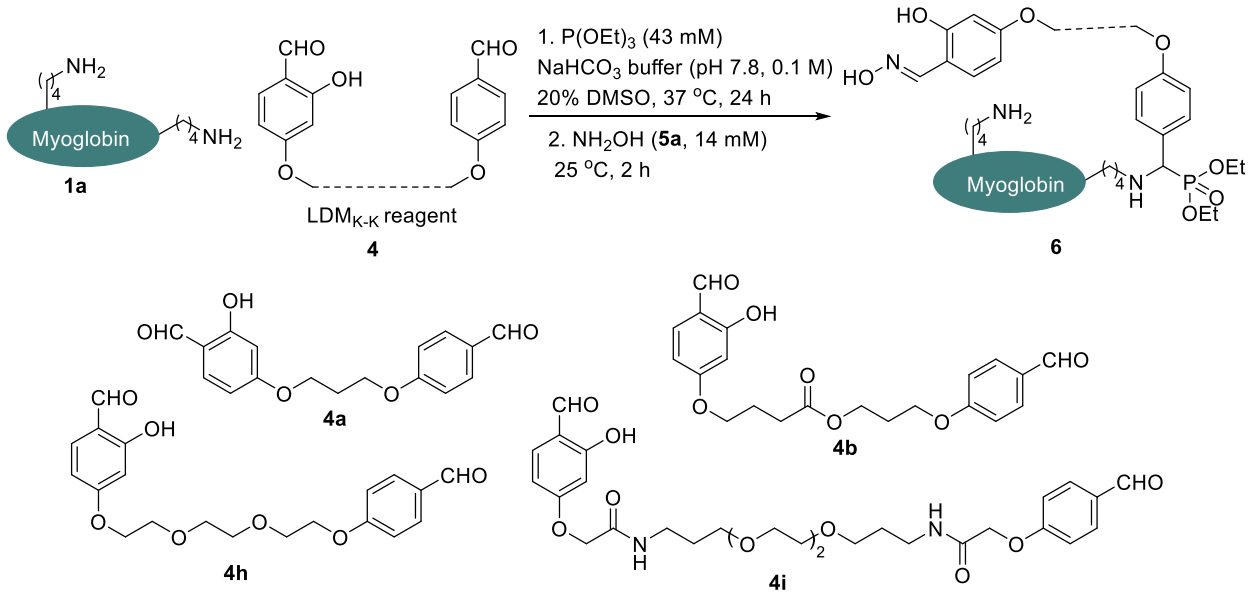


Figure S34. LC-ESI-MS spectra for site-selective labeling of cytochrome C (**1a**) enabled by LDM$_{K\text{-}K}$ reagent **4h** in 24 h (entry 10, Table S3).

Table S4. Reaction of myoglobin with diverse relative stoichiometry of $LDM_{K\text{-}K}$ reagents.



Entry	$LDM_{K\text{-}K}$ reagent	Equivalent(s)	Product (6)	% Conversion[a]
1	**4a**	1	**6n**	0
2		5		2 m
3		10		7 m
4		20		12 m
5	**4b**	1	**6c**	0
6		5		27 m
7		10		32 m
8		20		39 m
9	**4h**	1	**6o**	0
10		5		13 m
11		10		16 m
12		20		32 m
13	**4i**	1	**6d**	5 m
14		5		40 m
15		10		52 m, <5 b
16		20		65 m, 8 b

[a] % Conversions are estimated based on LC-ESI-MS. m = mono labeled, b = bis labeled. The data for entries 1-16 are given below in Figures S35-S50.

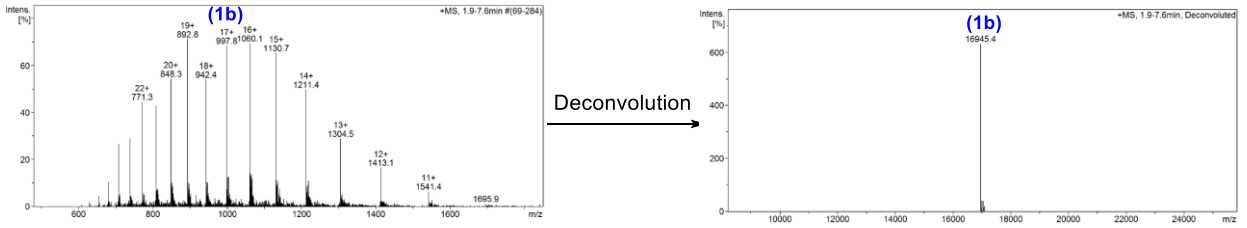


Figure S35. LC-ESI-MS spectra for the reaction between myoglobin (**1b**) and LDM$_{K\text{-}K}$ reagent **4a** in 24 h (entry 1, Table S4).

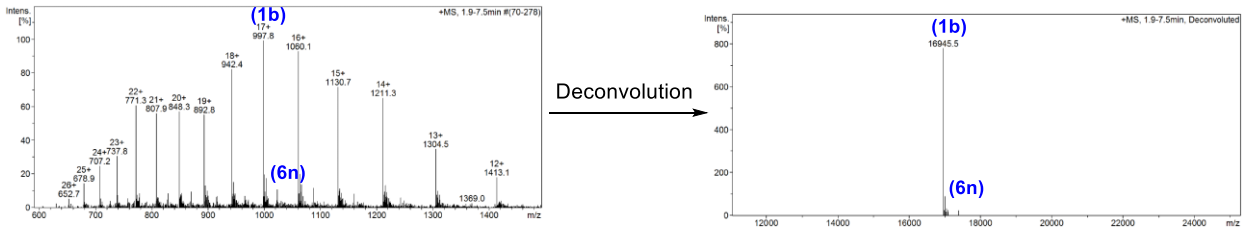


Figure S36. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by LDM$_{K\text{-}K}$ reagent **4a** in 24 h (entry 2, Table S4).

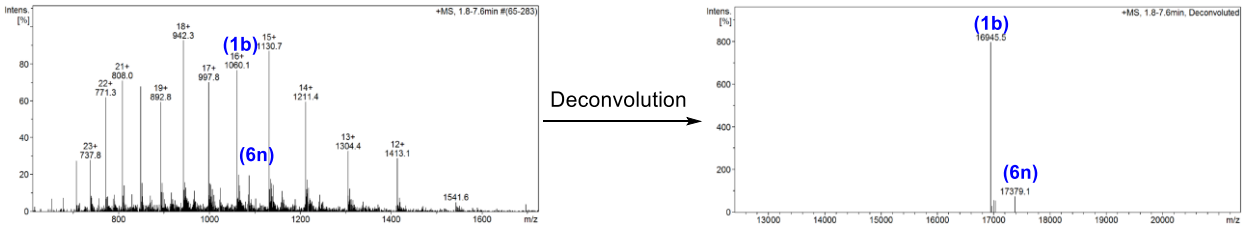


Figure S37. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by LDM$_{K\text{-}K}$ reagent **4a** in 24 h (entry 3, Table S4).

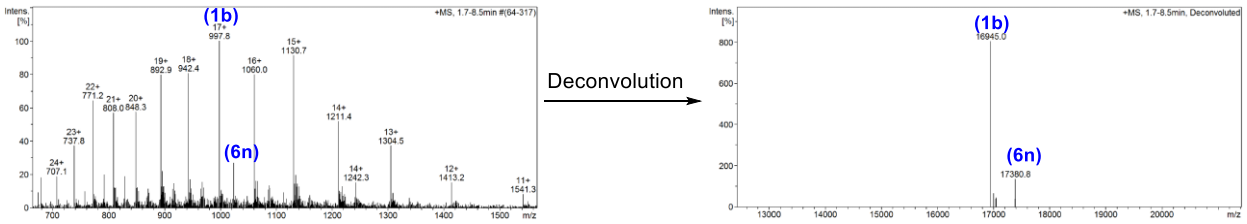


Figure S38. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by LDM$_{K\text{-}K}$ reagent **4a** in 24 h (entry 4, Table S4).

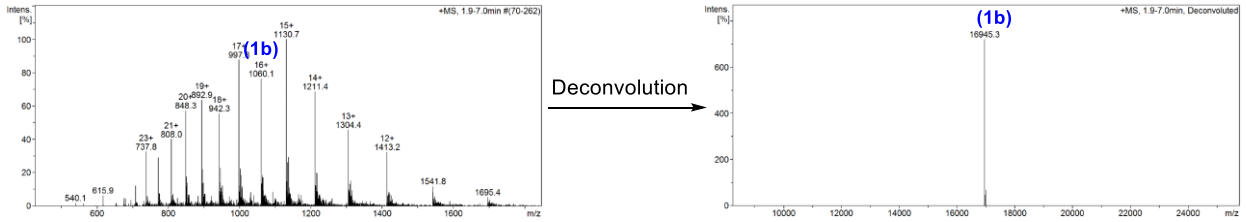


Figure S39. LC-ESI-MS spectra for the reaction between myoglobin (**1b**) and $LDM_{K\text{-}K}$ reagent **4b** in 24 h (entry 5, Table S4).

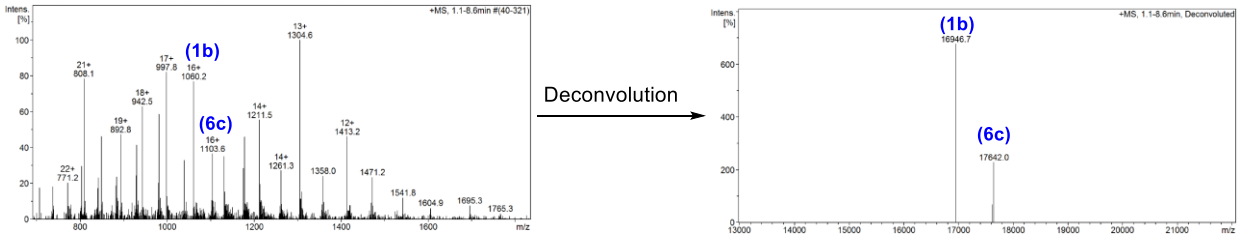


Figure S40. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4b** in 24 h (entry 6, Table S4).

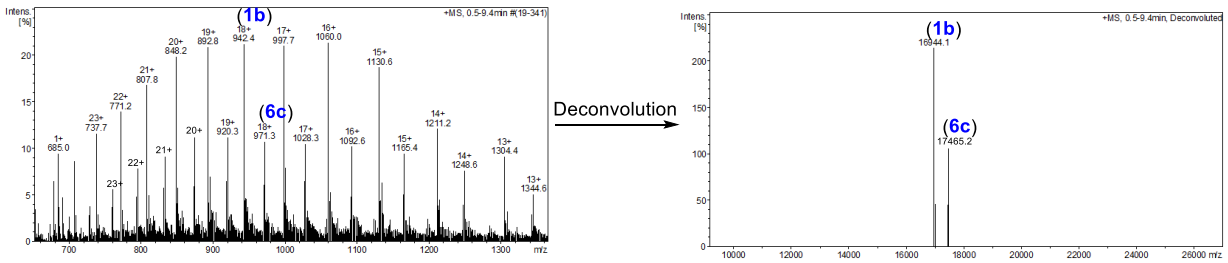


Figure S41. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4b** in 24 h (entry 7, Table S4).

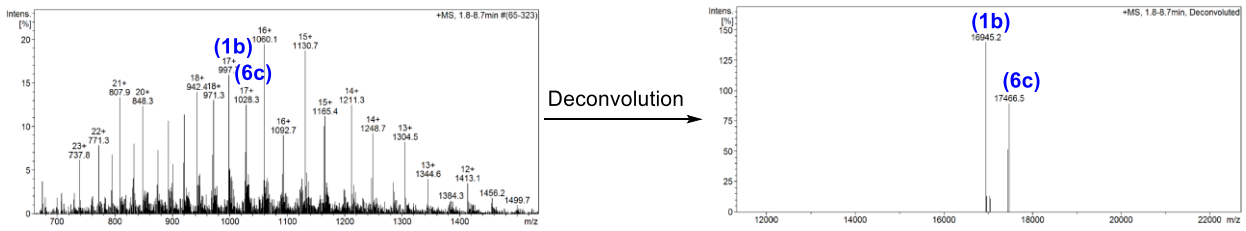


Figure S42. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4b** in 24 h (entry 8, Table S4).

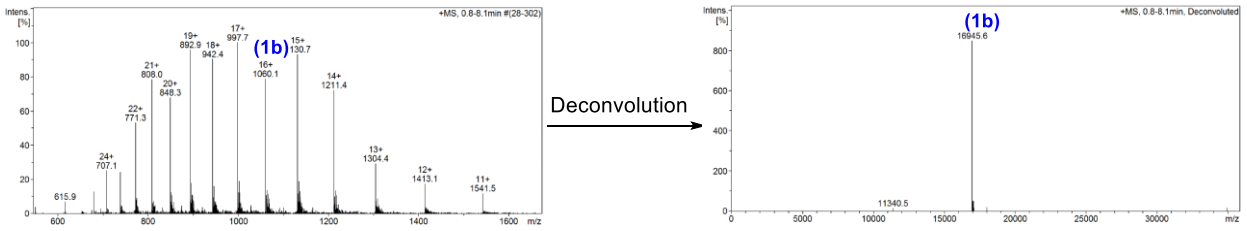


Figure S43. LC-ESI-MS spectra for the reaction between myoglobin (**1b**) and LDM$_{K\text{-}K}$ reagent **4h** in 24 h (entry 9, Table S4).

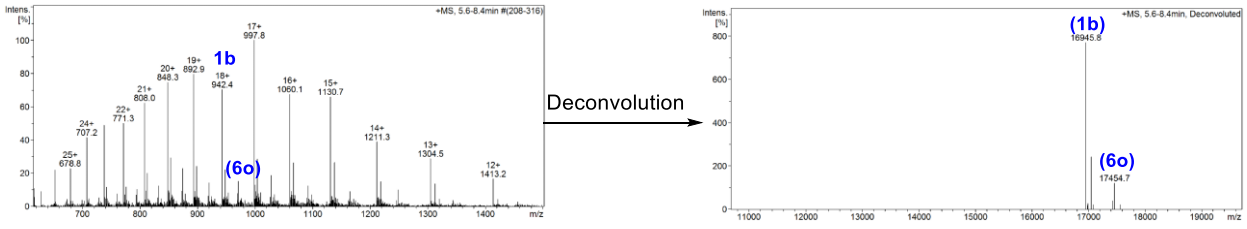


Figure S44. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by LDM$_{K\text{-}K}$ reagent **4h** in 24 h (entry 10, Table S4).

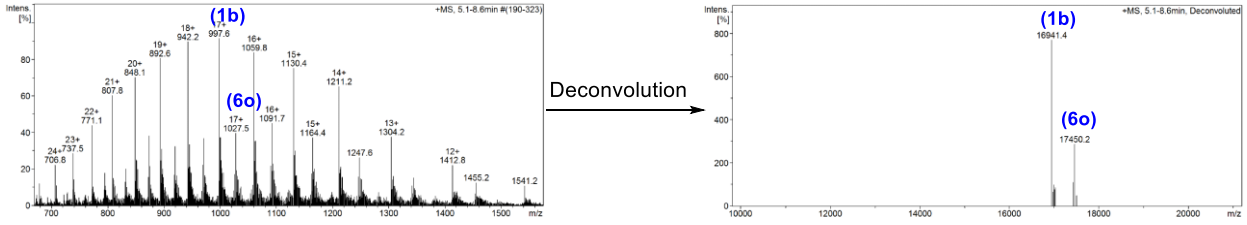


Figure S45. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by LDM$_{K\text{-}K}$ reagent **4h** in 24 h (entry 11, Table S4).

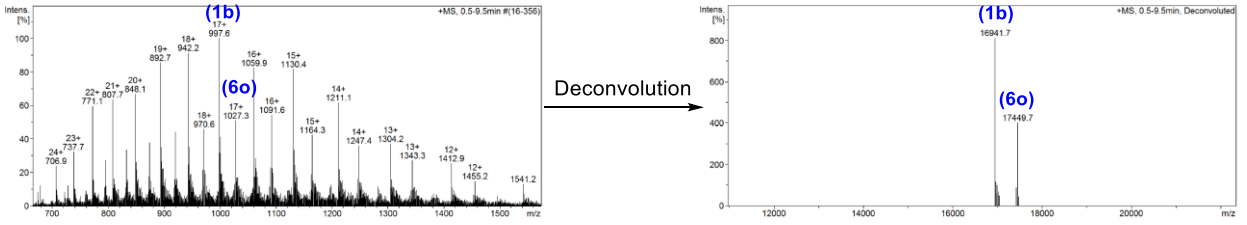


Figure S46. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by LDM$_{K\text{-}K}$ reagent **4h** in 24 h (entry 12, Table S4).

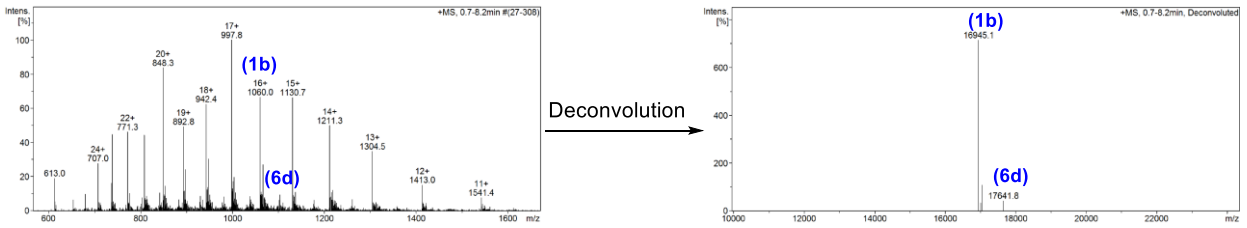


Figure S47. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 13, Table S4).

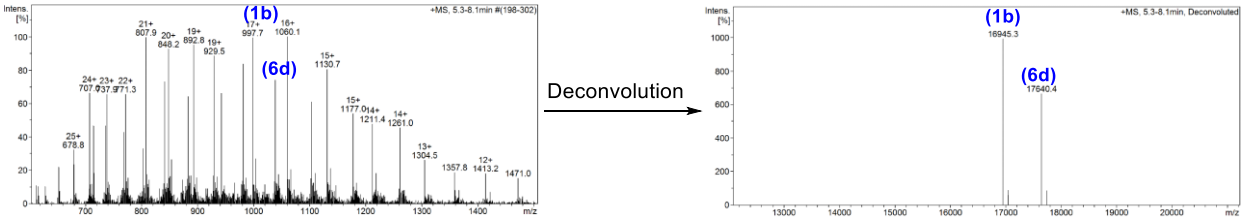


Figure S48. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 14, Table S4).

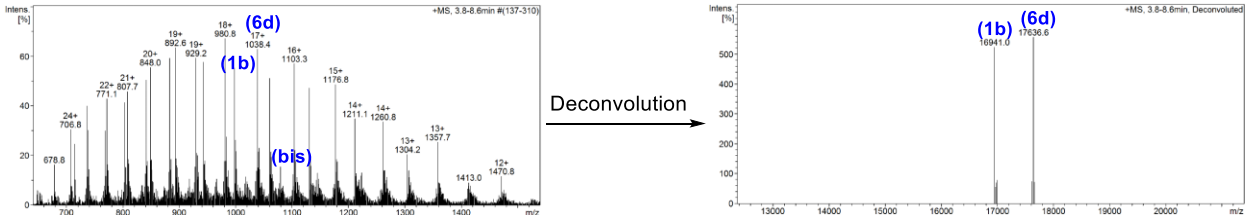


Figure S49. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 15, Table S4).

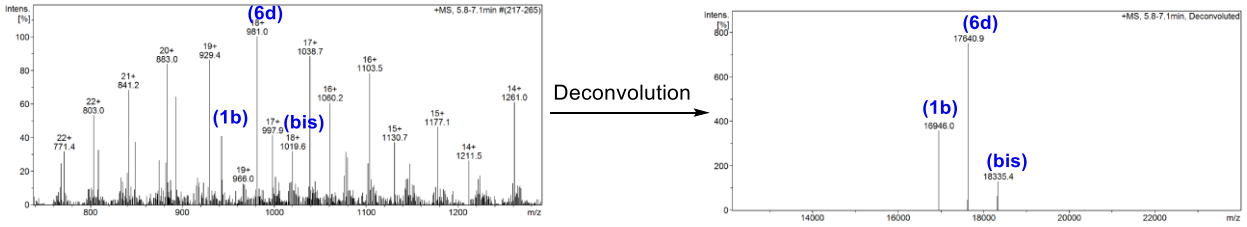
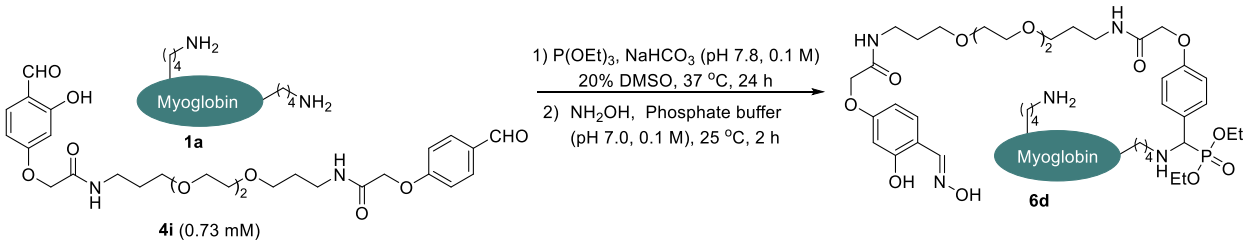


Figure S50. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** (entry 16, Table S4).

Table S5. Reaction of myoglobin with diverse relative stoichiometry of triethylphosphite.



Entry	$P(OEt)_3$ (equiv.)	% Conversion[a] (6d)	Site of modification
1	200	19 m	K42
2	400	32 m	K42
3	600	52 m, <5 b	K42
4	**800**	**72 m**	K42
5	1000	45 m	K42

[a] % Conversions are estimated based on LC-ESI-MS. m = mono-labeled, b = bis-labeled. The data for entries 1-5 are given below in Figures S51-S55.

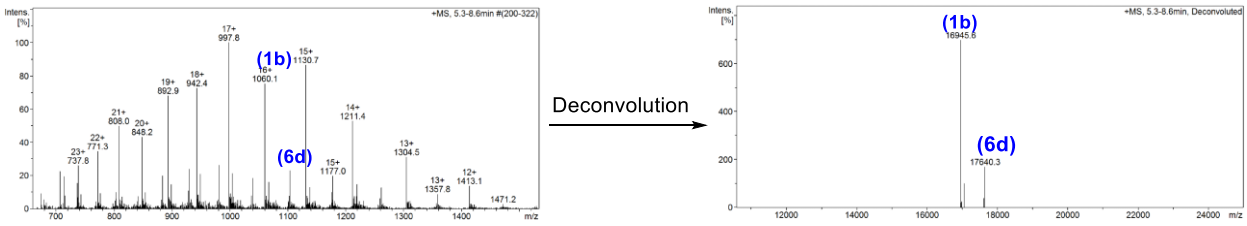


Figure S51. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 1, Table S5).

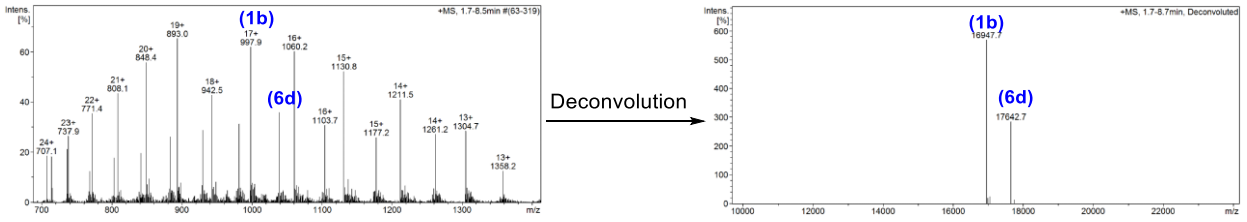


Figure S52. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 2, Table S5).

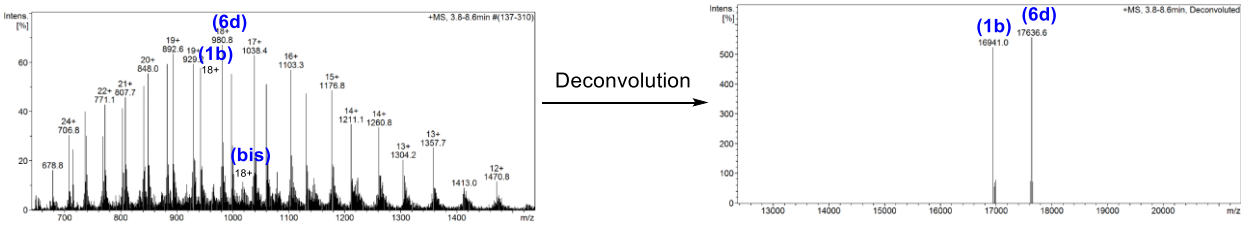


Figure S53. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 3, Table S5).

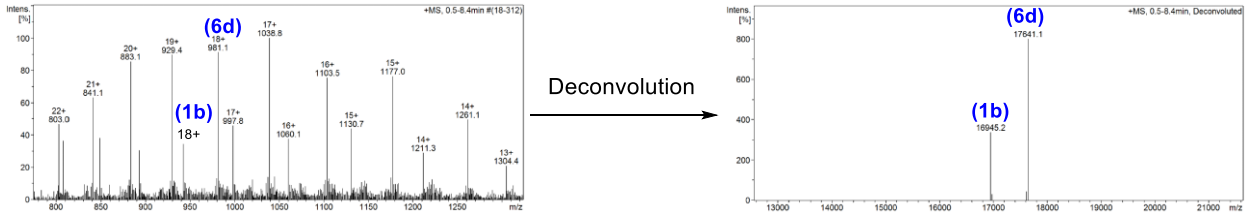


Figure S54. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 4, Table S5).

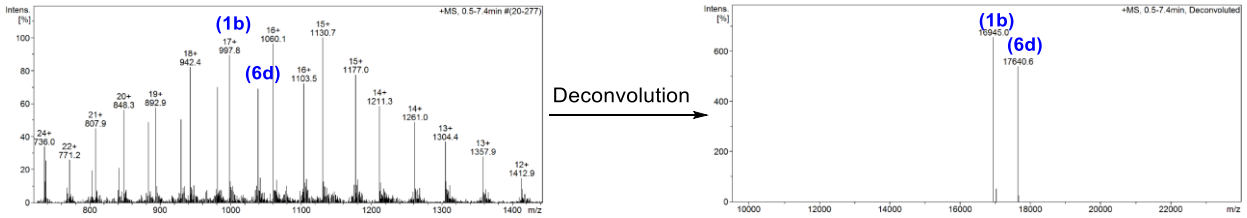


Figure S55. LC-ESI-MS spectra for site-selective labeling of myoglobin (**1b**) enabled by $LDM_{K\text{-}K}$ reagent **4i** in 24 h (entry 5, Table S5).

4.2 Structure and enzymatic activity of bioconjugate

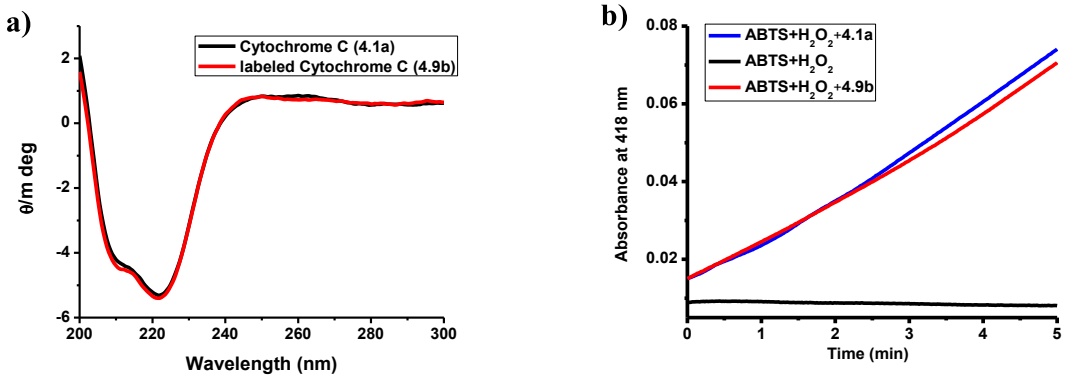
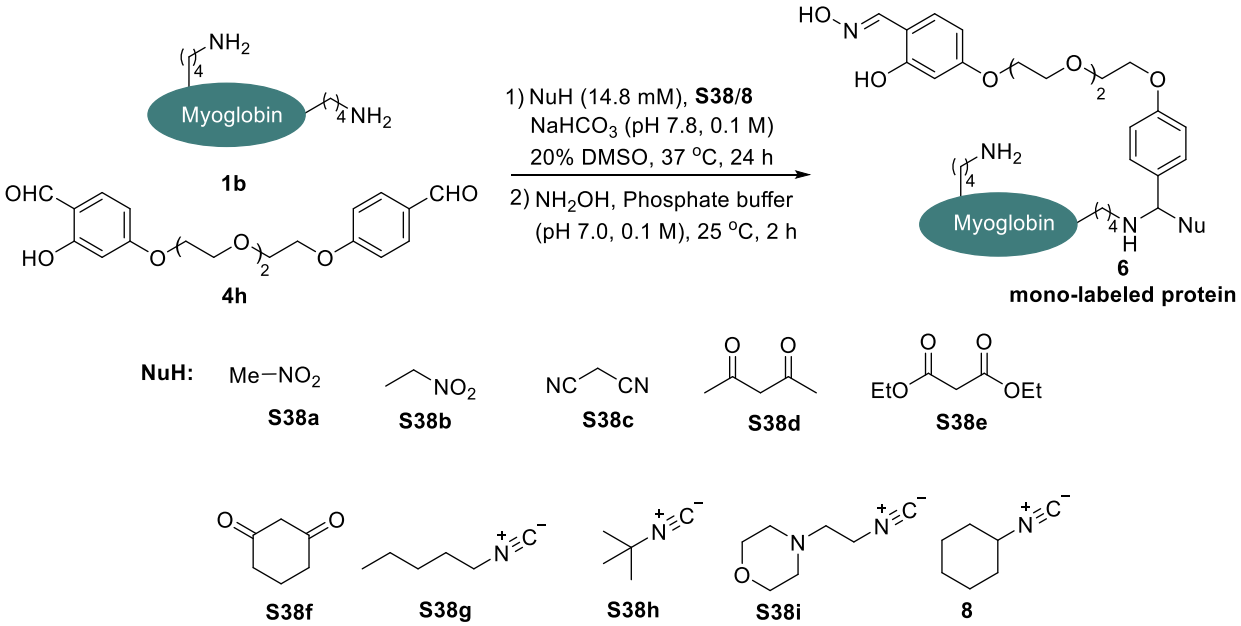


Figure S56. Effect of the bioconjugation on the secondary structure and activity of cytochrome C. (a) Circular dichroism spectra of cytochrome C (**1a**), K60-labeled cytochrome C (**6a**) at 5 μM. (b) Comparison of the enzymatic activity of native cytochrome C (**1a**) and K60-labeled cytochrome C (**6a**).

4.3 External nucleophiles for N-C-C bond formation

Table S6. C-centered external nucleophiles for capturing F_K^2-imine.



Entry	Nu	% Conversion (6)[a]
1	**S38a**	0 (**6p**)
2	**S38b**	0 (**6q**)
3	**S38c**	0 (**6r**)
4	**S38d**	0 (**6s**)
5	**S38e**	0 (**6t**)
6	**S38f**	0 (**6u**)
7	**S38g**	8 (**6v**)
8	**S38h**	0 (**6w**)
9	**S38i**	10 (**6x**)
10	**8**	**25 (6f)**

[a] % Conversion is based on LC-ESI-MS data. No higher order labeling is observed under the reaction conditions. For data, see Figure S70.

5. Spectral analysis

5.1 Mass data

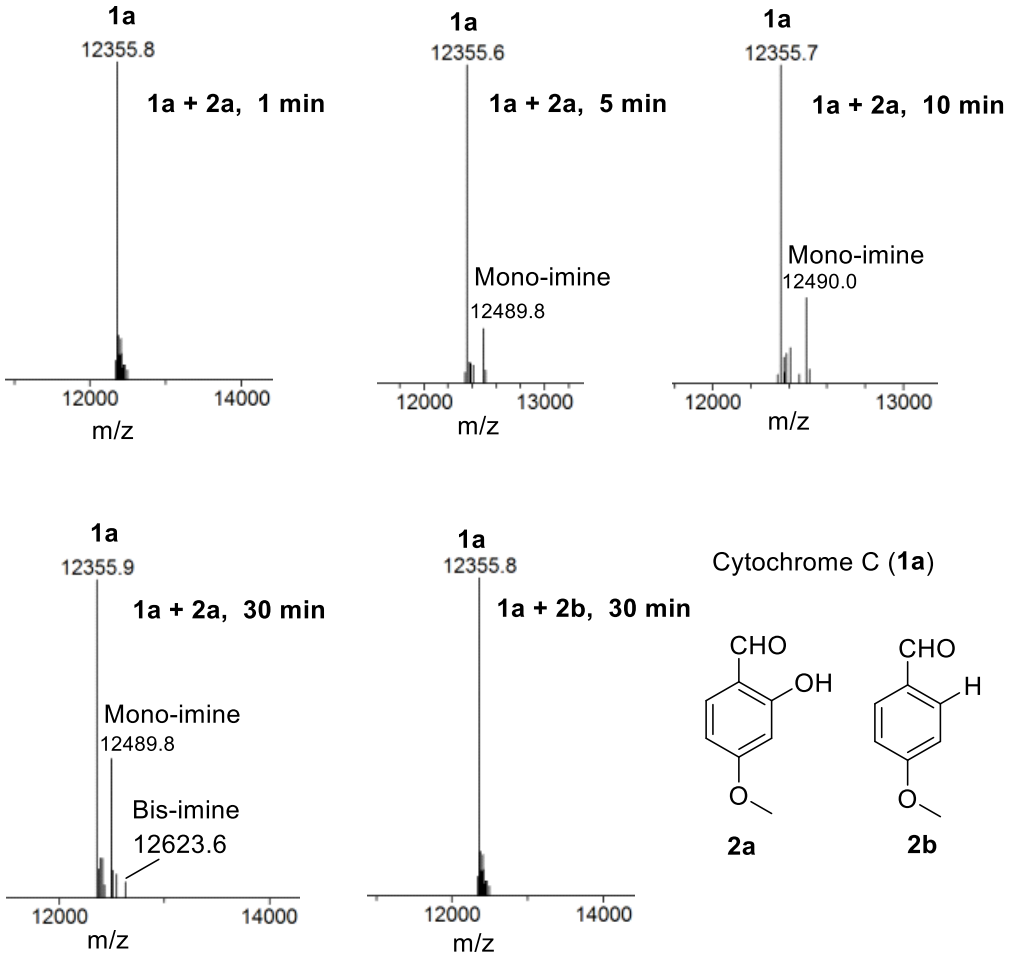


Figure S57. Comparative assessment of imine formation with cytochrome C (**1a**) and reagents **2a** and **2b** by LC-ESI-MS. Similar experiments on MALDI-ToF-ToF indicates the same order, but with higher extent of labeling and heterogeneity.[5]

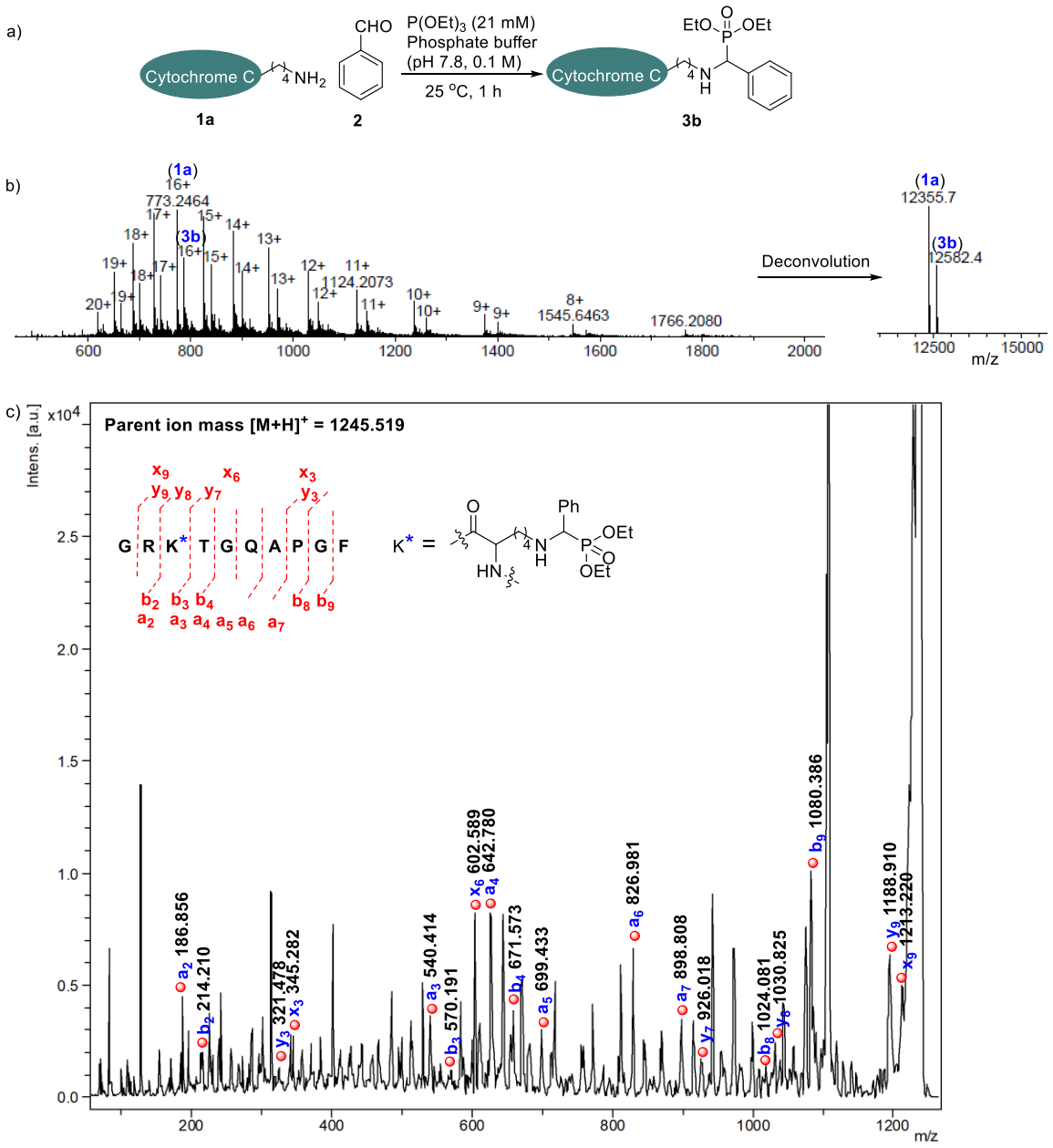
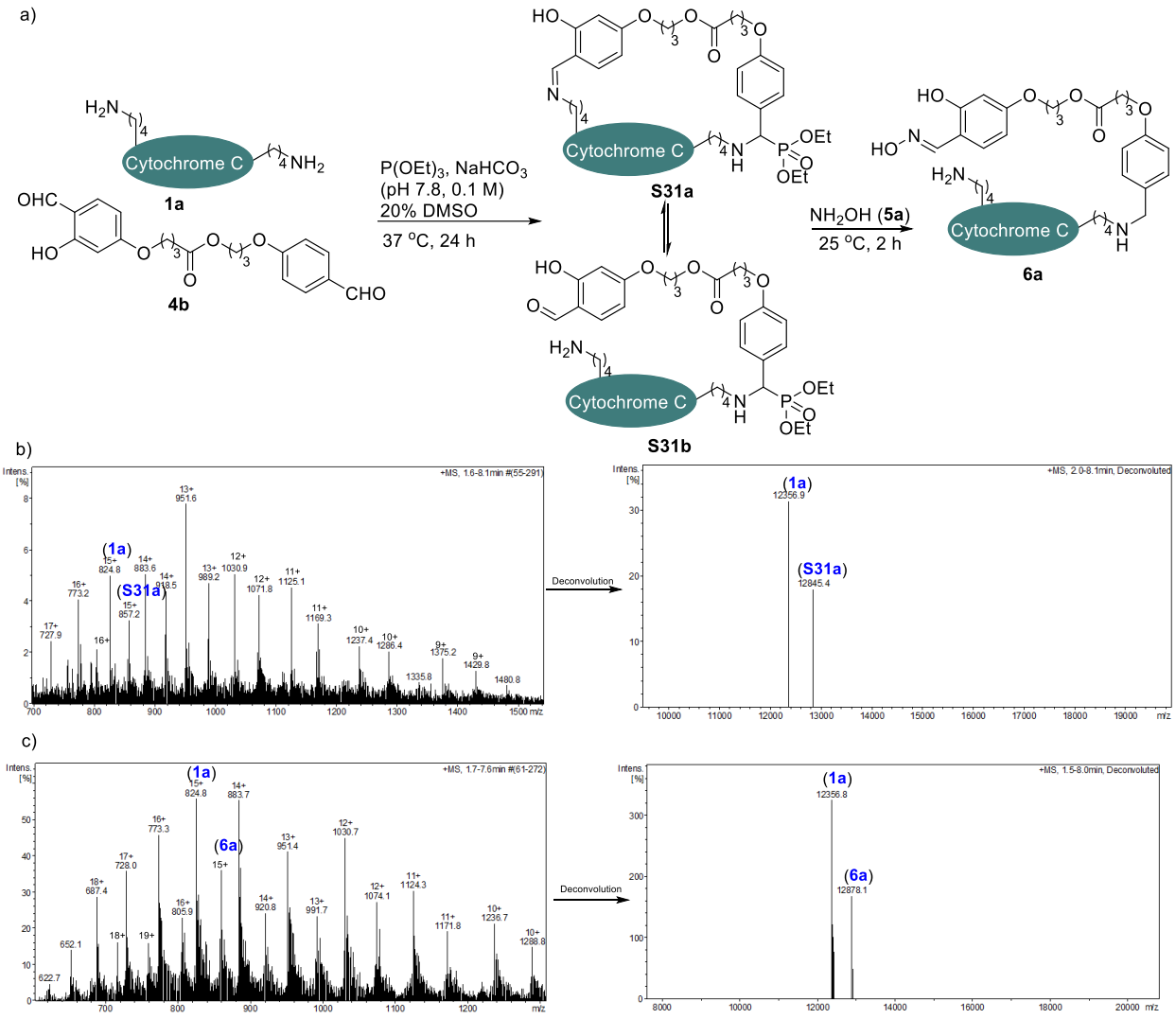


Figure S58. (a) Site-selective modification of cytochrome C with reagent **2**. (b) LC-ESI-MS spectra of reaction mixture containing cytochrome C (**1a**) and mono-labeled cytochrome C (**3b**). (c) MALDI-ToF MS-MS spectrum of labeled GRKTGQAPGF (G37-F46) confirms the site of modification (K39).


a)
1a
4b
P(OEt)3, NaHCO3
(pH 7.8, 0.1 M)
20% DMSO
37 °C, 24 h
S31a
S31b
NH2OH (5a)
25 °C, 2 h
6a
b)
(1a)
(S31a)
Deconvolution
(1a)
12356.9
(S31a)
12845.4
c)
(1a)
(6a)
Deconvolution
(1a)
12356.8
(6a)
12878.1

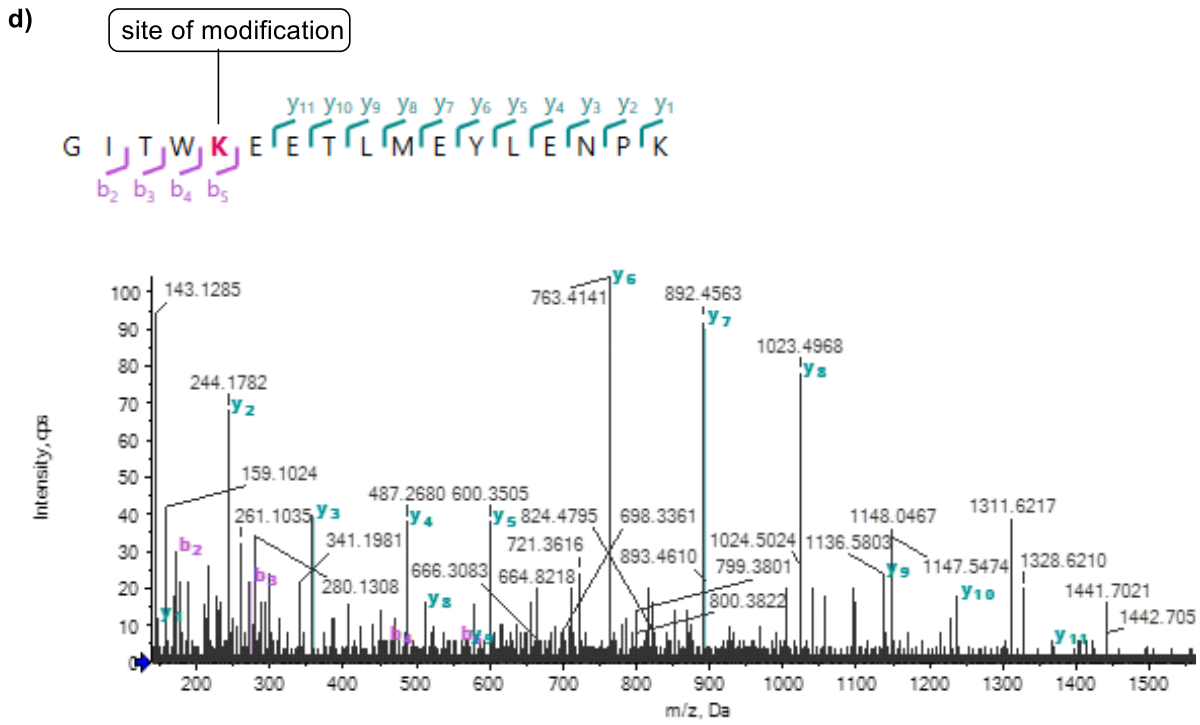


Figure S59. (a) Site-selective labeling of cytochrome C enabled by LDM$_{K\text{-}K}$ reagent **4b**. (b) LC-ESI-MS spectra for cytochrome C (**1a**) and cyclized intermediate (**S31a**). (c) LC-ESI-MS spectra for cytochrome C (**1a**) and mono-labeled cytochrome C (**6a)**. (d) MS-MS spectrum of labeled GITWKEETLMEYLENPK (G56-K72, *m/z* 868.0 [{(M+3)/3}+H]$^+$) confirms the site of modification (K60).

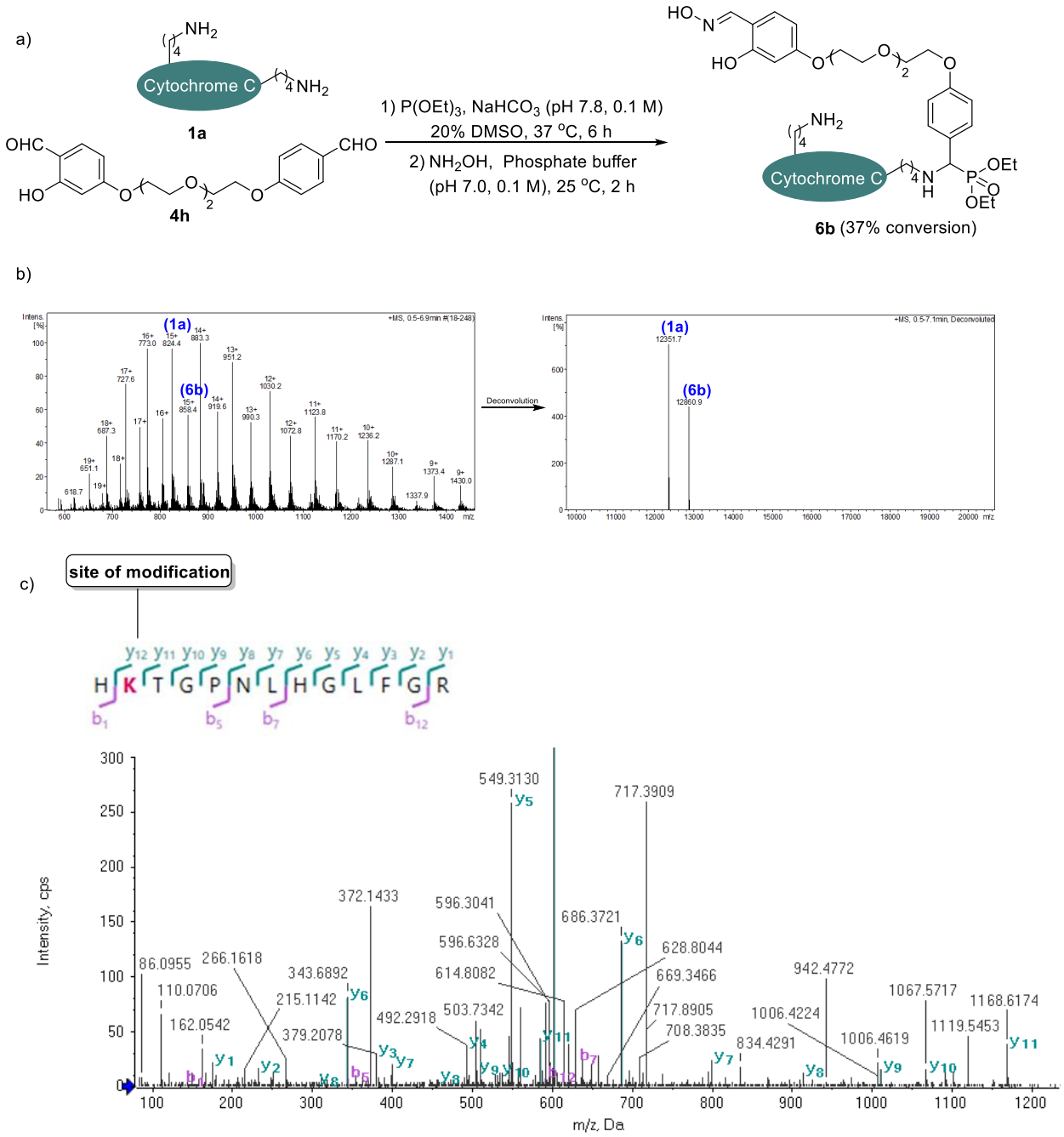


Figure S60. (a) Site-selective labeling of cytochrome C enabled by LDM$_{K\text{-}K}$ reagent **4h**. (b) LC-ESI-MS spectra for cytochrome C (**1a**) and mono-labeled cytochrome C (**6b**). (c) MS-MS spectrum of labeled HKTGPNLHGLFGR (H26-R38, *m/z* 648.4 [{(M+3)/3}+H]$^+$) confirms the site of modification (K27).

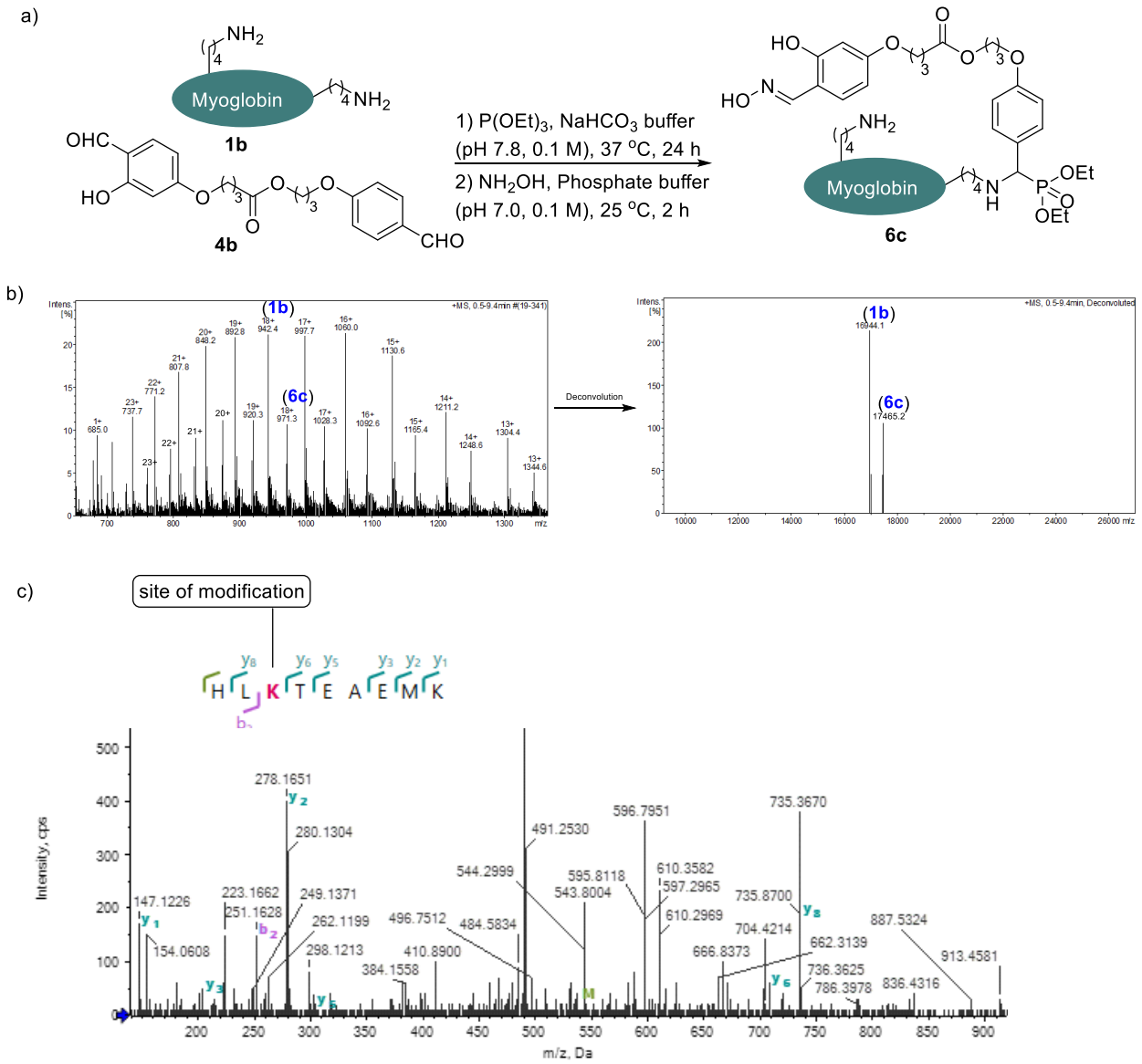


Figure S61. (a) Site-selective labeling of myoglobin enabled by $LDM_{K\text{-}K}$ reagent **4b**. (b) LC-ESI-MS spectra for myoglobin (**1b**) and single-site-labeled myoglobin (**6c**). (c) MS-MS spectrum of labeled HLKTEAEMK (H48-K56, *m/z* 536.5 [{(M+3)/3}+H]$^+$) confirms the site of modification (K50).

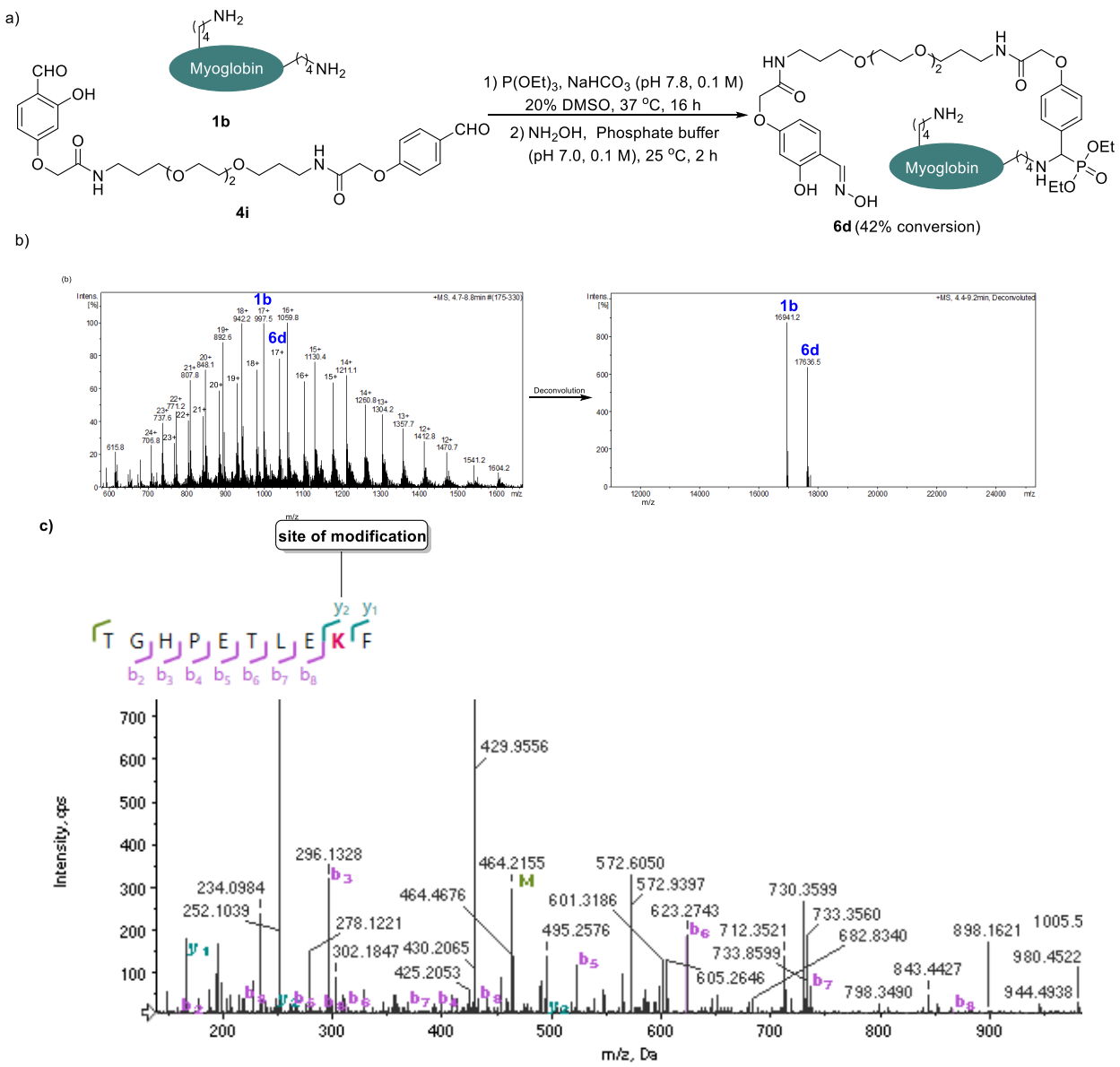


Figure S62. (a) Site-selective labeling of myoglobin enabled by $LDM_{K\text{-}K}$ reagent **4i**. (b) LC-ESI-MS spectra of the reaction mixture containing myoglobin (**1b**) and mono labeled myoglobin (**6d**). (c) MS-MS spectrum of labeled TGHPETLEKF (T34-F43, *m/z* 464.2 [{(M+4)/4}+H]$^+$) confirms the site of modification (K42).

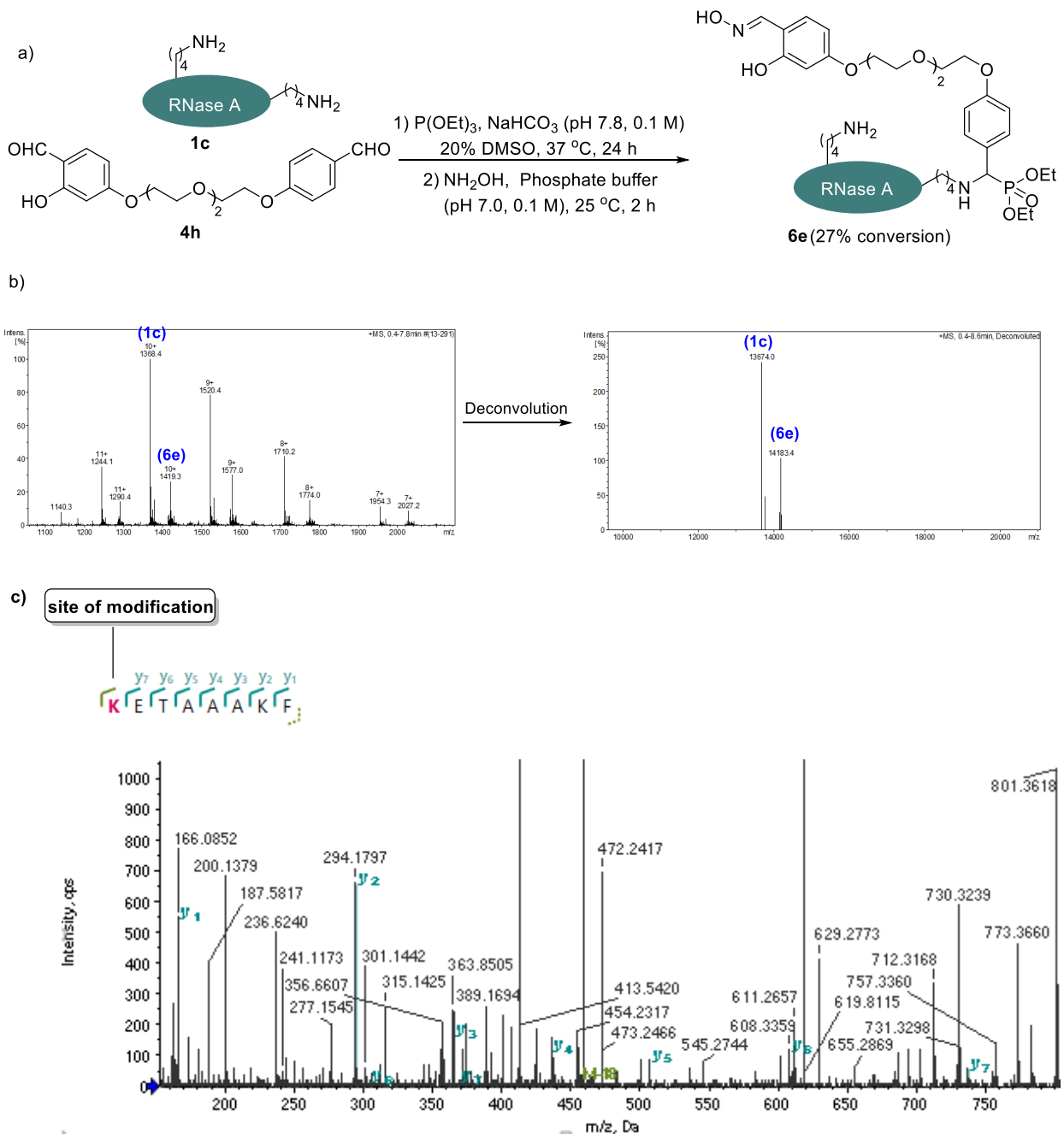


Figure S63. (a) Site-selective labeling of RNase A enabled by LDM$_{K\text{-}K}$ reagent **4h**. (b) LC-ESI-MS spectra of the reaction mixture containing RNase A (**1c**) and mono-labeled RNase A (**6e**). (c) MS-MS spectrum of labeled KETAAKF (K1-F8, *m/z* 458.8 [{(M+3)/3}+H]$^+$) confirms the site of modification (K1).

5.2 Late-stage modification

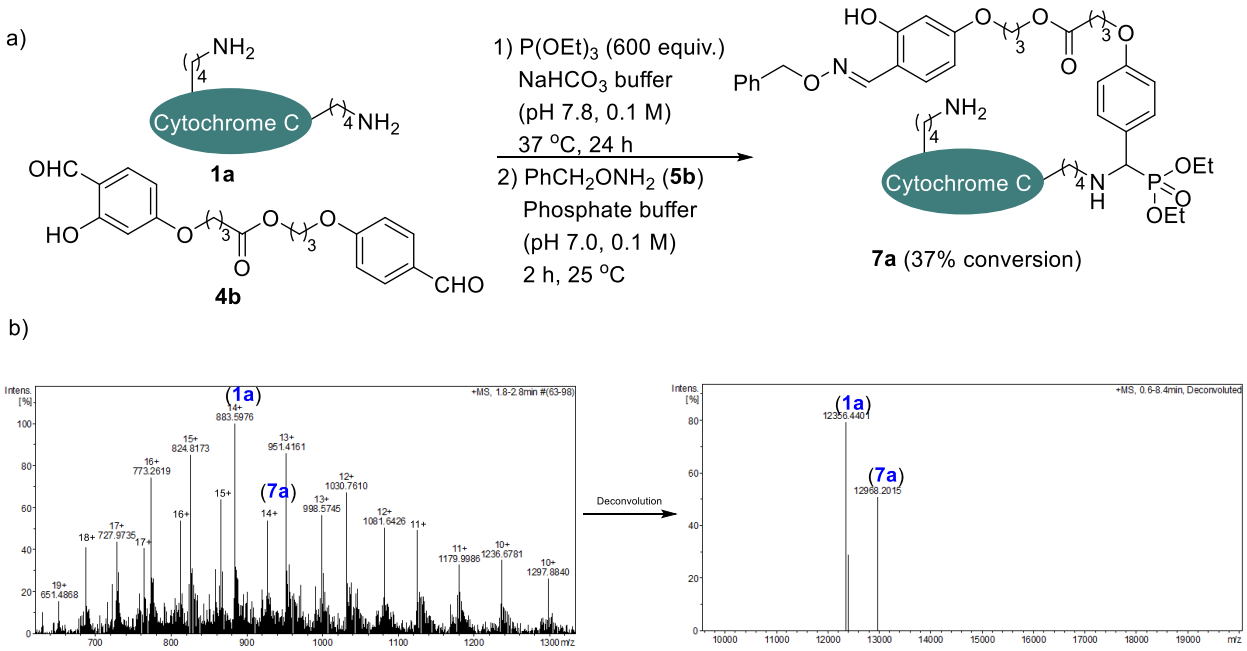


Figure S64. (a) Site-selective labeling of cytochrome C (**1a**) enabled by $LDM_{K\text{-}K}$ reagent **4b** and late-stage derivatization with **5b**. (b) LC-ESI-MS spectra of the reaction mixture containing cytochrome C (**1a**) and mono-labeled cytochrome C (**7a**).

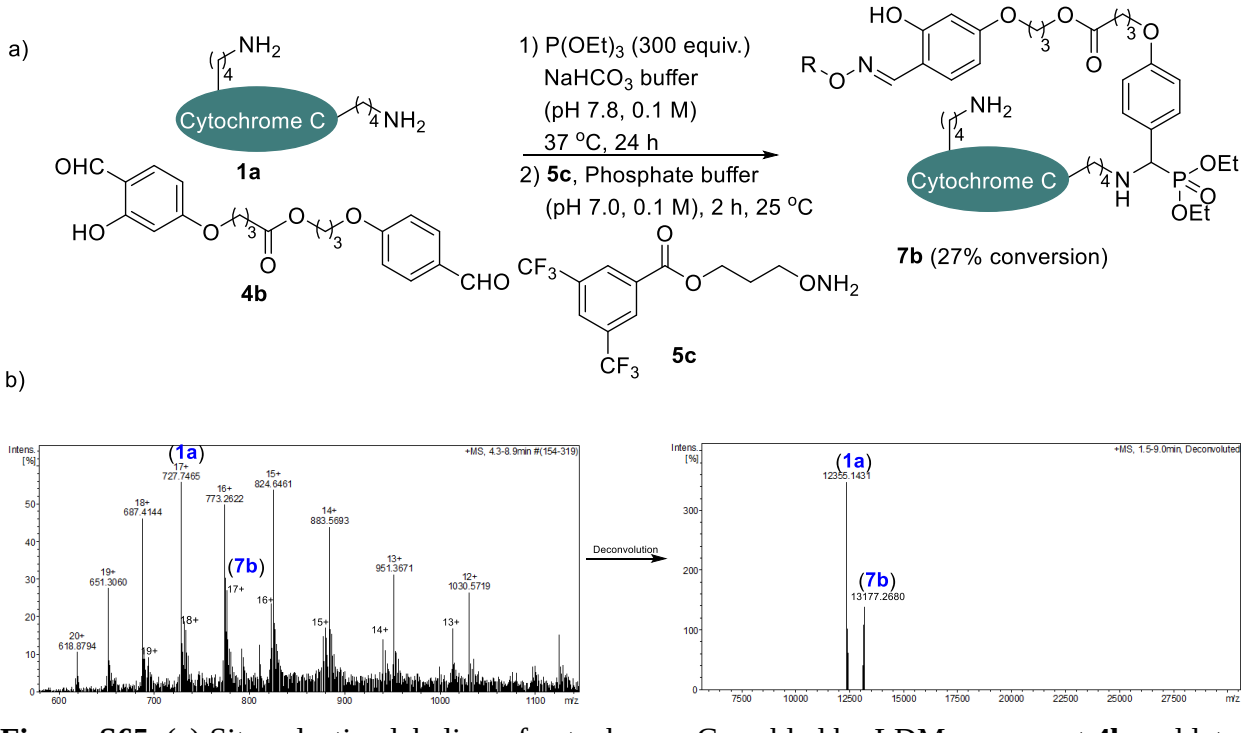


Figure S65. (a) Site-selective labeling of cytochrome C enabled by $LDM_{K\text{-}K}$ reagent **4b** and late-stage derivatization with **5c**. (b) LC-ESI-MS spectra of the reaction mixture containing cytochrome C (**1a**) and mono-labeled cytochrome C (**7b**).

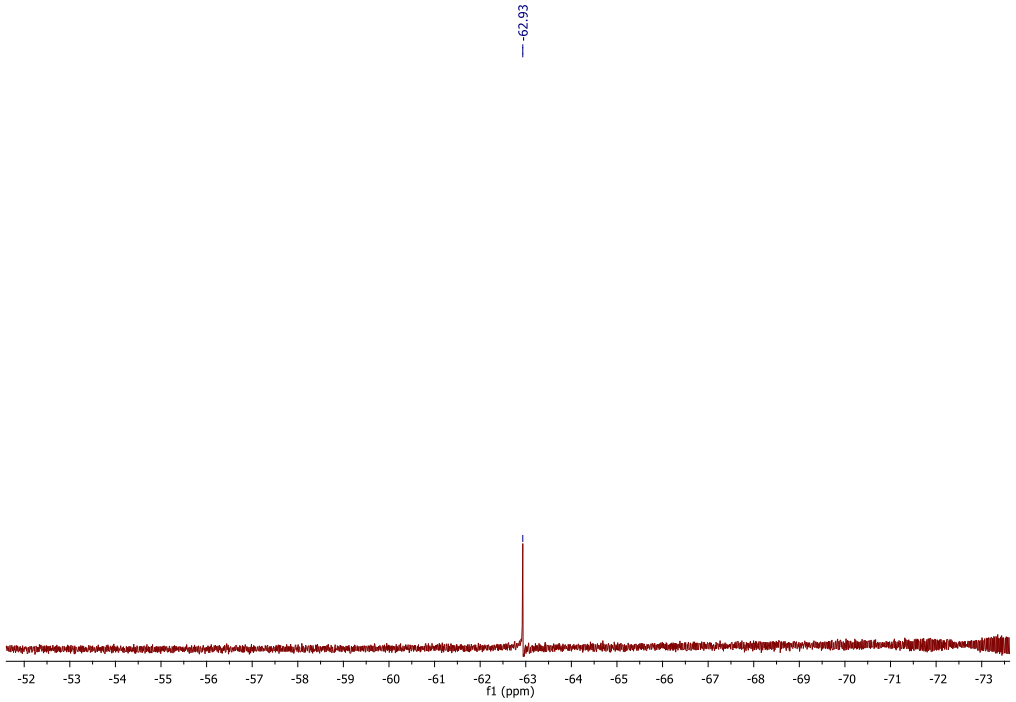


Figure S66. ^{19}F-NMR probe labeled cytochrome C (**7b**, shows a sharp signal at -62.9 ppm by ^{19}F NMR spectroscopy [TFA (0.2 mM) as internal standard at -75.45 ppm, phosphate buffer (0.1 M, pH 7.0):D_2O (9:1)].

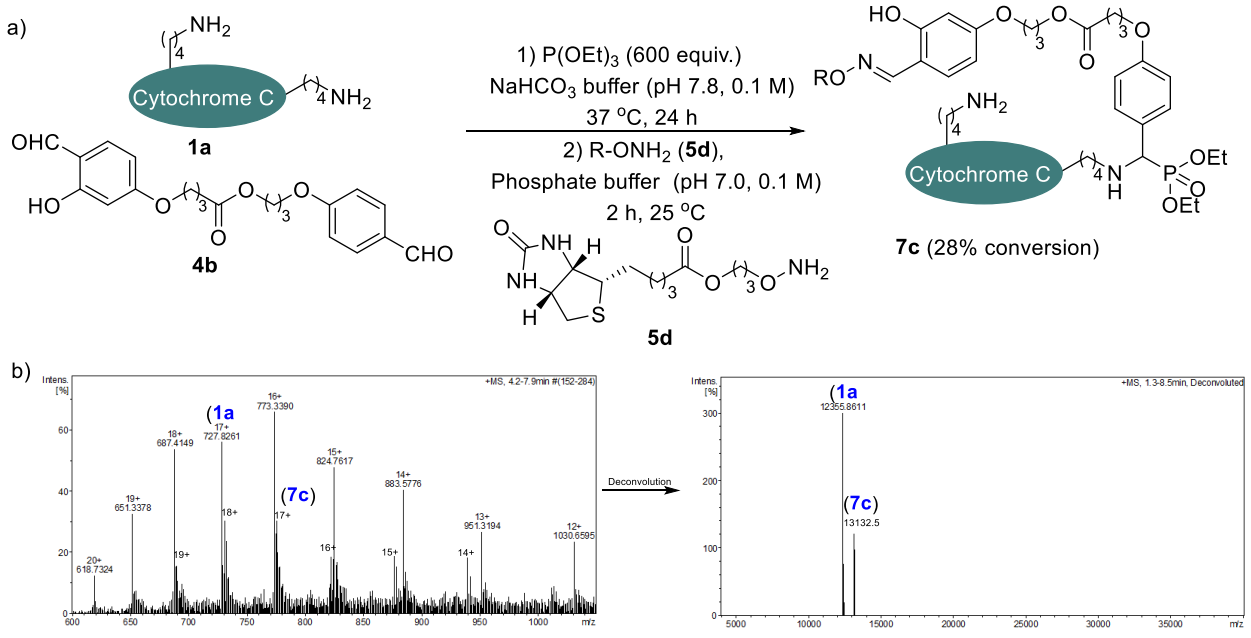


Figure S67. (a) Site-selective labeling of cytochrome C enabled by LDM$_{K\text{-}K}$ reagent **4b** and late-stage derivatization with **5d**. (b) LC-ESI-MS spectra of the reaction mixture containing cytochrome C (**1a**) and mono-labeled cytochrome C (**7c**).

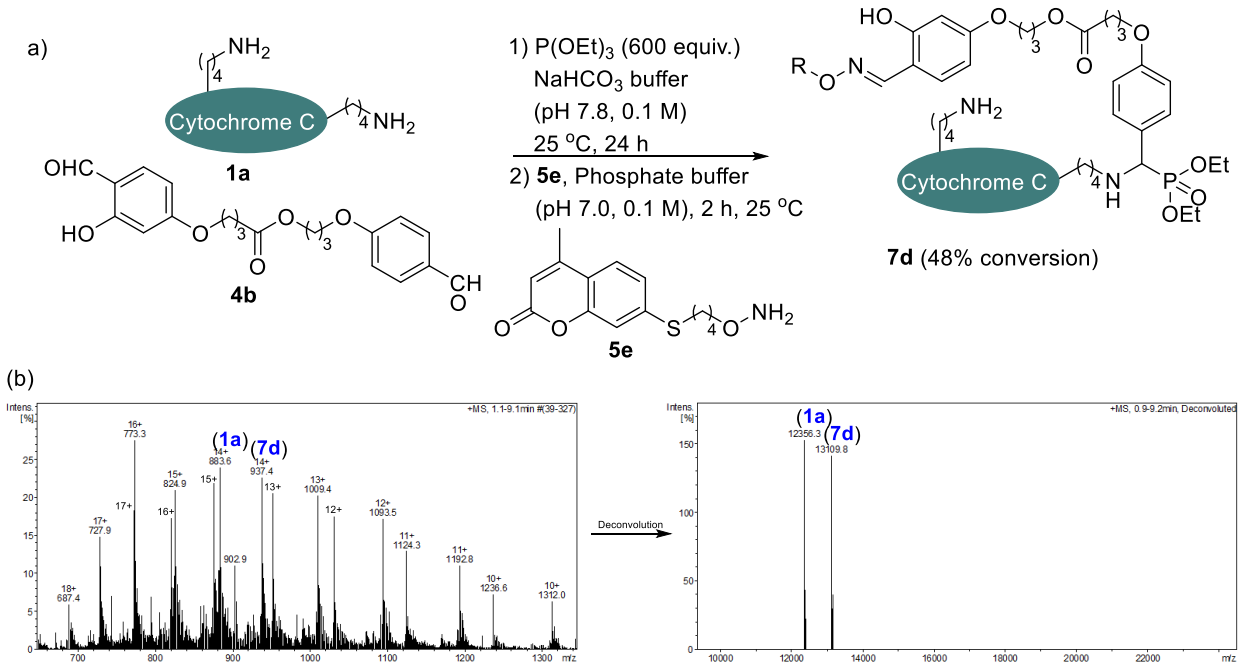


Figure S68. (a) Site-selective labeling of cytochrome C enabled by LDM$_{K\text{-}K}$ reagent **4b** and late-stage derivatization with **5e**. (b) LC-ESI-MS spectra of the reaction mixture containing cytochrome C (**1a**) and mono-labeled cytochrome C (**7d**).

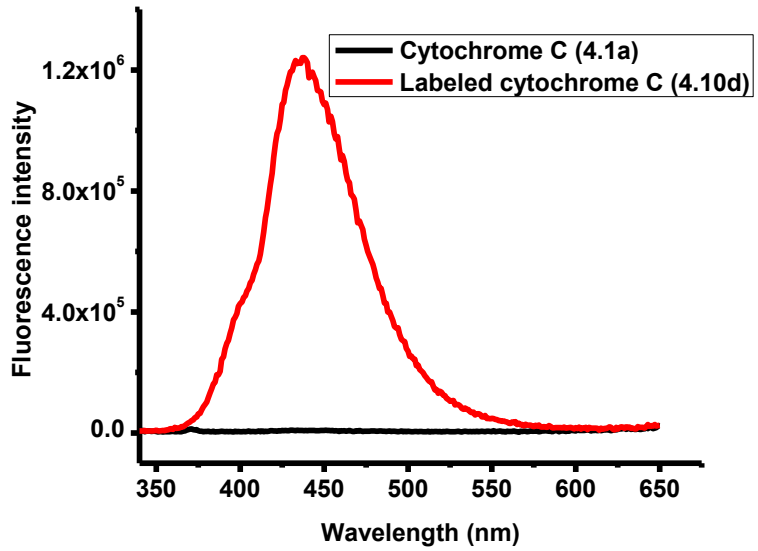
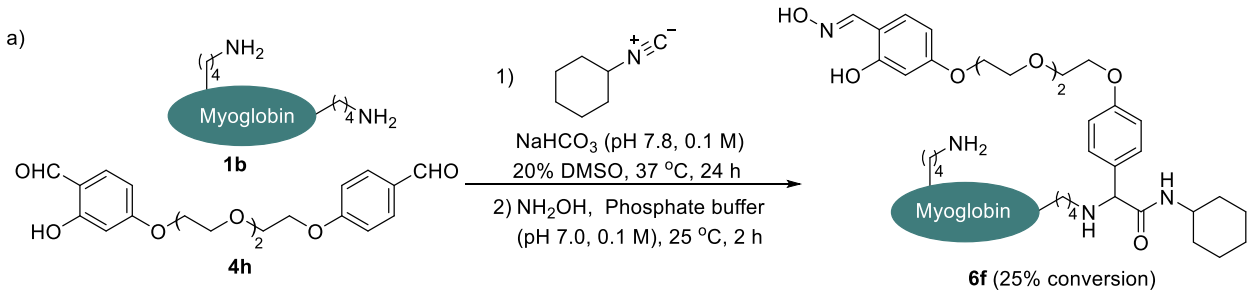
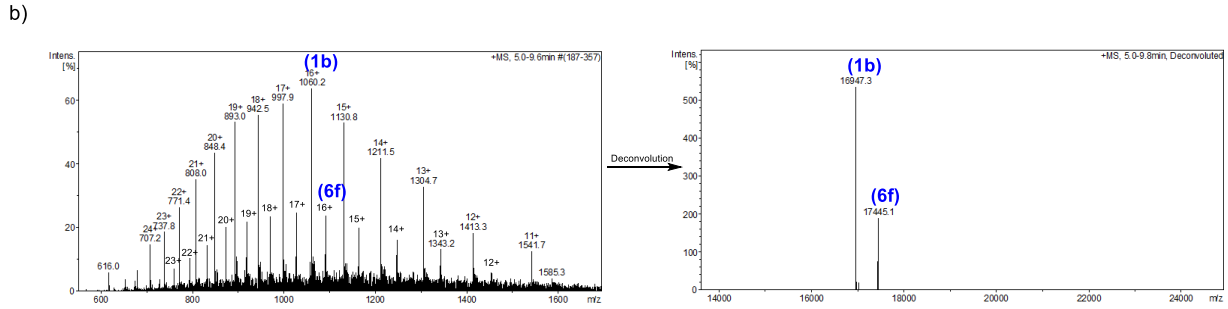


Figure S69. Steady-state fluorescence spectra of cytochrome C (**1a**) and coumarin tagged cytochrome C (**7d**). In phosphate buffer (0.1 M, pH 7.0), labeled cytochrome C (**7d**) exhibits emission band peaked at 430 nm (excitation at 333 nm).

5.3 N-C-C bond formation





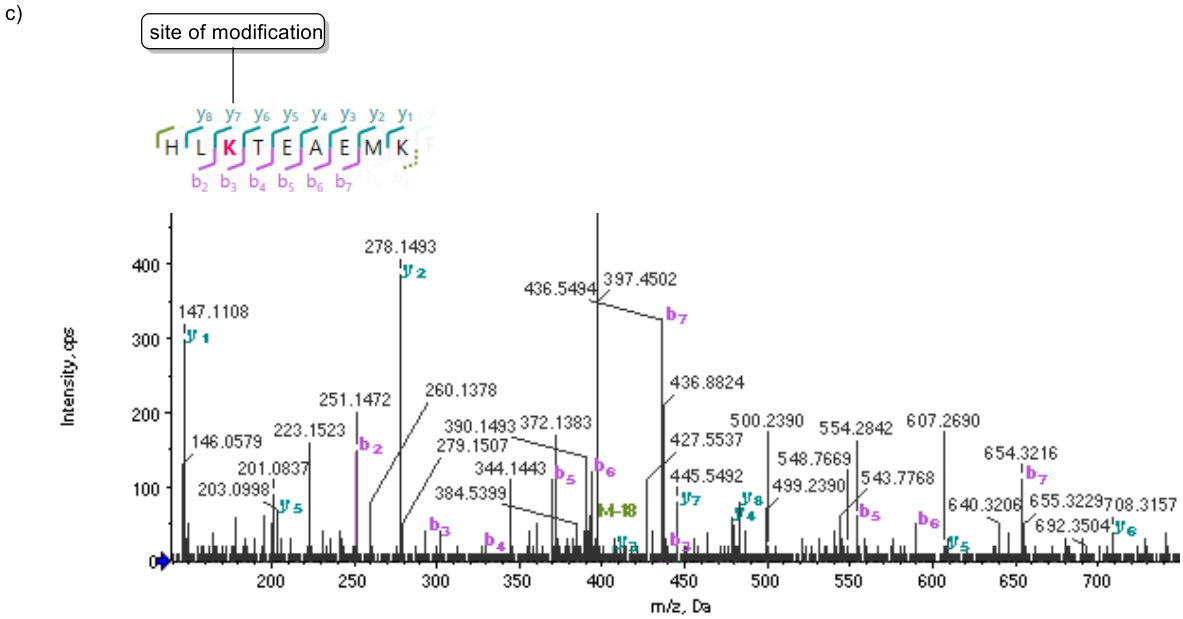
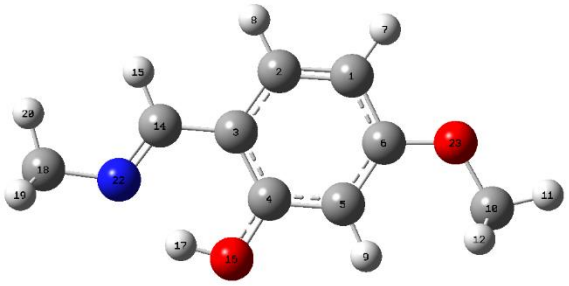


Figure S70. (a) Site-selective labeling of myoglobin enabled by $LDM_{K\text{-}K}$ reagent **4h**. (b) LC-ESI-MS spectra of the reaction mixture containing myoglobin (**1b**) and mono-labeled myoglobin (**6f**). (c) MS-MS spectrum of labeled HLKTEAEMK (H48-K56, m/z 397.20 [{(M+3)/3}+H]$^+$) confirms the site of modification (K50).

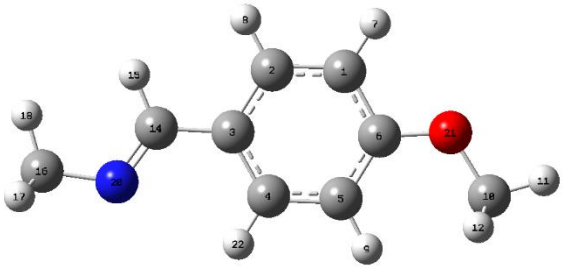
6. DFT calculations: Cartesian coordinates and method

Method for calculations: opt freq b3lyp/6-31+g(d,p) pop=(nbo,savenbo)



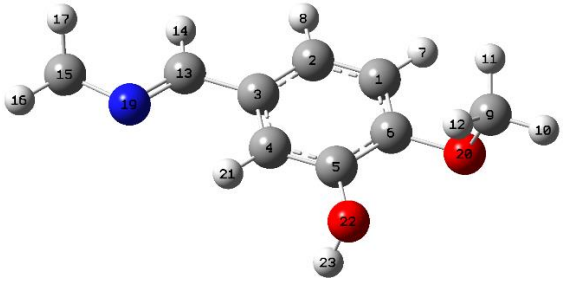
Imine (S39a) of 2-hydroxy-4-methoxybenzaldehyde

1	6	-1.472685	-1.409188	-0.234489
2	6	-0.097511	-1.601243	-0.098540
3	6	0.760307	-0.502058	-0.064476
4	6	0.243138	0.789887	-0.167580
5	6	-1.131673	0.981799	-0.303972
6	6	-1.989710	-0.117787	-0.337037
7	1	-2.148872	-2.275917	-0.260876
8	1	0.309715	-2.619453	-0.016943
9	1	-1.539570	1.999878	-0.385261
10	6	-4.160684	0.225877	0.856867
11	1	-5.261711	0.379901	0.748277
12	1	-3.995780	-0.694753	1.467870
13	1	-3.740233	1.096290	1.416934
14	6	2.278180	-0.713903	0.086468
15	1	2.587435	-1.765187	0.159580
16	8	1.122922	1.916681	-0.132699
17	1	1.951086	1.864282	0.350000
18	6	4.566934	-0.198357	0.232464
19	1	5.294796	0.648365	0.264905
20	1	4.696787	-0.805817	1.161016
21	1	4.820230	-0.842729	-0.644292
22	7	3.160811	0.313704	0.125862
23	8	-3.507589	0.094449	-0.487359



Imine (S39b) of 4-methoxybenzaldehyde

1	6	1.299937	1.271177	-0.207652
2	6	-0.081673	1.406865	-0.069036
3	6	-0.896797	0.275150	-0.065044
4	6	-0.330469	-0.992932	-0.200903
5	6	1.050772	-1.128479	-0.339957
6	6	1.866098	0.003618	-0.342942
7	1	1.942457	2.163535	-0.210328
8	1	-0.527642	2.406272	0.038375
9	1	1.497407	-2.127740	-0.447058
10	6	4.047843	-0.287767	0.845220
11	1	5.154032	-0.396667	0.734489
12	1	3.847401	0.609623	1.479720
13	1	3.660655	-1.187947	1.381724
14	6	-2.421768	0.424756	0.088839
15	1	-2.771100	1.461172	0.188794
16	6	-4.689200	-0.181614	0.217546
17	1	-5.384129	-1.056122	0.226709
18	1	-4.842837	0.395998	1.161387
19	1	-4.966399	0.475251	-0.642549
20	7	-3.264431	-0.636546	0.099984
21	8	3.391092	-0.146395	-0.496213
22	1	-0.956020	-1.861020	-0.197869



Imine (S39c) of 3-hydroxy-4-methoxybenzaldehyde

1	6	-1.071475	-1.535072	-0.319322
2	6	0.316999	-1.608063	-0.204063
3	6	1.068568	-0.440288	-0.074961
4	6	0.431770	0.801301	-0.062269
5	6	-0.956333	0.874223	-0.177981
6	6	-1.708072	-0.294057	-0.306124
7	1	-1.663877	-2.455906	-0.420650
8	1	0.818512	-2.586650	-0.213729
9	6	-3.888879	-0.262338	0.918506
10	1	-5.000575	-0.203667	0.826504
11	1	-3.632119	-1.213430	1.445359
12	1	-3.545757	0.591239	1.552342
13	6	2.601117	-0.520793	0.053153
14	1	3.008096	-1.540655	0.032362
15	6	4.833041	0.193018	0.238849
16	1	5.478826	1.098342	0.344320
17	1	5.028986	-0.478682	1.109770
18	1	5.136173	-0.345199	-0.692015
19	7	3.384139	0.577866	0.181181
20	8	-3.240657	-0.213220	-0.433584
21	1	1.008570	1.697070	0.036729
22	8	-1.609292	2.146377	-0.165018
23	1	-1.286637	2.679642	-0.895170

Note: The F_K^1-derived imine (**S39a**) is more stable by 14.5 kcal/mol than its counterpart that does not have 2-hydroxy group for stabilization. **S39c** was selected in place of **S39b** for this comparative analysis to maintain the consistency of number and type of atoms.

7. Protein sequence

A. Cytochrome C from equine heart
PDB ID: 1HRC
amino acid sequence:
Ac-GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFTYTDANK
NKGITWKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE

B. Myoglobin from equine skeletal muscle
PDB ID: 1WLA
amino acid sequence:
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKAS
EDLKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSK
HPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

C. RNase A from bovine pancreas
PDB ID: 2AAS
amino acid sequence:
KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAV
CSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEG
NPYVPVHFDASV

8. NMR data

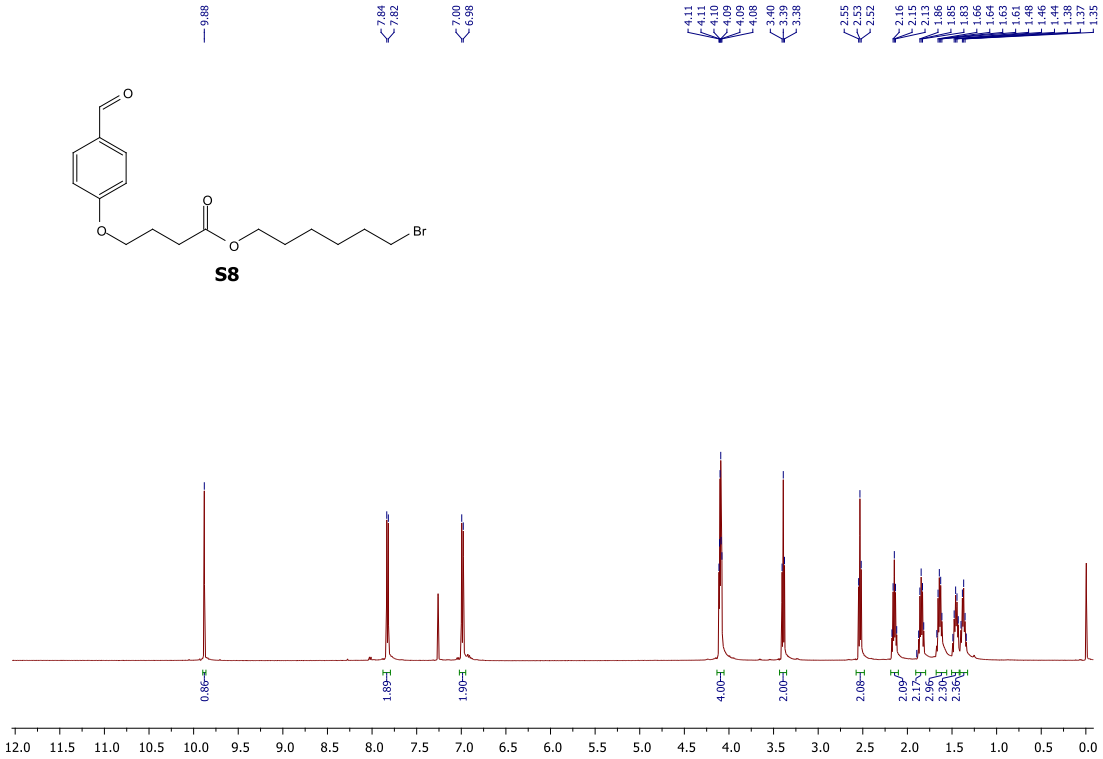


Figure S71. ^{1}H-NMR spectrum of compound **S8**.

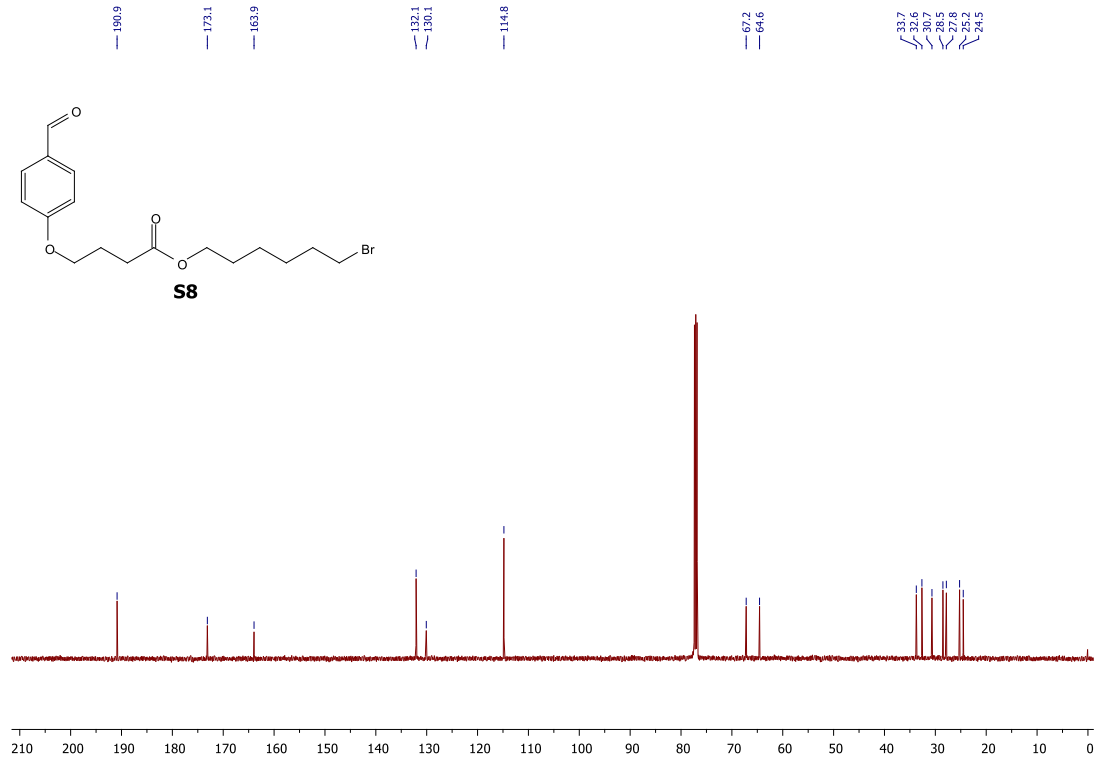


Figure S72. ^{13}C-NMR spectrum of compound **S8**.

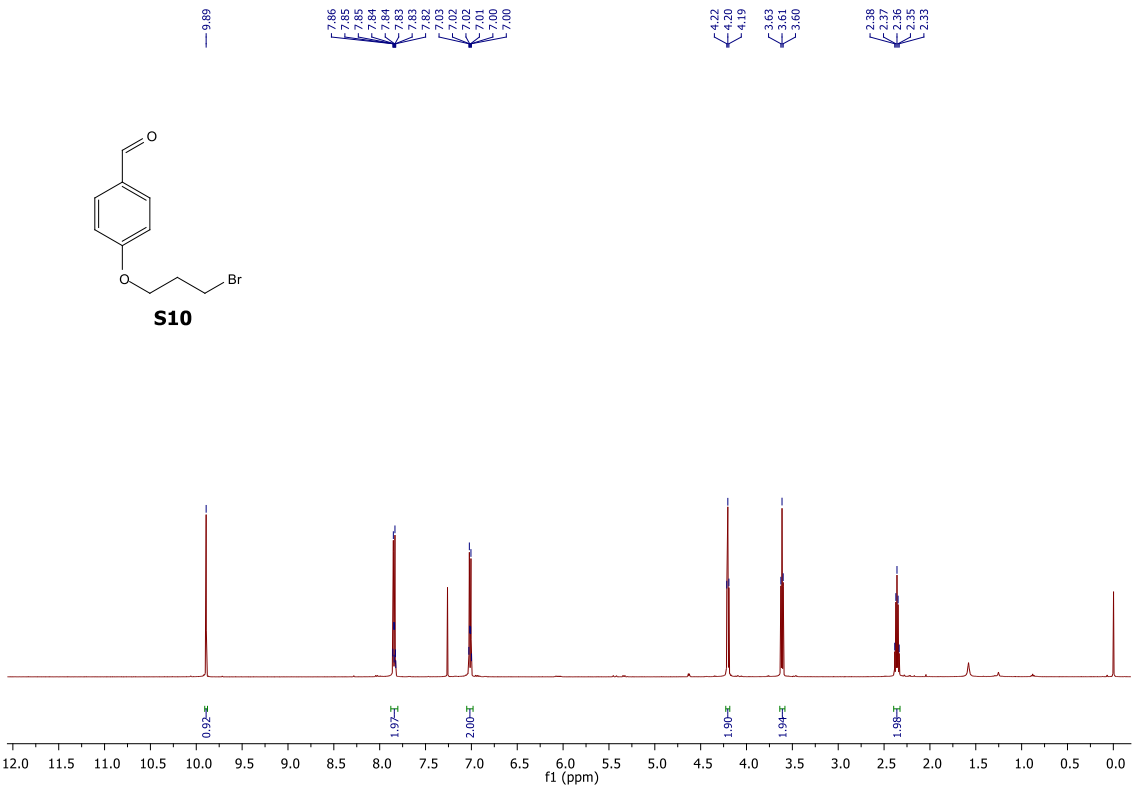


Figure S73. ^{1}H-NMR spectrum of compound **S10**.

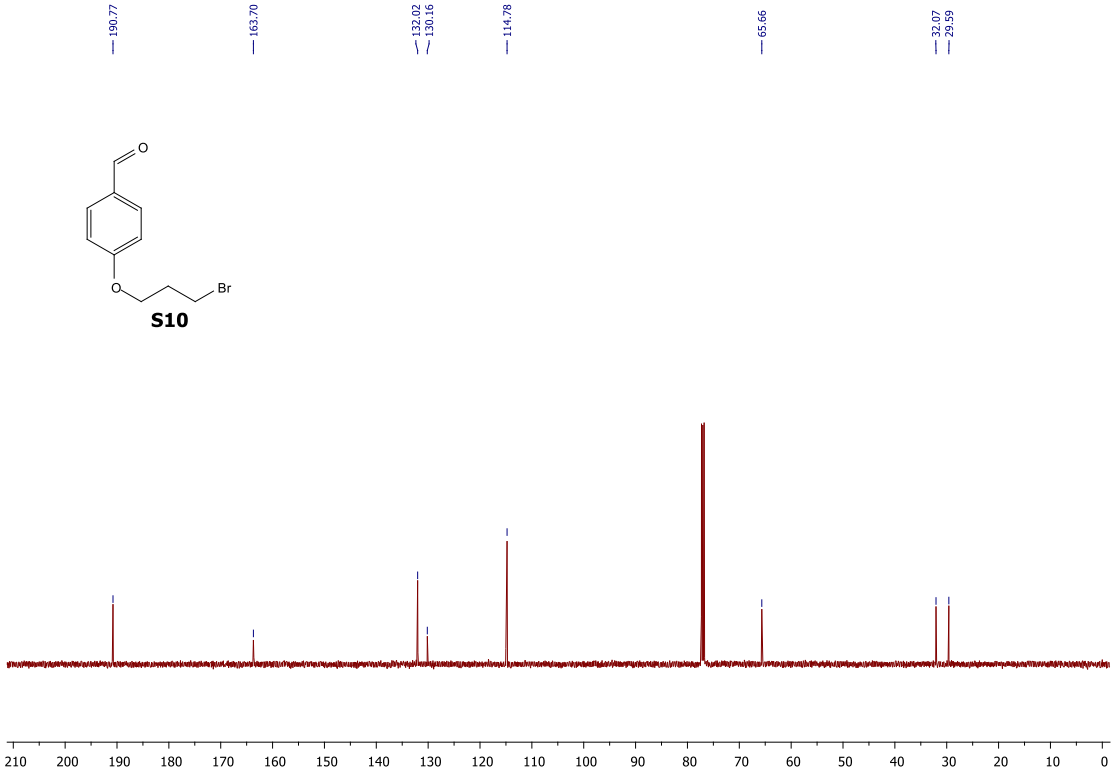


Figure S74. ^{13}C-NMR spectrum of compound **S10**.

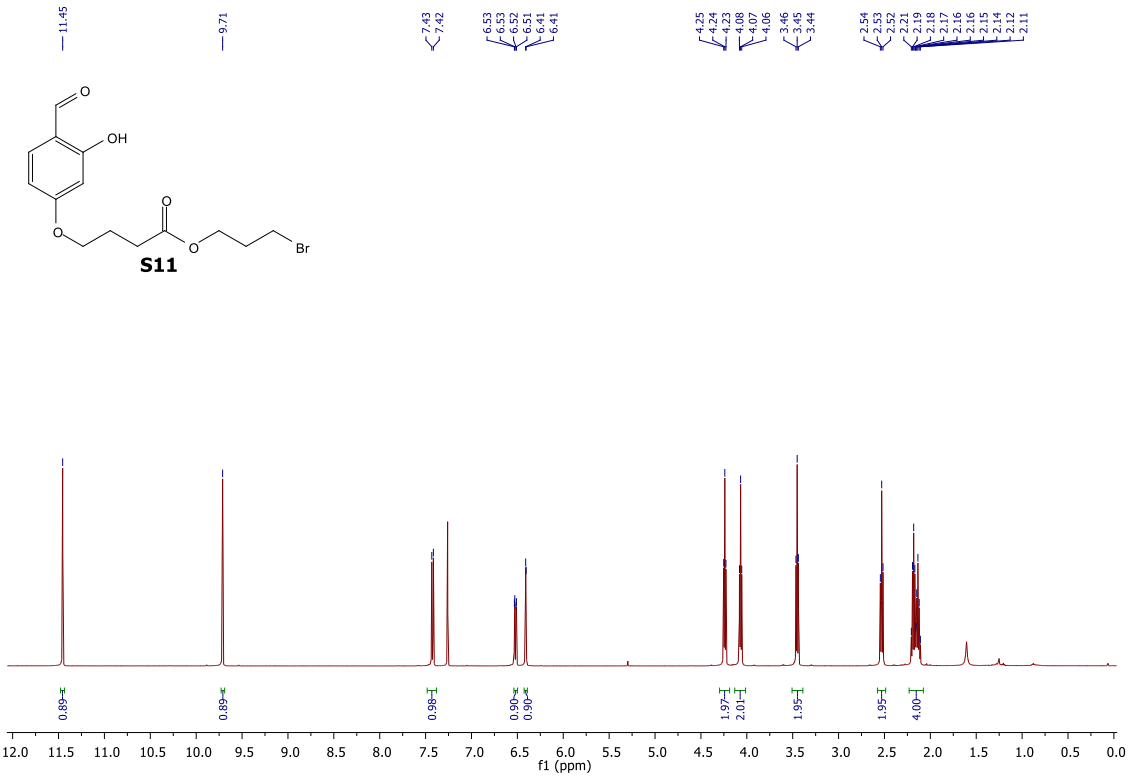


Figure S75. ^{1}H-NMR spectrum of compound **S11**.

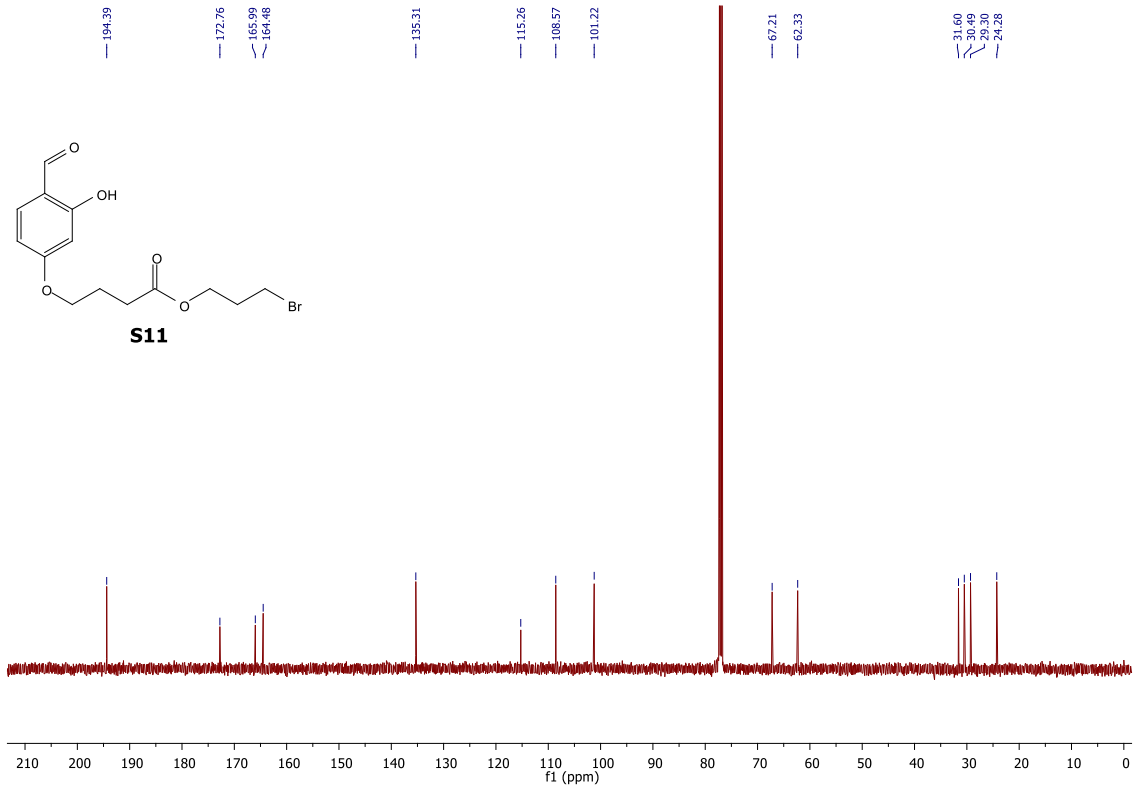


Figure S76. ^{13}C-NMR spectrum of compound **S11**.

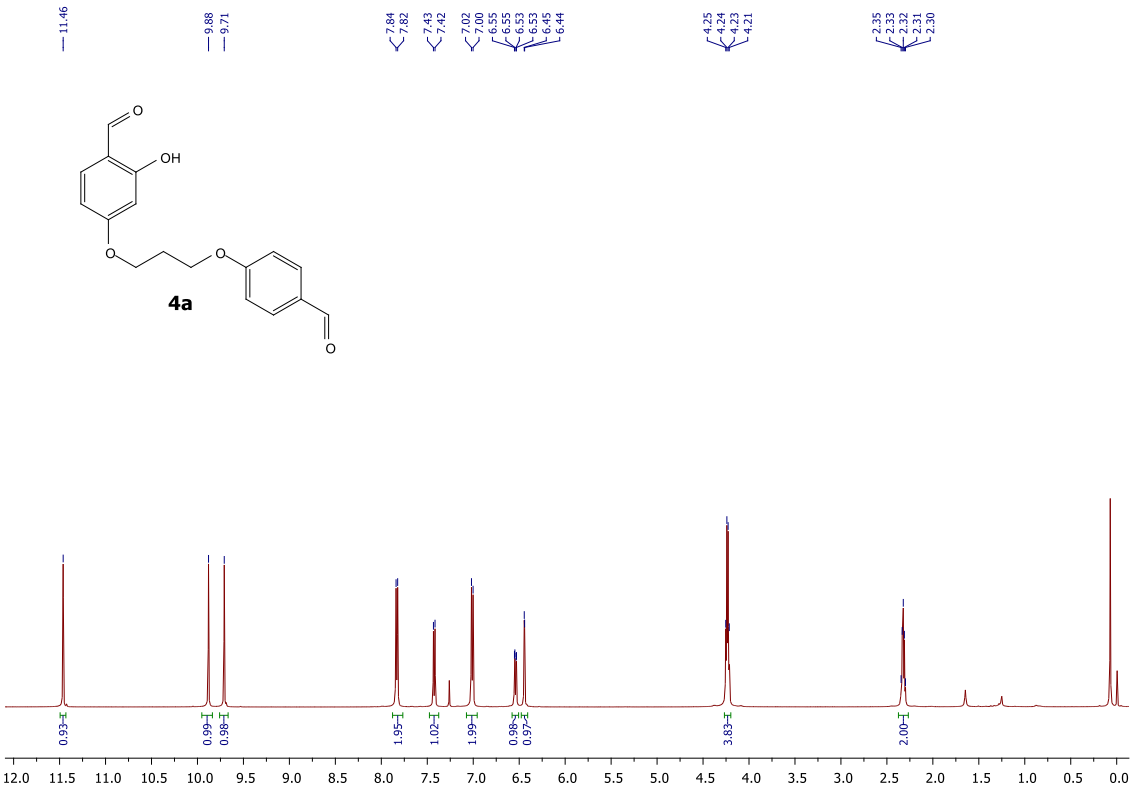


Figure S77. ^{1}H-NMR spectrum of compound **4a**.

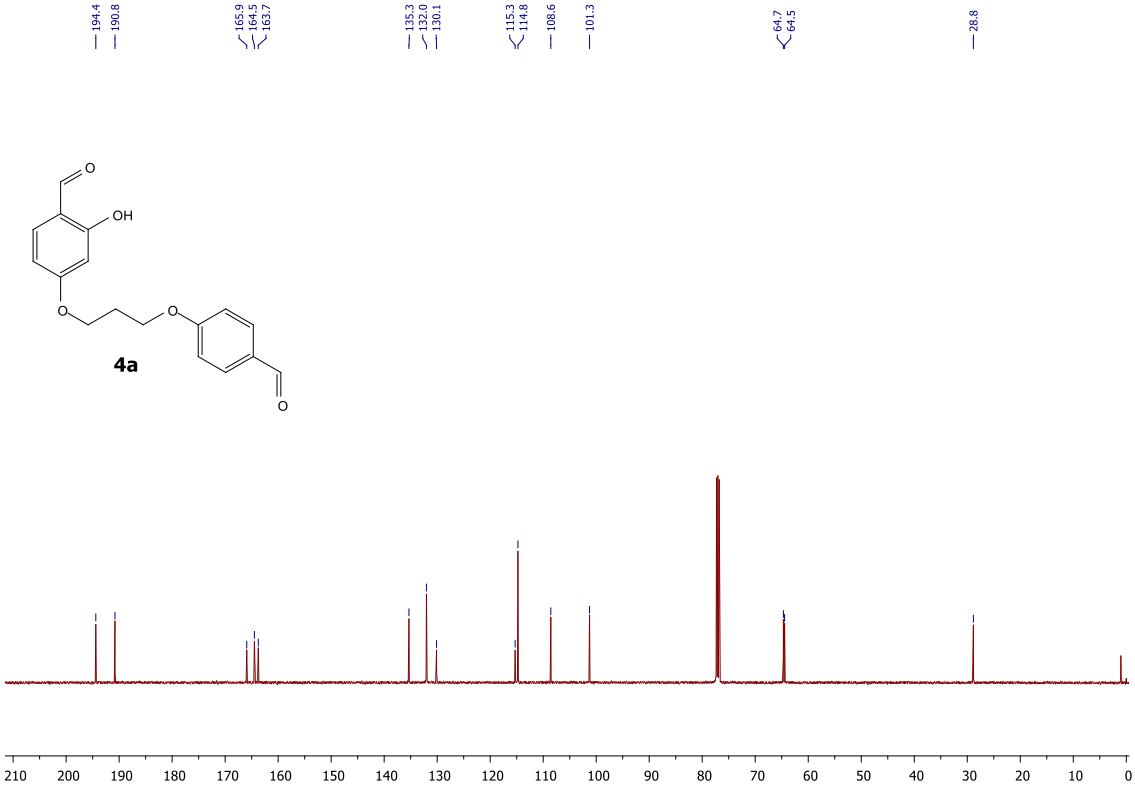


Figure S78. ^{13}C-NMR spectrum of compound **4a**.

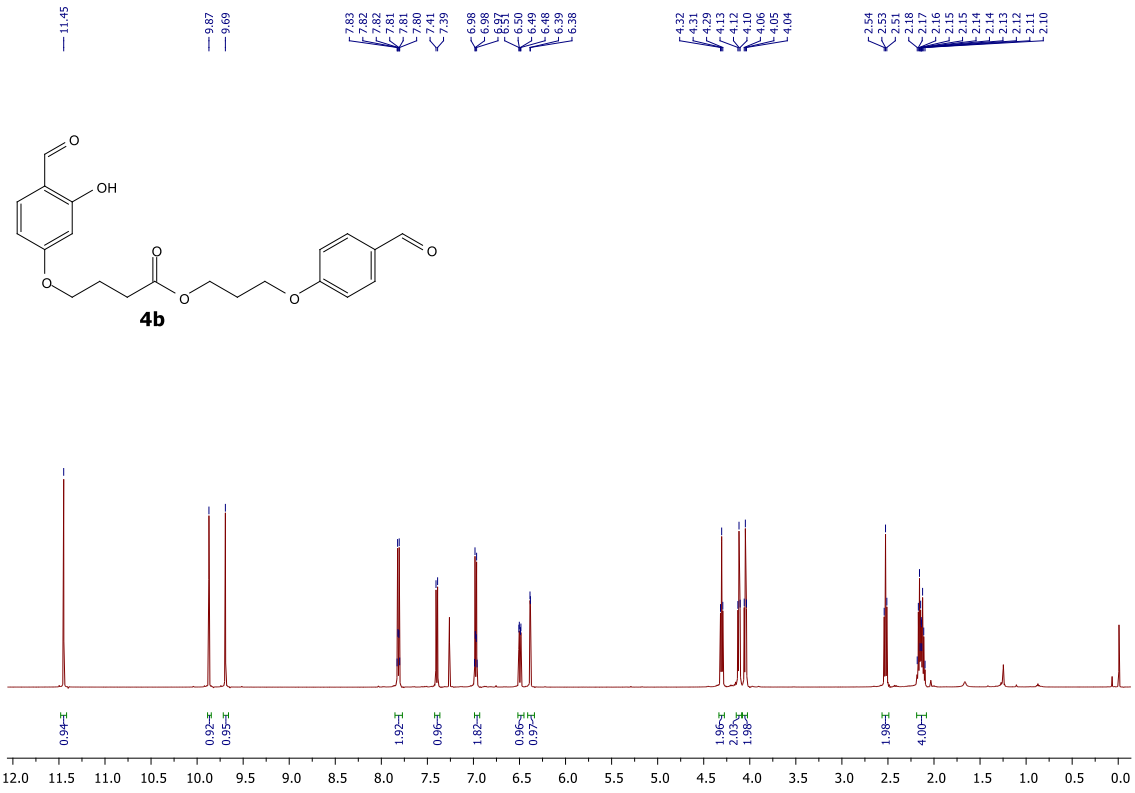


Figure S79. ^{1}H-NMR spectrum of compound **4b**.

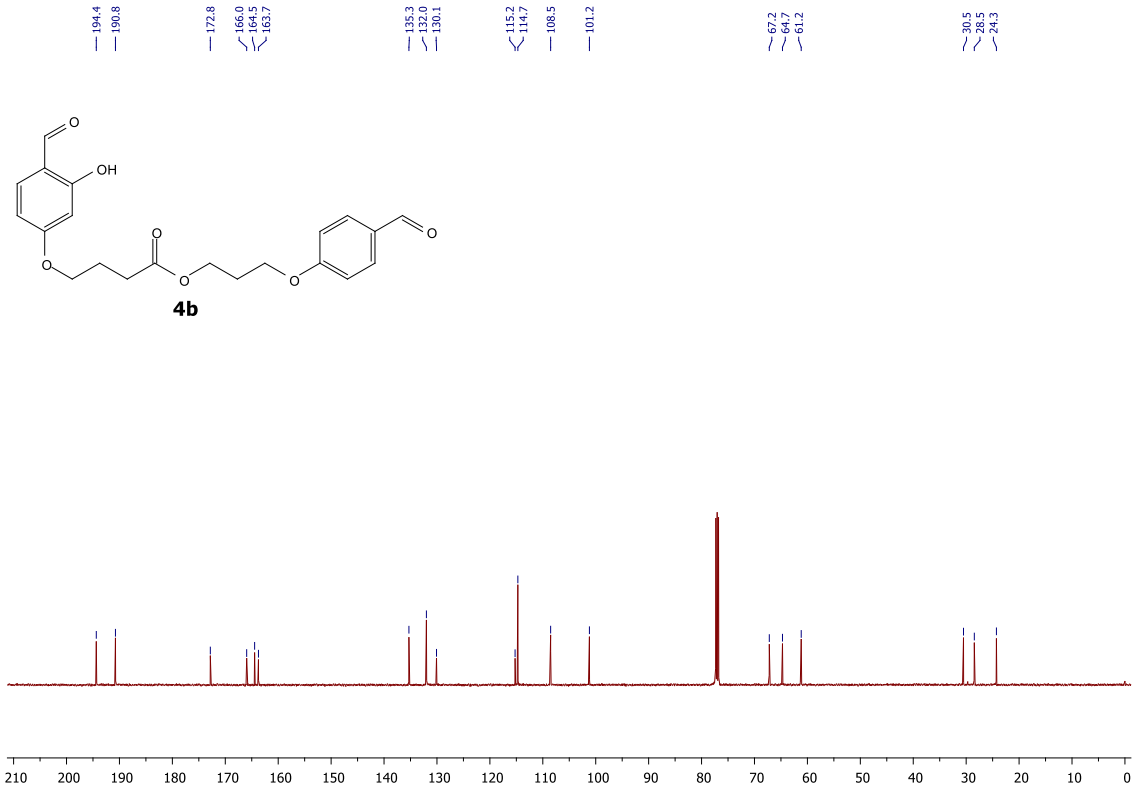


Figure S80. ^{13}C-NMR spectrum of compound **4b**.

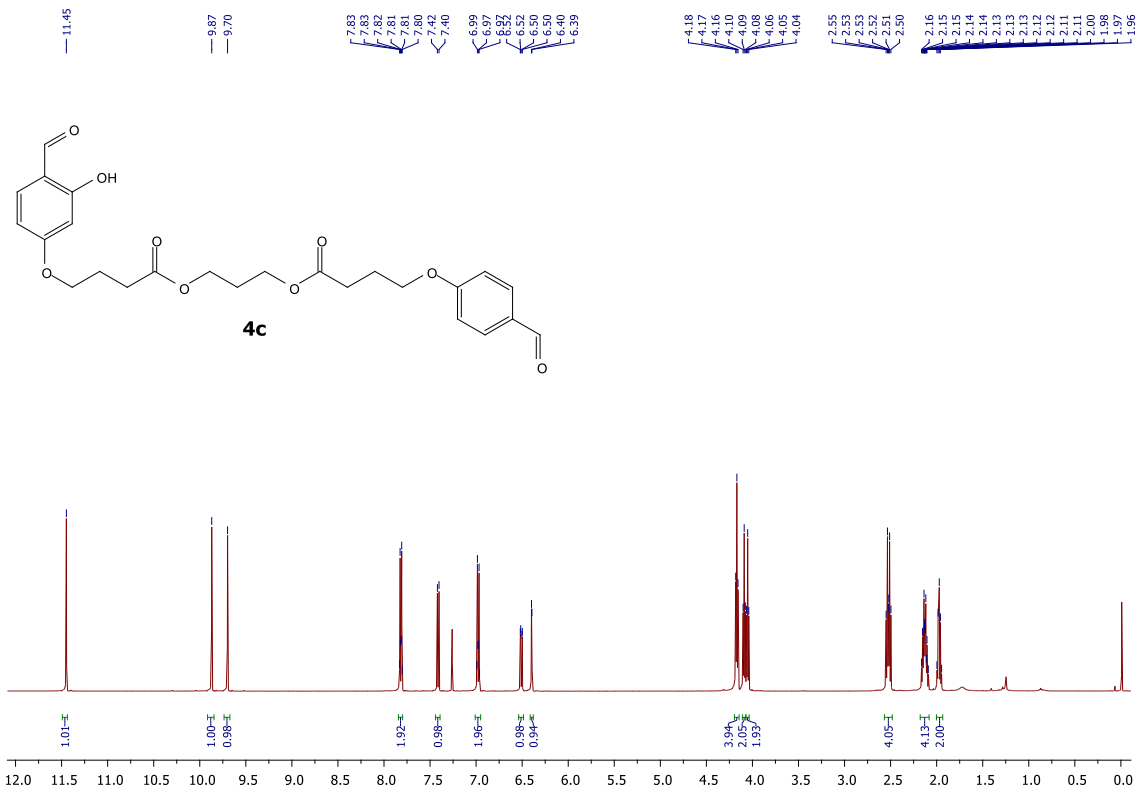


Figure S81. ^{1}H-NMR spectrum of compound **4c**.

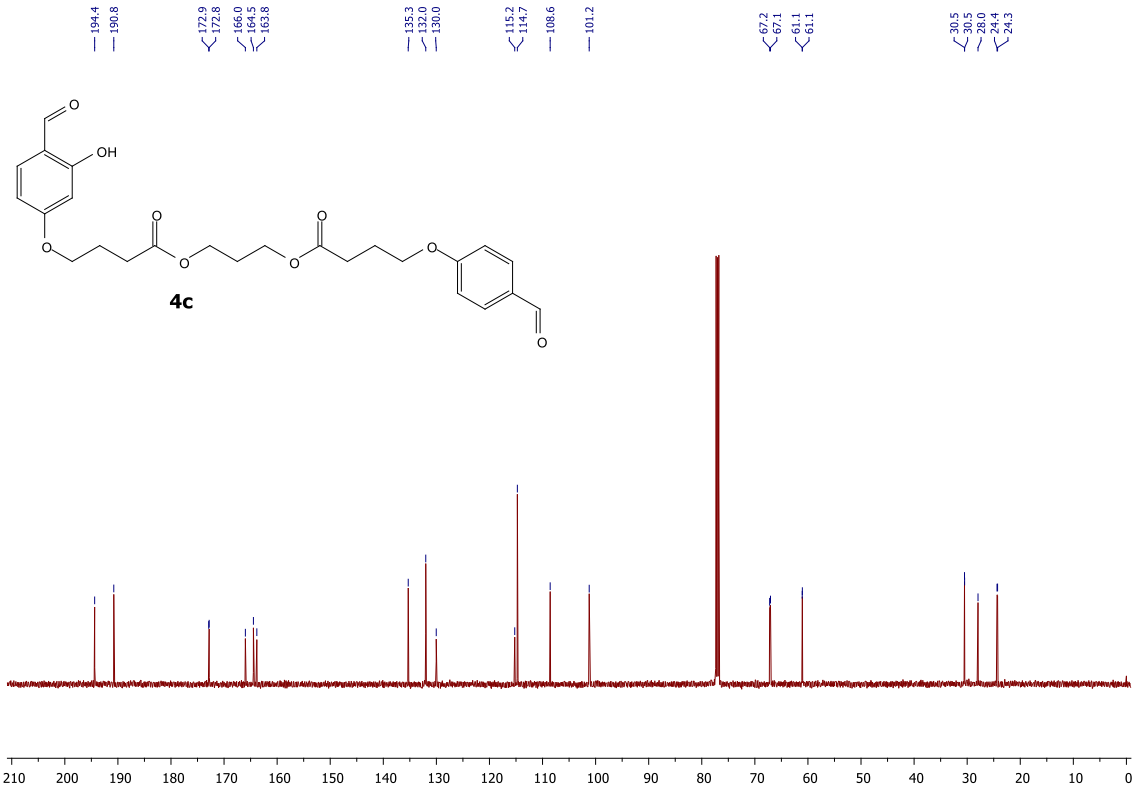


Figure S82. ^{13}C-NMR spectrum of compound **4c**.

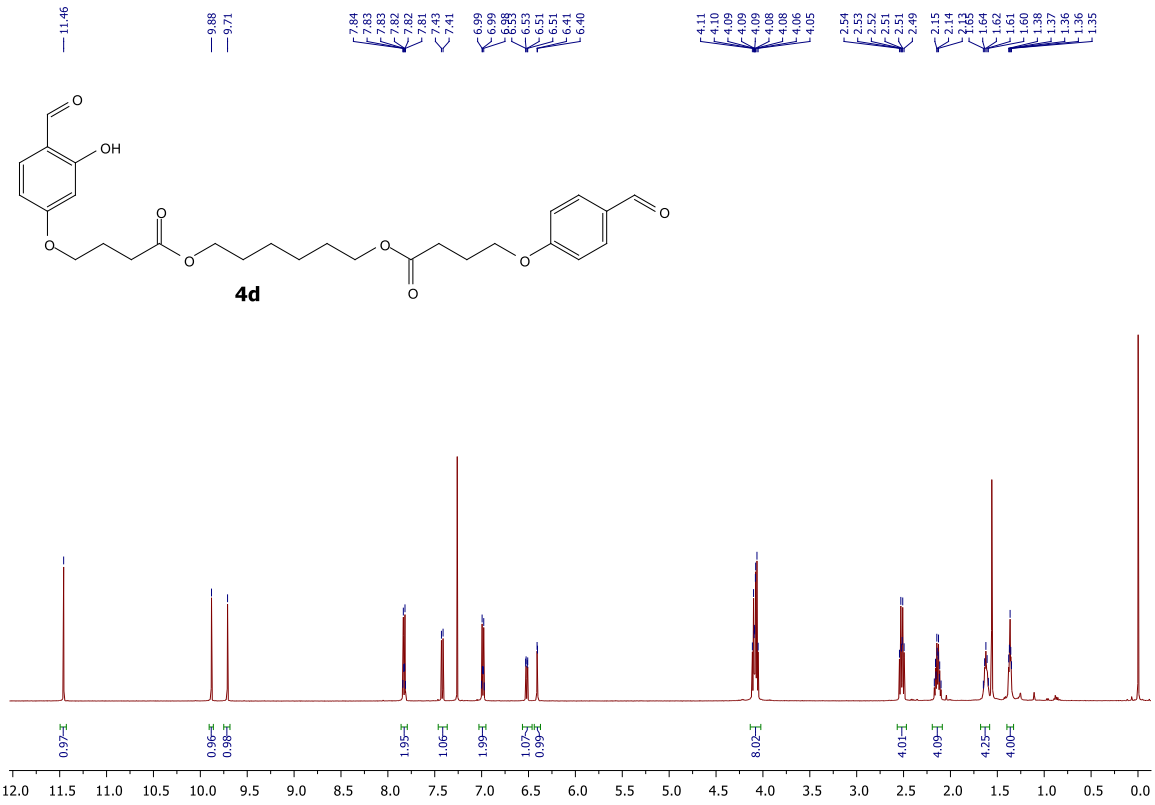


Figure S83. ^{1}H-NMR spectrum of compound **4d**.

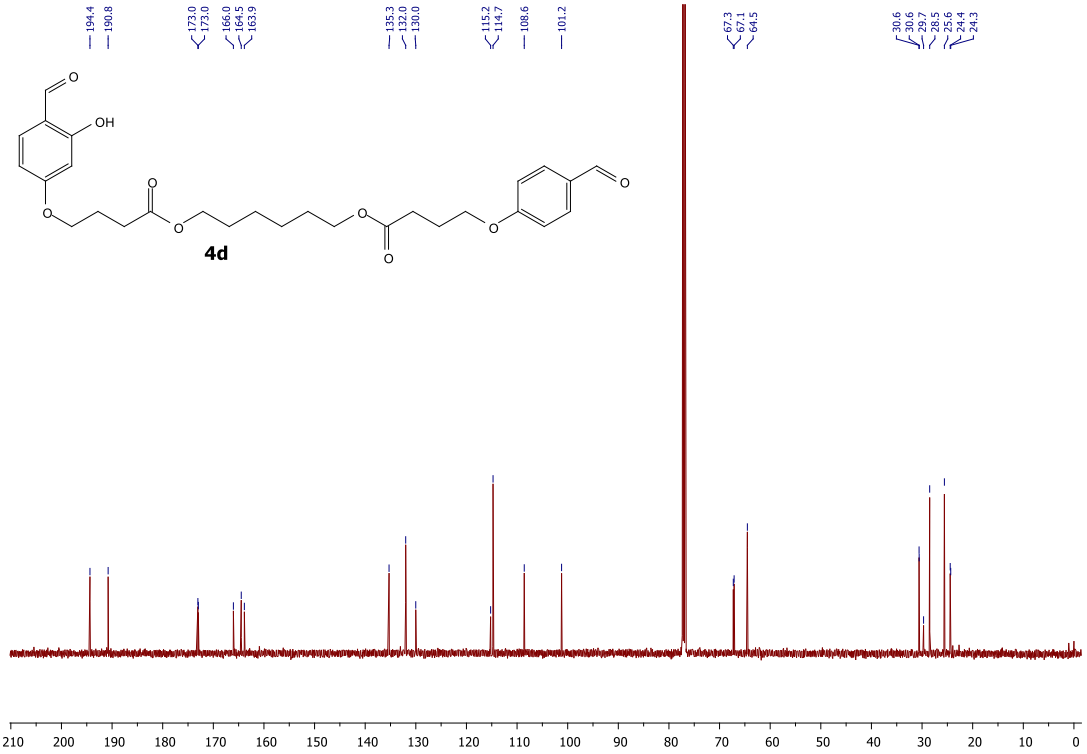


Figure S84. ^{13}C-NMR spectrum of compound **4d**.

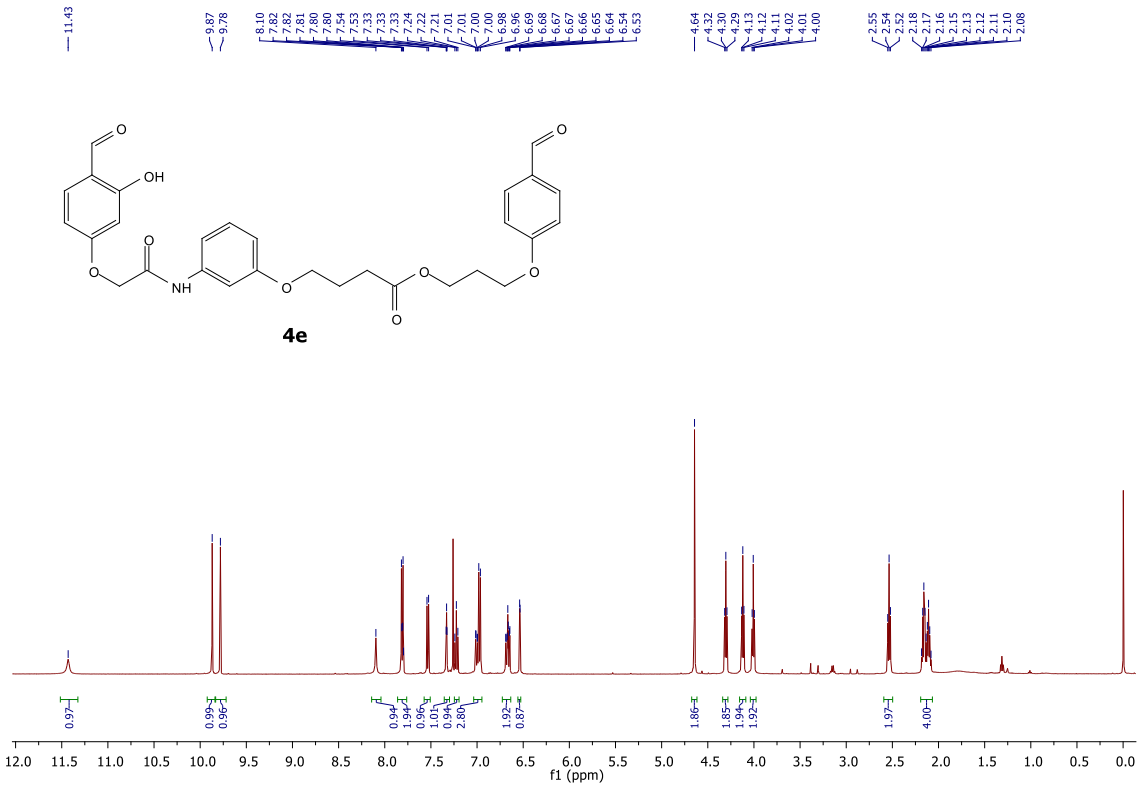


Figure S85. ^{1}H-NMR spectrum of compound **4e**.

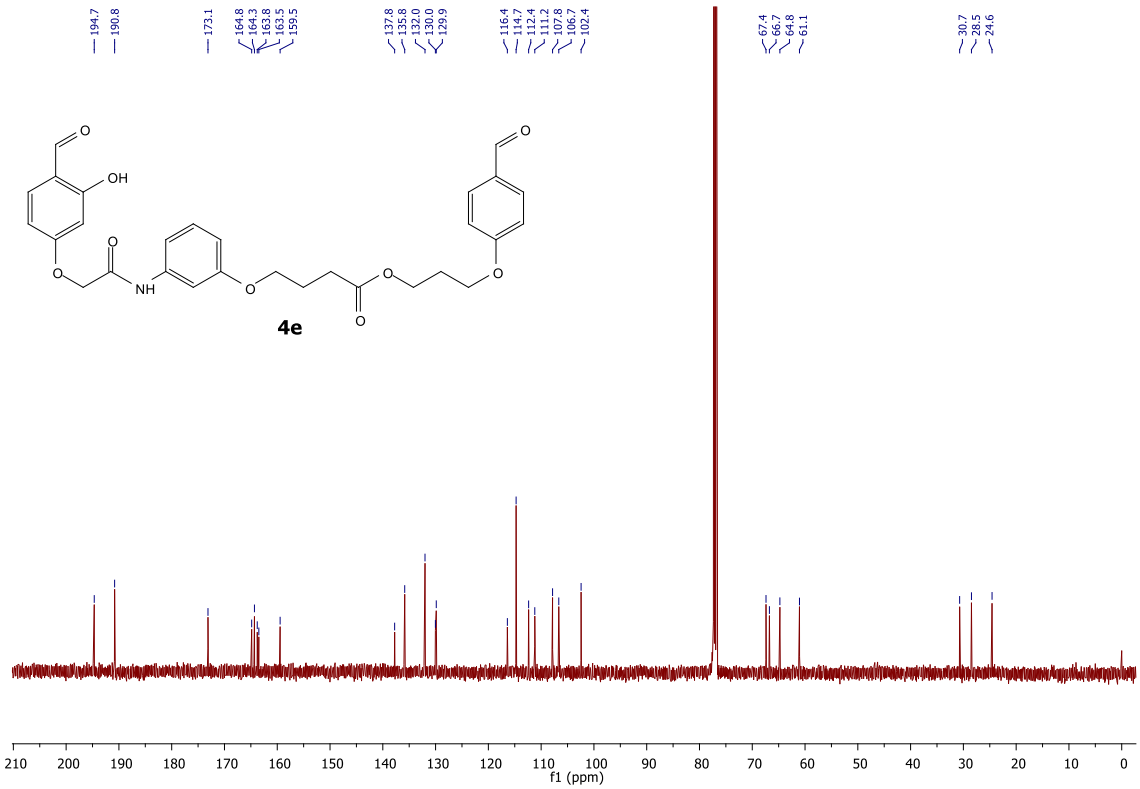


Figure S86. ^{13}C-NMR spectrum of compound **4e**.

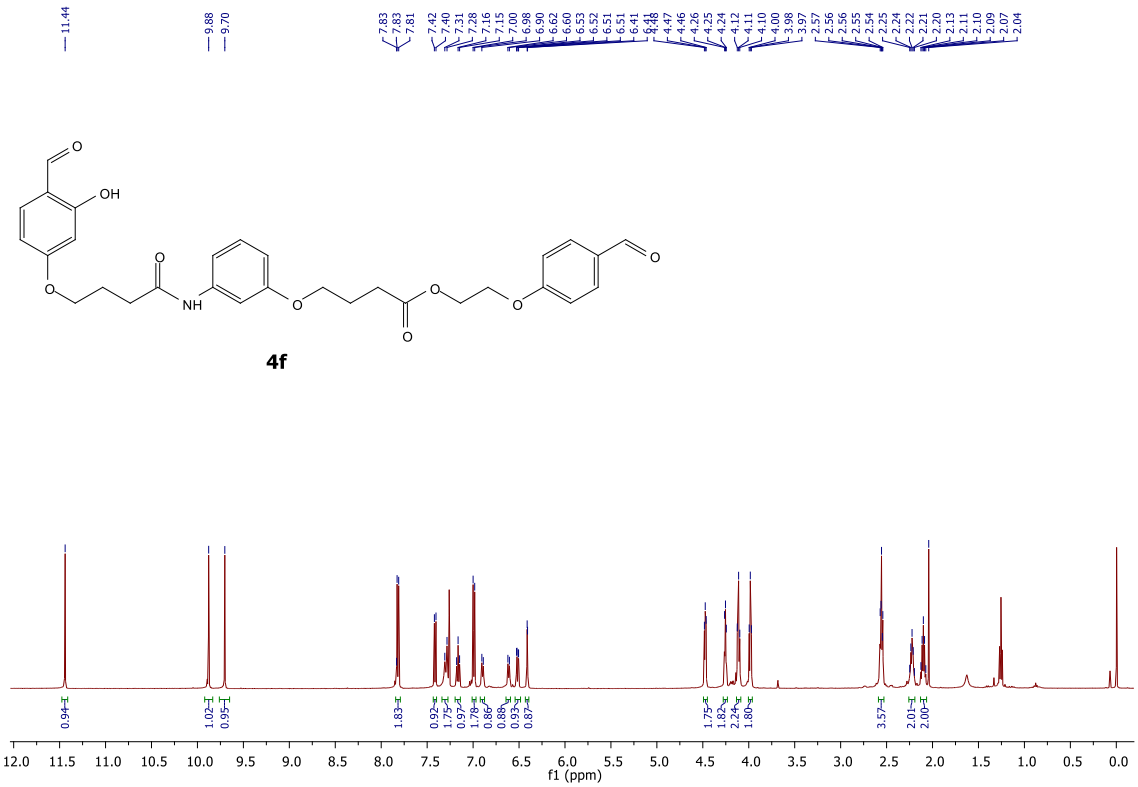


Figure S87. ^{1}H-NMR spectrum of compound **4f**.

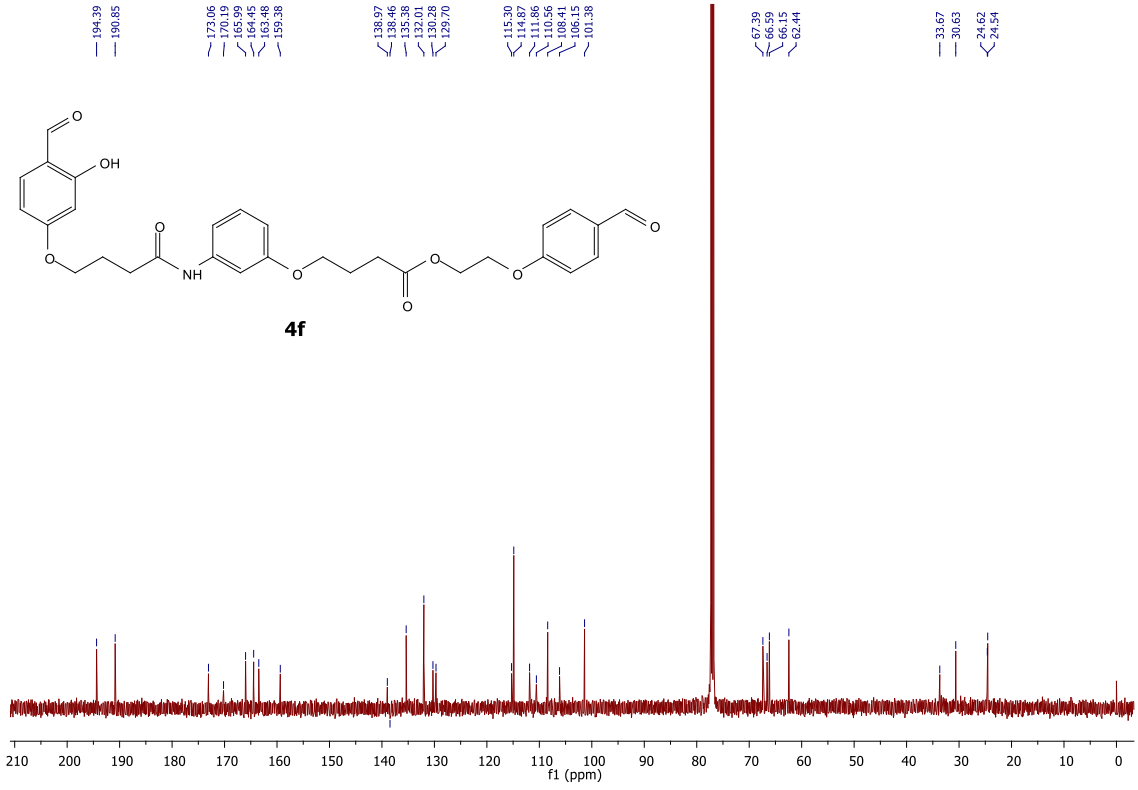


Figure S88. ^{13}C-NMR spectrum of compound **4f**.

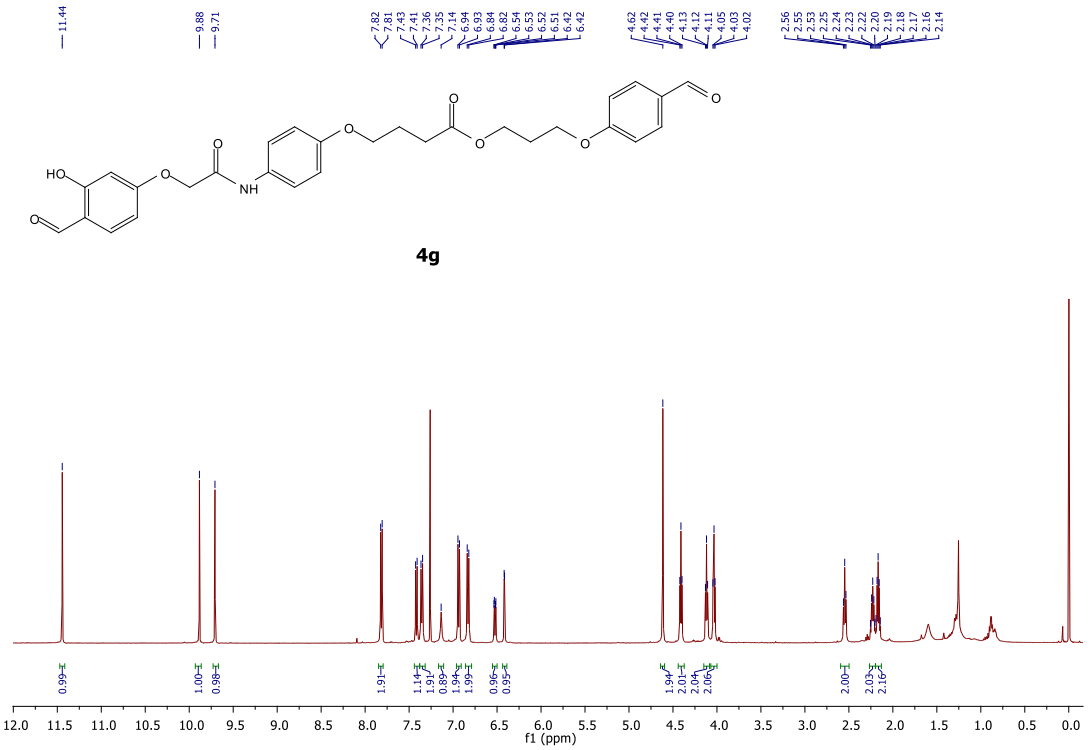


Figure S89. ^{1}H-NMR spectrum of compound **4g**.

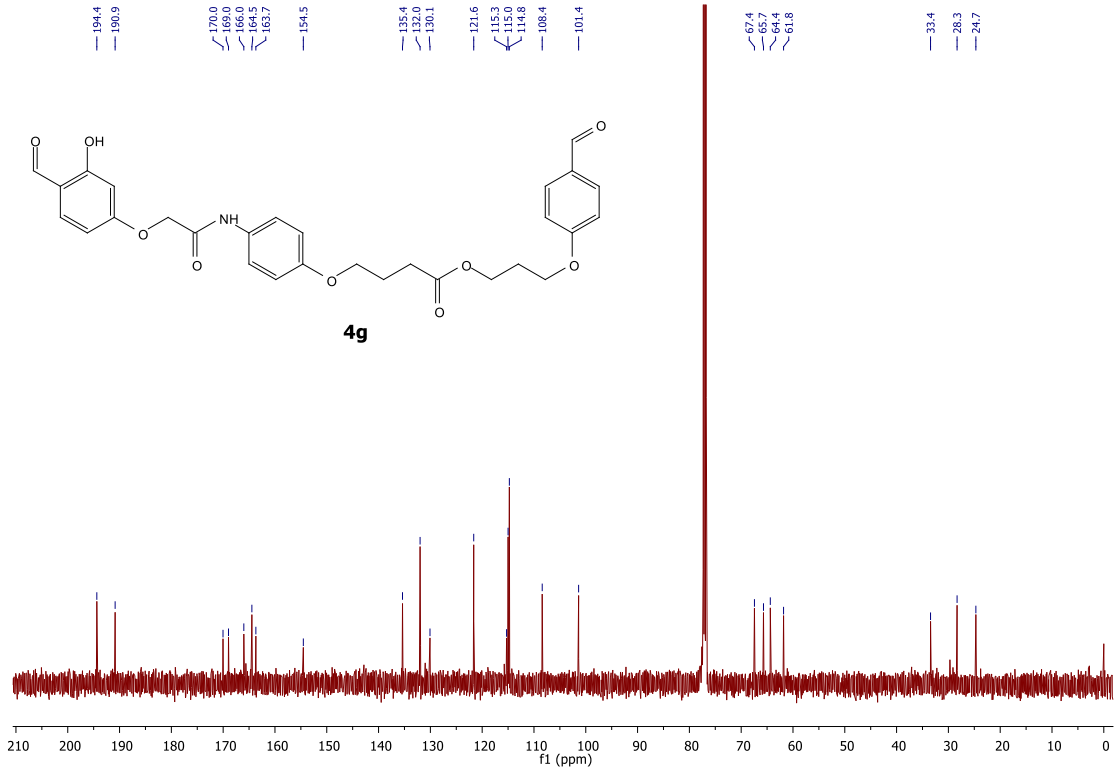


Figure S90. ^{13}C-NMR spectrum of compound **4g**.

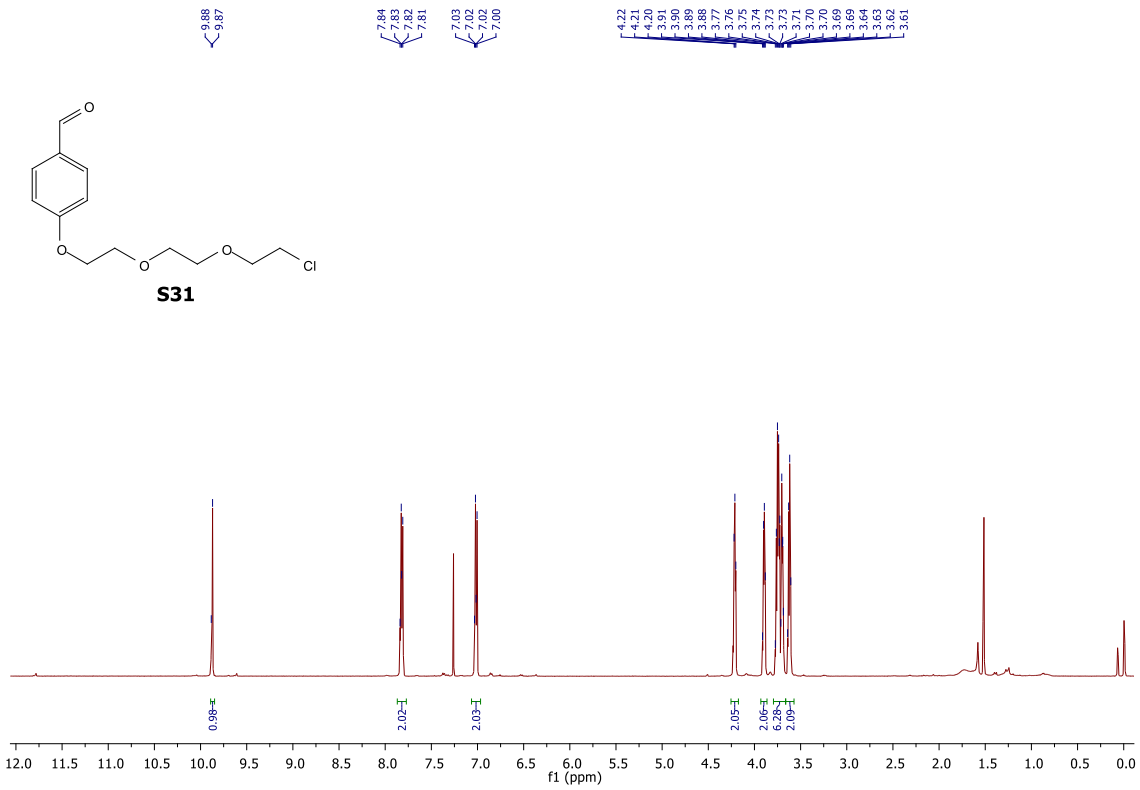


Figure S91. ^{1}H-NMR spectrum of compound **S31**.

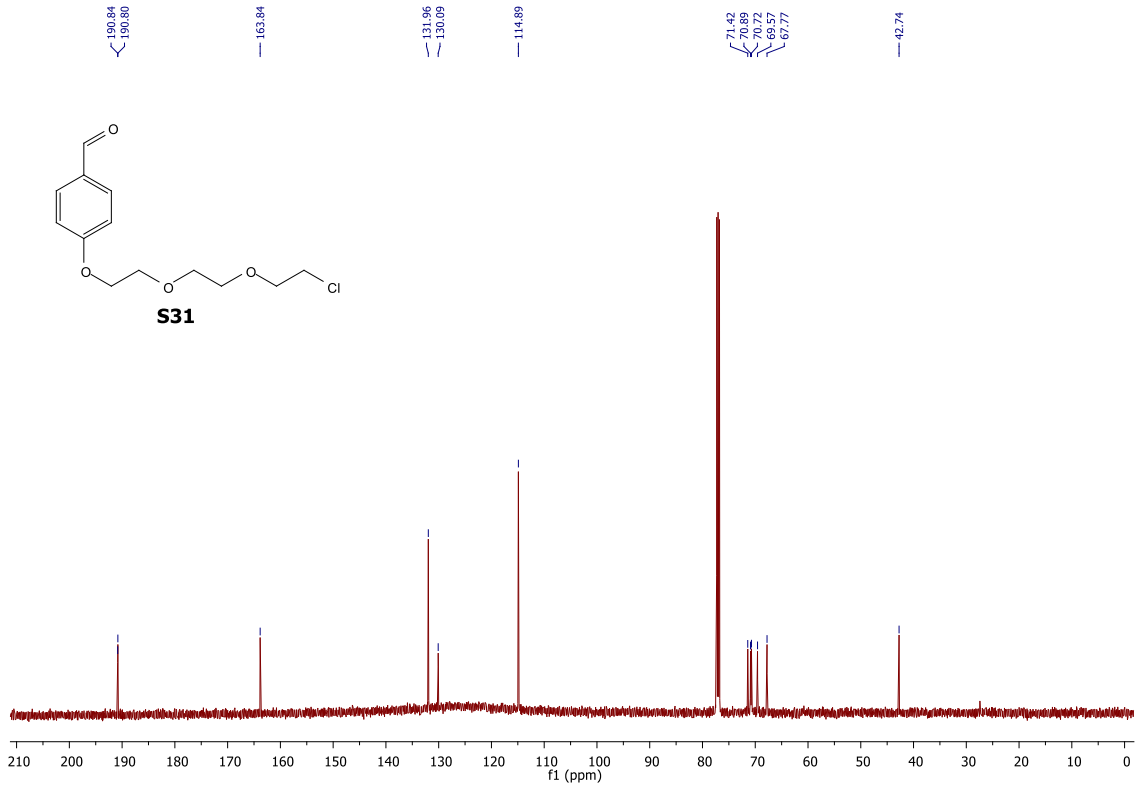


Figure S92. ^{13}C-NMR spectrum of compound **S31**.

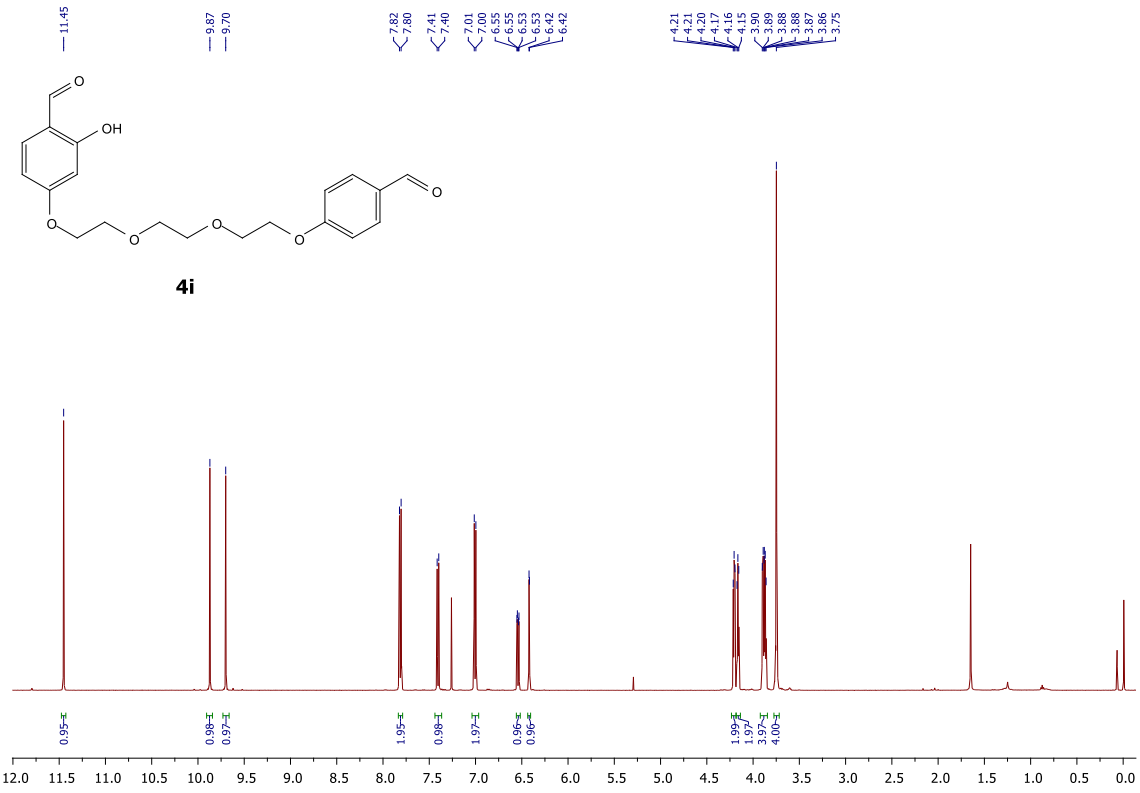


Figure S93. ^{1}H-NMR spectrum of compound **4i**.

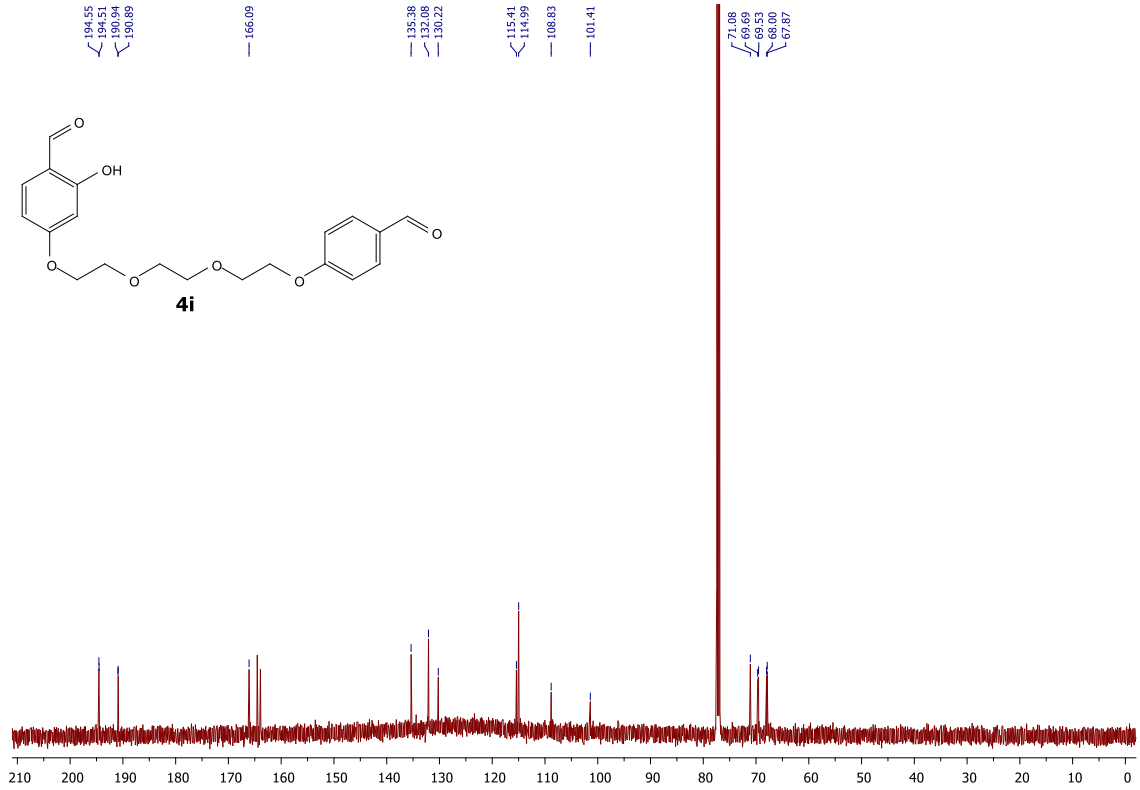


Figure S94. ^{13}C-NMR spectrum of compound **S4i**.

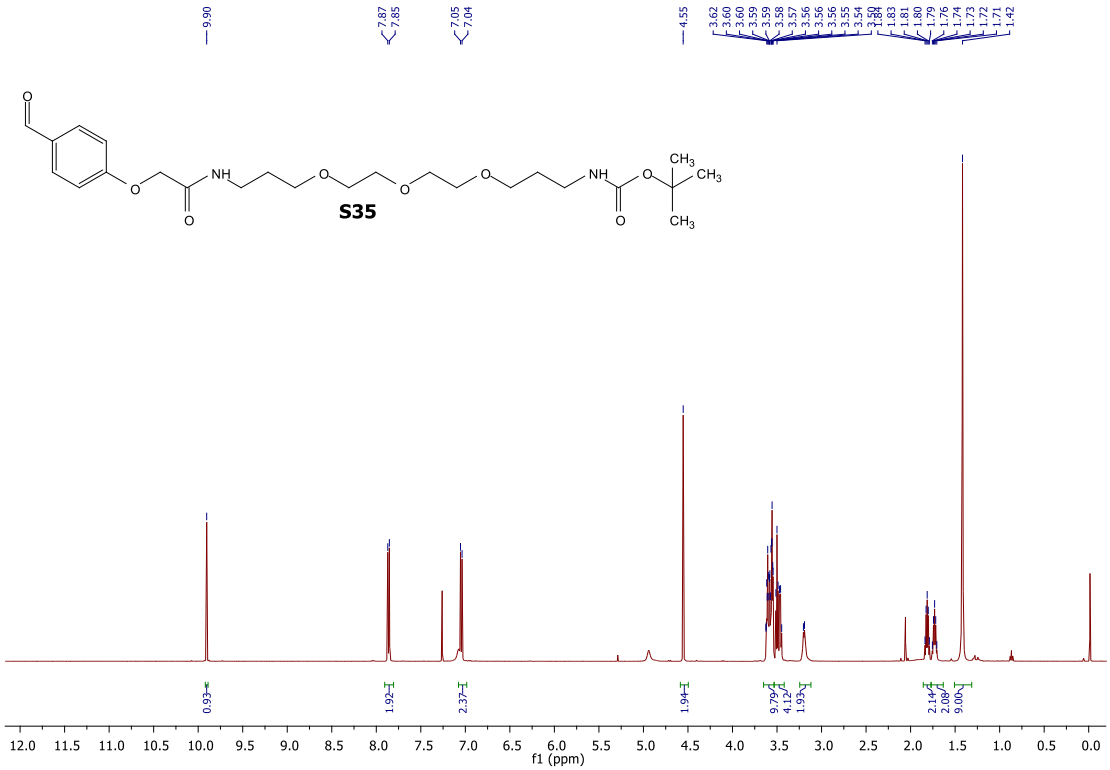


Figure S95. ^{1}H-NMR spectrum of compound **S35**.

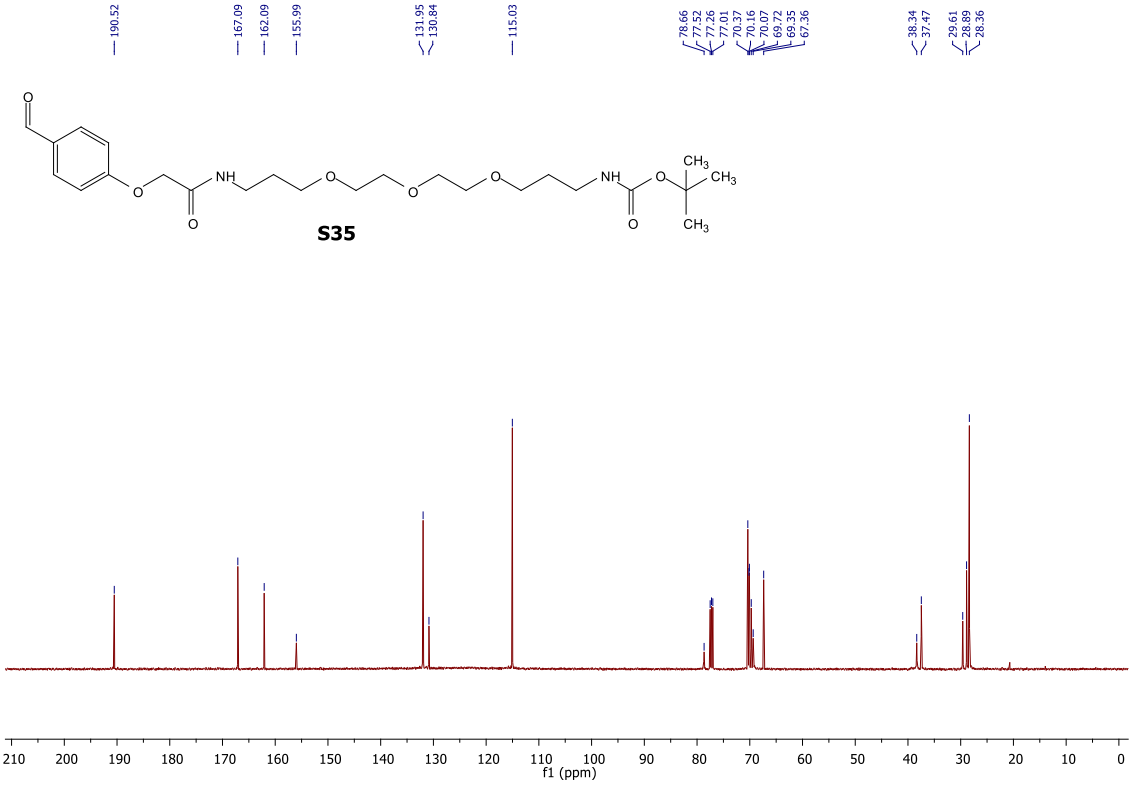


Figure S96. ^{13}C-NMR spectrum of compound **S35**.

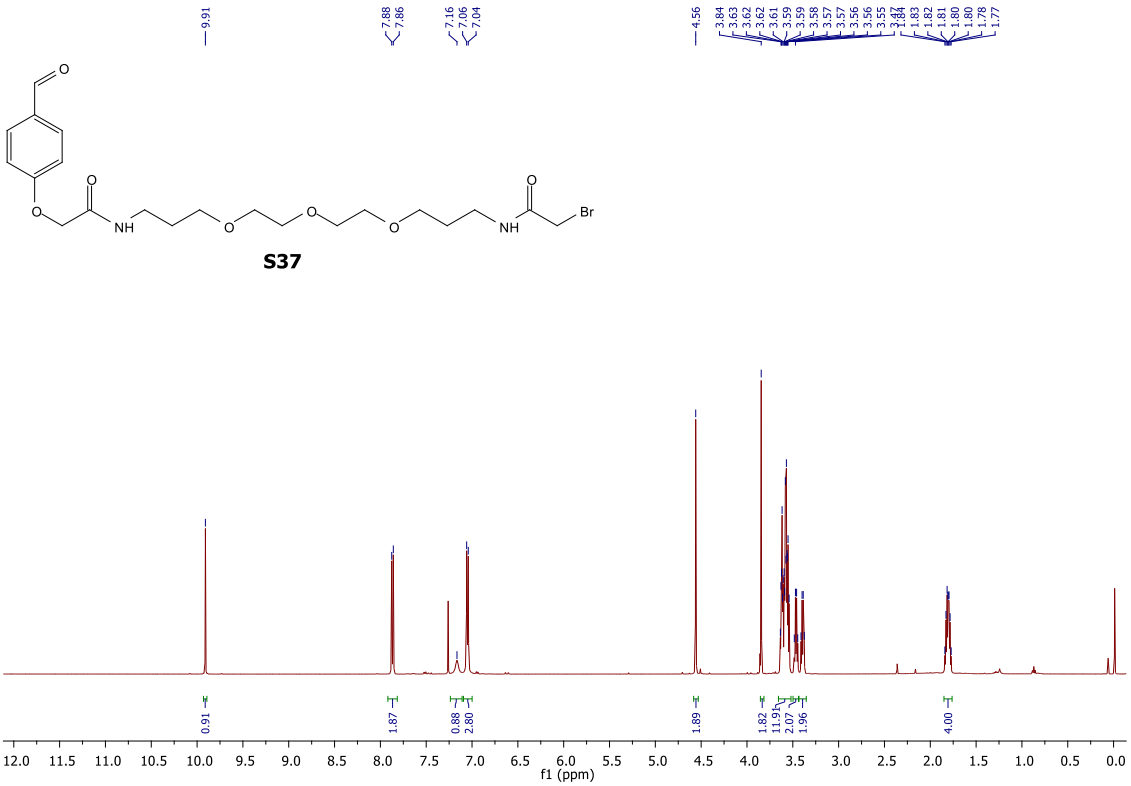


Figure S97. ^{1}H-NMR spectrum of compound **S37**.

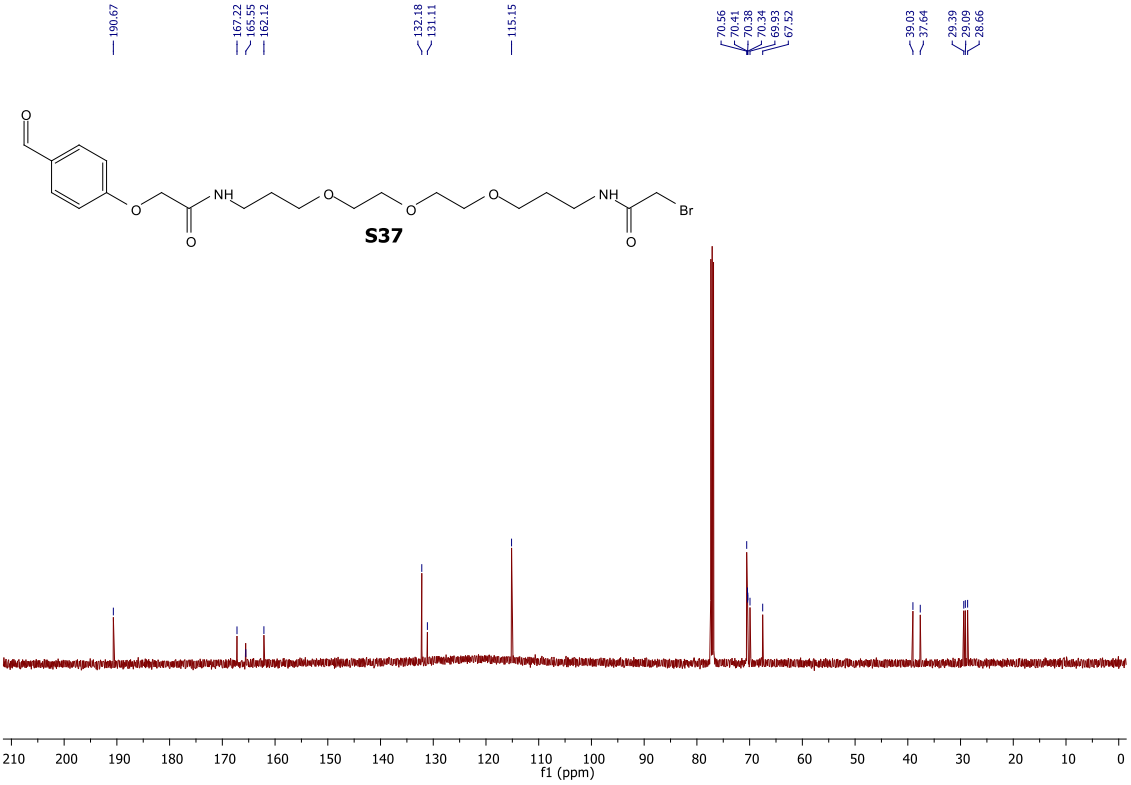


Figure S98. ^{13}C-NMR spectrum of compound **S37**.

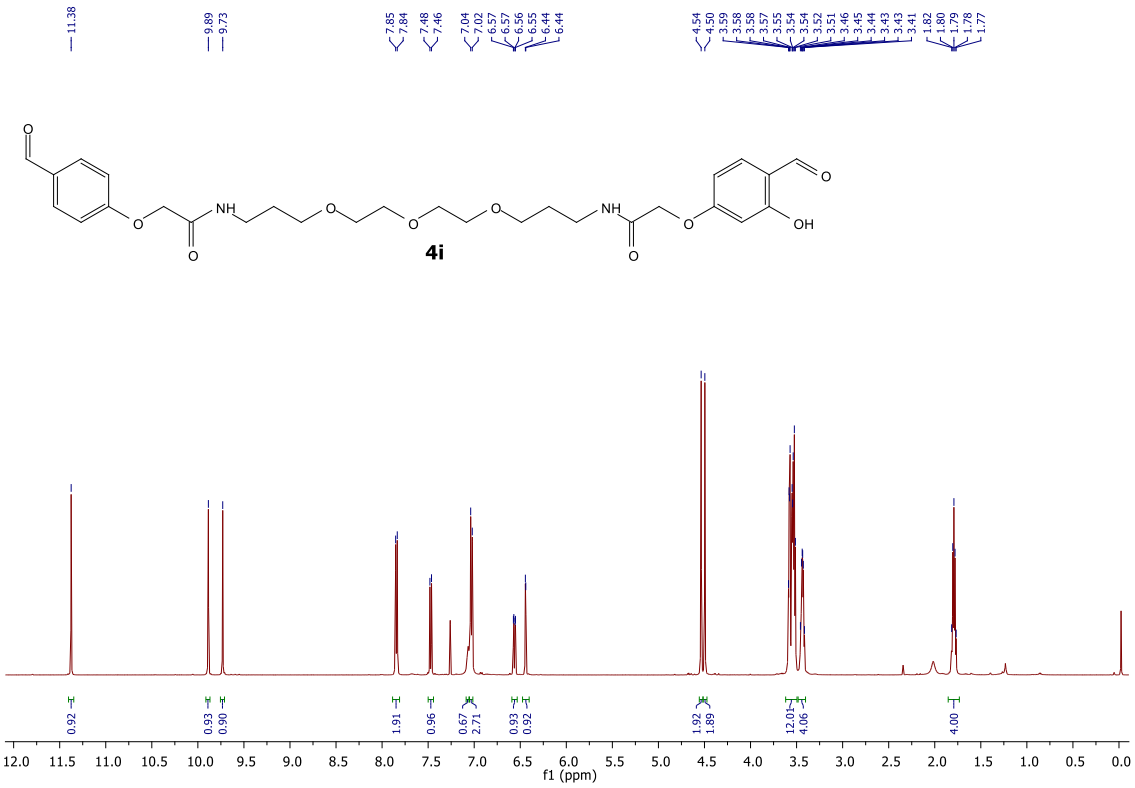


Figure S99. ^{1}H-NMR spectrum of compound **4i**.

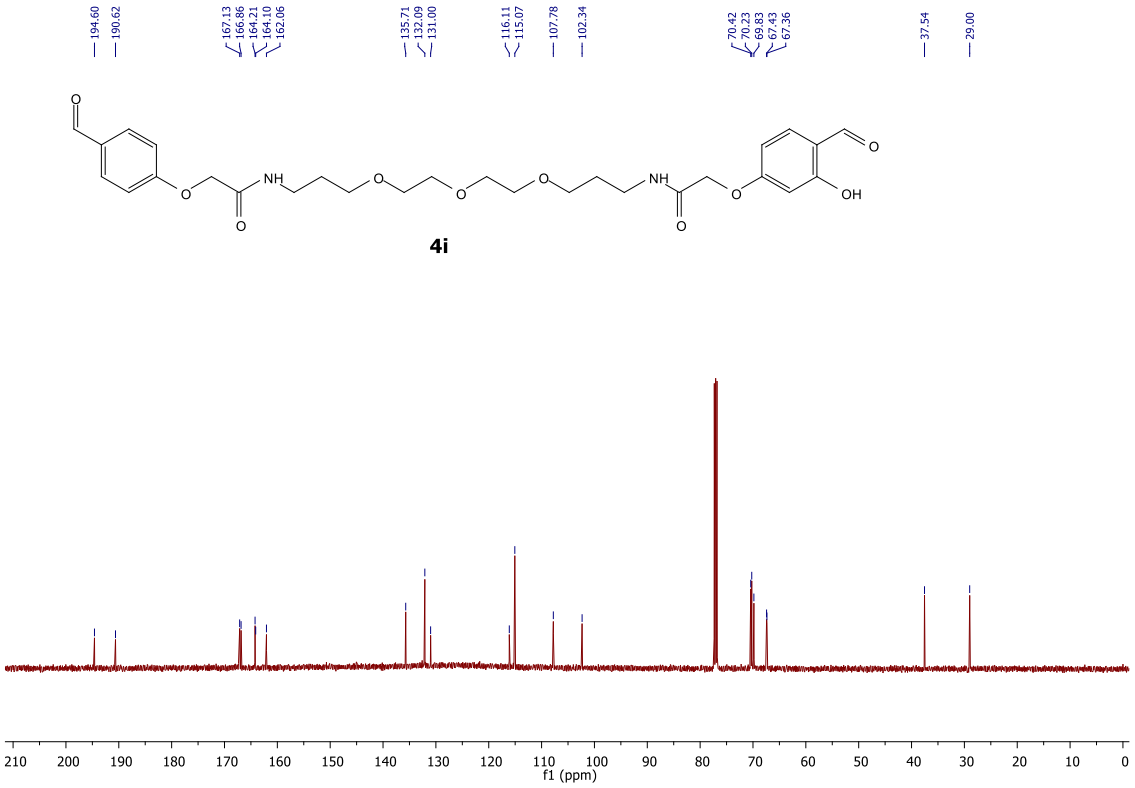


Figure S100. ^{13}C-NMR spectrum of compound **4i**.

9. Additional references

[1] N. H. Kim, M. S. Jeong, S. Y. Choi, and J. H. Kang, *Bull. Korean Chem. Soc.,* 2004, **25**, 1889-1892.
[2] S. P. Velagapudi, S. J. Seedhouse, J. French, and M. D. Disney, *J. Am. Chem. Soc.*, 2011, **133**, 10111-10118.
[3] S. R. Adusumalli, D. G. Rawale, K. Thakur, L. Purushottam, N. Reddy. N. Kalra, S. Shukla and V. Rai, *Angew. Chem. Int. Ed.*,2020, **59**, 10332-10336.
[4] 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.
[5] 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.

Additional relevant references:

[6] T. Nagaya, Y. Nakamura, S. Okuyama, F. Ogata, Y. Maruoka, P. L. Choyke and H. Kobayashi, *Mol. Cancer Res.*, 2017, **15**, 1153-1162.
[7] S. E. Bosma, P. B. A. A. van Driel, P. C. W. Hogendoorn, P. D. S. Dijkstra and C. F. M. Sier, *J. Surg. Oncol.*, 2018, **118**, 906-914.
[8] M. J. Matos, B. L. Oliveira, N. Martínez-Sáez, A. Guerreiro, P. M. S. D. Cal, J. Bertoldo, M. Maneiro, E. Perkins, J. Howard, M. J. Deery, J. M. Chalker, F. Corzana, G. Jiménez-Osés and G. J. L. Bernardes, *J. Am. Chem. Soc.*, 2018, **140**, 4004-4017.
[9] T. Tamura and I. Hamachi, *J. Am. Chem. Soc.*, 2019, **141**, 2782-2799.
[10] T. Tamura, A. Fujisawa, M. Tsuchiya, Y. Shen, K. Nagao, S. Kawano, Y. Tamura, T. Endo, M. Umeda and I. Hamachi, *Nat. Chem. Biol.*, 2020, **16**, 1361-1367.
[11] Z. Chen, F. Vohidov, J. M. Coughlin, L. J. Stagg, S. T. Arold, J. E. Ladbury and Z. T. Ball, *J. Am. Chem. Soc.* 2012, **134**, 10138-10145.
[12] A. O.-Y. Chan, C. M. Ho, H.-C. Chong, Y.-C. Leung, J.-S. Huang, M.-K. Wong and C.-M. Che, *J. Am. Chem. Soc.*, 2012, **134**, 2589-2598.
[13] M. Raj, H. Wu, S. L. Blosser, M. A. Vittoria and P. S. Arora, *J. Am. Chem. Soc.*, 2015, **137**, 6932-6940.
[14] O. Boutureira and G. J. L. Bernardes, *Chem. Rev.*, 2015, **115**, 2174-2195.