Supporting Information

Amine-Selective Bioconjugation Using Arene Diazonium Salts

Stefan Diethelm, Michael A. Schafroth, Erick M. Carreira*

Laboratorium für Organische Chemie, ETH Zürich, HCI, H335, Vladimir-Prelog-Weg 3, CH-8093 Zürich, Switzerland

erickm.carreira@org.chem.ethz.ch

Supporting information:

Table of content

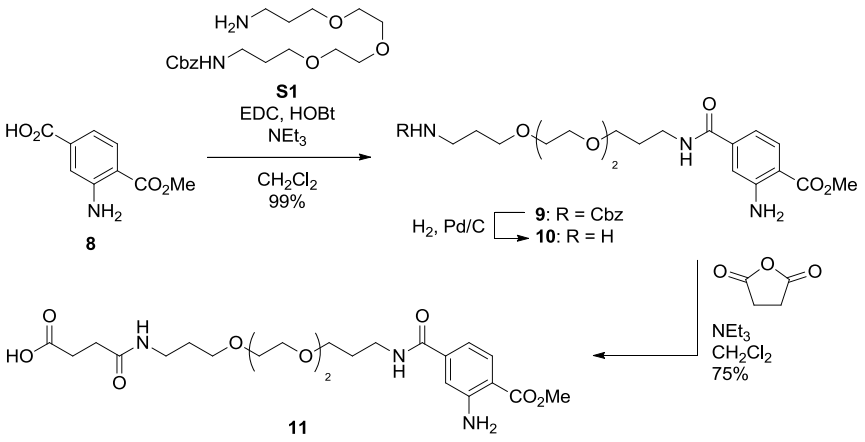
1. General Methods

Chemicals and Solvents: All chemicals were purchased from Acros, Aldrich, Fluka, TCI or ABCR and used as such unless stated otherwise. Proteins were purchase form Sigma-Aldrich (myoglobin from equine skeletal muscle (No. M0630); lysozyme from egg white (No. 62970); cytochrome c from equine heart (No. C2506); ribonuclease a from bovine pancreas (No. 83831)) or from Axon Lab (α-chymotrypsinogen (No. A3975)). For flash chromatography technical grade solvents were used without further purification. For reactions tetrahydrofuran, and dichloromethane were purified by passage over activated alumina under argon atmosphere. Methanol was distilled from magnesium turnings under an atmosphere of nitrogen. Triethylamine and diisopropyl ethylamine were distilled under nitrogen from CaH_2. pH 7 buffer for workup was prepared with $NaH_2PO_4 \cdot 2H_2O$ (8g), $Na_2HPO_4 \cdot 12H_2O$ (18g) in water (1000 mL). pH buffers for bioconjugation reactions were prepared with NaOAc or NaH_2PO_4 (100 mM in water). Deuterated solvents were obtained from ARMAR chemicals (Switzerland).

Reactions: All non-aqueous reactions were carried out using oven dried glassware under an atmosphere of argon unless otherwise stated. Reactions were magnetically stirred and monitored by TLC unless otherwise stated. Chromatographic purification was performed as flash chromatography (Fluka silica gel, 60 Å pore size) using the solvents indicated as eluent with 0.3-0.5 bar pressure. Thin layer chromatography (TLC) was performed on Merck silica gel 60 F_{254} TLC glass plates and visualized with UV light or stained in ceric ammonium molybdate or potassium permanganate solutions. The yields given refer to chromatographically purified and spectroscopically pure compounds unless otherwise stated.

Analysis: ^{1}H- and ^{13}C- NMR spectra were recorded on VARIAN Mercury (300 MHz), BRUKER DRX (400 MHz, 500MHz, 600 MHz) or BRUKER Avance (400 MHz, 600 MHz) spectrometers in the solvents indicated. All signals are reported in ppm with the internal chloroform signal at 7.26 ppm or 77.0 ppm, the internal DMSO signal at 2.50 ppm or 39.5 ppm or the internal methanol signal at 3.31 ppm or 49.0 ppm as standard. The data is being reported as (s=singlet, d=doublet, t=triplet, q=quadruplet, m=multiplet or unresolved, bs=broad signal, coupling constant(s) in Hz, integration). Service measurements were performed by the NMR service team of the Laboratorium für Organische Chemie at ETH Zürich by Mr. Philipp Zumbrunnen, Mr. Rainer Frankenstein and Mr. René Arnold under the direction of Dr. Marc-Olivier Ebert. Infrared spectra were recorded on a Perkin Elmer RXI FT-IR Spectrophotometer as thin films. Absorptions are given in wavenumbers (cm^{-1}). Mass spectrometric measurements were performed as high resolution ESI measurements on a Bruker maXis ESI-Q-TOF by the mass spectrometry service of the LOC at the ETHZ.

2. Reagent Synthesis



To a solution of 1-methyl-2-aminoterephthalate (**8**) (1 g, 5.12 mmol) in CH_2Cl_2 (20 mL) was added amine **S1**[1] (2.2 g, 6.15 mmol), EDC (1.47 g, 7.69 mmol), HOBt (1.18 g, 7.69 mmol) and diisopropyl ethylamine (2.7 mL, 15.4 mmol). The mixture was stirred overnight and then diluted with sat. $NaHCO_3$. The phases were separated and the aqueous phase was extracted three times with CH_2Cl_2. The combined organic phases were washed with brine and dried over $MgSO_4$. The solvent was removed and the residue was purified by flash column chromatography (hexanes/EtOAc 1:1 → EtOAc) to give aniline **9** (2.69 g, 5.06 mmol, 99%).

9: $\mathbf{R_f}$ 0.35 (EtOAc); **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 7.81 (d, *J* = 8.3 Hz, 1H), 7.34-7.24 (m, 6H), 7.14 (s, 1H), 6.94 (d, *J* = 8.3 Hz, 1H), 5.96 (bs-s, 2H, N*H*$_2$), 5.46-5.40 (bs-m, 1H, N*H*), 5.09-5.02 (m, 2H), 3.82 (s, 3H), 3.63-3.42 (m, 14H), 3.24 (q, *J* = 6.2 Hz, 2H), 1.81 (p, *J* = 5.5 Hz, 2H), 1.70 (p, *J* = 6.0 Hz, 2H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 167.9, 166.4, 156.3, 150.5, 139.6, 136.6, 131.3, 128.3 (2C), 127.9 (2C), 127.8, 115.4, 113.7, 112.0, 70.7, 70.2, 70.0, 69.9, 69.9, 69.4, 66.3, 51.5, 39.0, 39.0, 29.3, 28.5; **IR** ν_{max} (film)/cm^{-1}: 3460, 3347, 2949, 2870, 1696, 1619, 1532, 1437, 1310, 1243, 1103; **HRMS** (ESI) *m/z* calculated for $C_{27}H_{38}N_3O_8$ ($[M+H]^+$) 532.2653, found 532.2651.

[1] Prepared according to: Harris, T.D.; Kalogeropoulos, S.; Nguyen, T.; Dwyer, G.; Edwards, D.S.; Liu, S.; Bartis, J.; Ellars, C.; Onthank, D.; Yalamanchili, P.; Heminway, S.; Robinson, S.; Lazewatsky, J.; Barrett, J. *Bioconj. Chem.* **2006**, *17*, 1294-1313. **S1:** $\mathbf{R_f}$ 0.05 (9:1, CH_2Cl_2/MeOH); **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 7.25-7.13 (m, 5H), 5.89-5.68 (bs-m, 1H, N*H*), 5.02-4.94 (m, 2H), 3.51-3.34 (m, 12H), 3.20-3.11 (m, 2H), 2.64 (t, *J* = 6.7 Hz, 2H), 1.70-1.49 (m, 6H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 156.1, 136.5, 128.0 (2C), 127.6, 127.5 (2C), 70.1, 70.1, 69.7, 69.7, 69.0, 69.0, 65.8, 39.1, 38.6, 32.6, 29.0; **IR** ν_{max} (film)/cm^{-1}: 3335, 2923, 2867, 1709, 1525, 1455, 1349, 1249, 1103, 1044, 1028; **HRMS** (ESI) *m/z* calculated for $C_{18}H_{31}N_2O_5$ ($[M+H]^+$) 355.2227, found 355.2227.

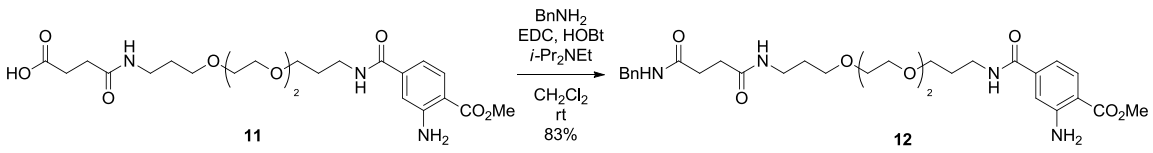
To a solution of carbamate **9** (3.50 g, 6.6 mmol) in MeOH (20 mL) was added Pd/C (100 mg, 10wt% Pd). The mixture was set under vacuum and the flask was purged with nitrogen gas. After applying vacuum again, the flask was backfilled with hydrogen (balloon). After 1h the mixture was filtered over celite and the solvent was evaporated. Amine **10** was obtained in sufficient purity for use in subsequent reactions.

10: $\mathbf{R_f}$ 0.05 (9:1, CH_2Cl_2/MeOH); **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 7.98 (bs-s, 2H, N*H*$_2$), 7.77 (d, *J* = 8.3 Hz, 1H), 7.72-7.64 (m, 1H, N*H*), 7.21 (s, 1H), 6.92 (d, *J* = 8.3 Hz, 1H), 5.45 (bs-s, 2H, N*H*$_2$), 3.82 (s, 3H), 3.60-3.39 (m, 14H), 3.08 (bs-s, 2H), 1.93-1.83 (m, 2H), 1.83-1.74 (m, 2H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 168.0, 167.4, 150.7, 138.9, 131.4, 115.8, 113.7, 112.2, 70.1, 69.8, 69.6 (2C), 69.5, 69.4, 51.6, 39.5, 38.0, 28.9, 26.3; **IR** ν_{max} (film)/cm^{-1}: 3460, 3356, 2875, 1678, 1618, 1594, 1542, 1438, 1312, 1245, 1200, 1180, 1127; **HRMS** (ESI) *m/z* calculated for $C_{19}H_{32}N_3O_6$ ([M+H]$^+$) 398.2286, found 398.2287.

To a solution of amine **10** (1.8 g, 4.53 mmol) in CH_2Cl_2 (20 mL) was added succininc anhydride (453 mg, 4.53 mmol) and triethylamine (1.26 mL, 9.06 mmol). The mixture was allowed to stir over night. The solvent was removed and the residue was directly subjected to flash column chromatography (CH_2Cl_2/MeOH/NEt_3 10:1:0.01) to give acid **11** (1.69 g, 3.40 mmol, 75%).

11: $\mathbf{R_f}$ 0.21 (9:1, CH_2Cl_2/MeOH); **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 11.62 (bs-s, 1H, CO_2*H*), 7.76 (d, *J* = 8.3 Hz, 1H), 7.54 (t, *J* = 5.0 Hz, 1H, N*H*), 7.13 (d, *J* = 1.4 Hz, 1H), 6.98 (bs-s, 1H, N*H*), 6.91 (dd, *J* = 8.4, 1.6 Hz, 1H), 3.78 (s, 3H), 3.60-3.48 (m, 8H), 3.48-3.41 (m, 4H), 3.38 (t, *J* = 6.0 Hz, 2H), 3.19 (q, *J* = 6.3 Hz, 2H), 2.53-2.45 (m, 2H), 2.38 (t, *J* = 6.6 Hz, 2H), 1.80 (p, *J* = 5.8 Hz, 2H), 1.63 (p, *J* = 6.3 Hz, 2H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 177.2, 172.9, 167.9, 166.7, 150.6, 139.5, 131.2, 115.4, 113.7, 111.7, 70.2, 70.1, 70.0, 69.8, 69.7, 69.2, 51.4, 38.5, 37.1, 31.8, 31.5, 28.9, 28.7; **IR** ν_{max} (film)/cm^{-1}: 3454, 3317, 2949, 2869, 1693, 1645, 1620, 1593, 1538, 1437, 1310, 1243, 1191, 1102; **HRMS** (ESI) *m/z* calculated for $C_{23}H_{36}N_3O_9$ ([M+H]$^+$) 498.2446, found 498.2454.

Preparation of Test Substrate 12:



To a solution of acid **11** (100 mg, 0.20 mmol) in CH_2Cl_2 (2 mL) was added benzylamine (26 µL, 0.24 mmol), EDC (46 mg, 0.24 mmol), HOBt (37 mg, 0.24 mmol) and Hünig's base (53 µL, 0.30 mmol). The

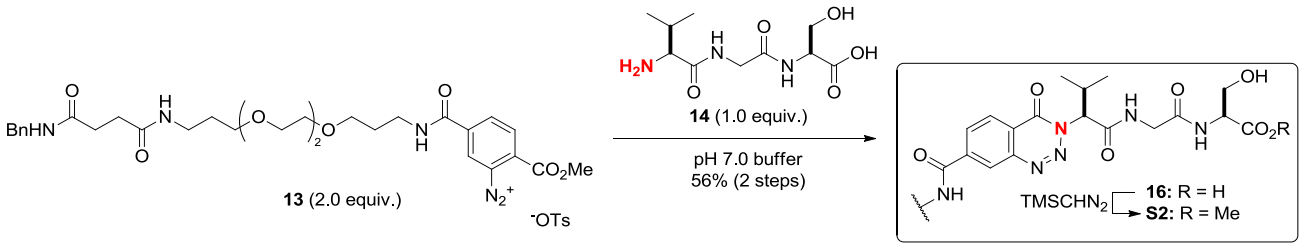
reaction was allowed to stir overnight. The mixture was then diluted with sat. $NaHCO_3$ and CH_2Cl_2. The phases were separated and the aqueous phase was washed three times with CH_2Cl_2. The combined organic phases were washed with brine and dried over $MgSO_4$. The solvent was removed and the residue was subjected to flash column chromatography (CH_2Cl_2/MeOH 20:1 → 10:1) to give amide **12** (98 mg, 0.17 mmol, 83%).

12: $\mathbf{R_f}$ 0.50 (9:1, CH_2Cl_2/MeOH); **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 7.78 (d, *J* = 8.3 Hz, 1H), 7.48 (bs-t, *J* = 5.4 Hz, 1H, N*H*), 7.31-7.13 (m, 7H), 6.90 (dd, *J* = 8.3, 1.6 Hz, 1H), 6.90-6.86 (bs-m, 1H, N*H*), 6.07 (bs-s, 2H, NH_2), 4.28 (d, *J* = 5.8 Hz, 2H), 3.80 (s, 3H), 3.61-3.51 (m, 8H), 3.46-3.35 (m, 6H), 3.18 (q, *J* = 6.4 Hz, 2H), 2.46 (qt, *J* = 9.2, 4.6 Hz, 4H), 1.79 (p, *J* = 5.9 Hz, 2H), 1.62 (p, *J* = 6.2 Hz, 2H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 172.3, 172.3, 167.9, 166.9, 150.6, 139.3, 138.3, 131.3, 128.3 (2C), 127.3 (2C), 127.0, 115.5, 113.6, 111.9, 70.1, 70.1, 70.0, 69.7, 69.7, 69.2, 51.5, 43.2, 38.4, 37.3, 31.4 (2C), 28.8, 28.7; **IR** ν_{max} (film)/cm^{-1}: 3298, 2866, 1630, 1540, 1492, 1474, 1436, 1340, 1310, 1247, 1216, 1194, 1137, 1106; **HRMS** (ESI) *m/z* calculated for $C_{30}H_{43}N_4O_8$ ($[M+H]^+$) 587.3075, found 587.3066.

3. Preliminary Studies and Optimization

3.1. Preliminary Studies

We first tested the reaction of tripeptide **14** (H_2N-VGS-CO_2H) with diazonium salt **13**. The diazonium reagent was prepared from aniline **12** ($NaNO_2$, *p*-TsOH). 2 equivalents of **13** were added to a solution of peptide **14** in pH 7.0 buffer (100 mM NaH_2PO_4) and the reaction was monitored by LC-MS analysis. Formation of a single product with a mass corresponding to benzotriazeneone **16** was observed. After 2 hours the product was isolated by extraction. For easier purification, the carboxylic acid **16** was transformed into the corresponding methyl ester **S2** by treatment with $TMSCHN_2$. Purification by column chromatography provided benzotriazeneone **S2** as the only product in 56% yield.



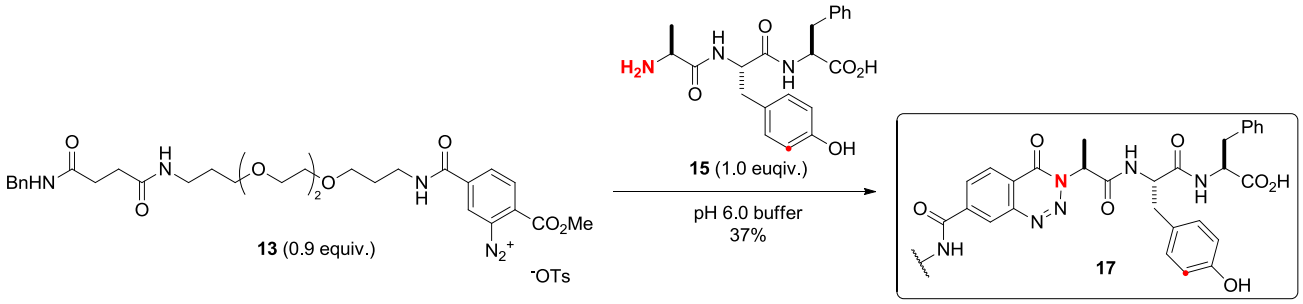
Scheme S1. Reaction of diazonium salt **13** with tripeptide **14**.

To a solution of aniline **12** (22 mg, 38 µmol) in MeOH/water (1 mL, 1:3) was added *p*-toluenesulfonic acid (44 mg, 230 µmol) followed by sodium nitrite (11 mg, 153 µmol). The solution was stirred for 30 min. In a separate flask tripeptide **14** (5 mg, 19 µmol) was dissolved in pH 7 buffer (5 mL, 100 mM NaH_2PO_4). The diazonium solution was transferred to this flask and the reaction was allowed to stir at ambient temperature for 2 h. EtOAc was added to this mixture and the aqueous phase was extracted three times with EtOAc. The combined organic layers were dried over $MgSO_4$ and the solvent was removed. The crude product was dissolved in CH_2Cl_2/MeOH (1 mL, 10:1) and trimethylsilyl diazomethane (excess) was added. After 15 min AcOH was added until the mixture turned form yellow to colorless. The solvent was removed and the residue was purified by flash column chromatography (CH_2Cl_2/MeOH 9:1) to give ester **S2** (9 mg, 11 µmol, 56%).

S2: $\mathbf{R_f}$ 0.22 (9:1, CH_2Cl_2/MeOH); **^{1}H-NMR** (600 MHz, CD_3OD): δ 8.60 (dd, *J* = 1.7, 0.6 Hz, 1H), 8.39 (dd, *J* = 8.3, 0.5 Hz, 1H), 8.28 (dd, *J* = 8.3, 1.7 Hz, 1H), 7.30-7.24 (m, 4H), 7.23-7.20 (m, 1H), 5.31 (d, *J* = 9.7 Hz, 1H), 4.50 (t, *J* = 4.4 Hz, 1H), 3.34 (s, 2H), 3.99 (d, *J* = 16.8 Hz, 1H), 3.92 (d, *J* = 16.8 Hz, 1H), 3.86 (dd, *J* = 11.4, 4.7 Hz, 1H), 3.77 (dd, *J* = 11.3, 4.1 Hz, 1H), 3.70 (s, 3H), 3.67-3.65 (m, 2H), 3.64-3.61 (m,

6H), 3.57-3.53 (m, 4H), 3.47 (t, *J* = 6.2 Hz, 2H), 3.22 (t, *J* = 6.8 Hz, 2H), 3.03-2.94 (m, 1H), 2.54-2.50 (m, 2H), 2.50-2.47 (m, 2H), 1.95-1.90 (m, 2H), 1.73-1.68 (m, 2H), 1.21 (d, *J* = 6.7 Hz, 3H), 0.97 (d, *J* = 6.6 Hz, 2H); **^{13}C-NMR** (150 MHz, CD_3OD): δ 174.5, 174.5, 172.0, 171.3, 171.1, 167.6, 157.2, 144.9, 142.6, 140.0, 132.2, 129.5 (2C), 128.5 (2C), 128.2, 128.1, 126.8, 122.2, 71.6, 71.5, 71.3, 71.2, 70.2, 69.9, 68.2, 62.8, 56.2, 52.8, 44.1, 43.6, 39.1, 37.9, 32.3, 32.3, 30.4, 30.3, 30.0, 20.1, 19.8; **IR** ν_{max} (film)/cm^{-1}: 3313, 2926, 1743, 1651, 1548, 1454, 1087; **HRMS** (ESI) *m/z* calculated for $C_{40}H_{57}N_8O_{12}$ ($[M+H]^+$) 841.4090, found 841.4088.

In order to assess the reactivity pattern of the reported bioconjugation protocol towards tyrosine residues, we tested tyrosine containing tripeptide **15** as a substrate in the reaction with diazonium salt **13** at pH 6.0 (Scheme S2). LS-MS analysis of this reaction indicated formation of amide **17** as the major product (Figure S1). Only trace amounts of diazo compound **S3** could be detected, probably as a result of a reaction of the product **17** with excess reagent. Isolation and characterization of amide **17** confirmed this analysis.



Scheme S2. Reaction of tripeptide **15** with diazonium salt **13**.

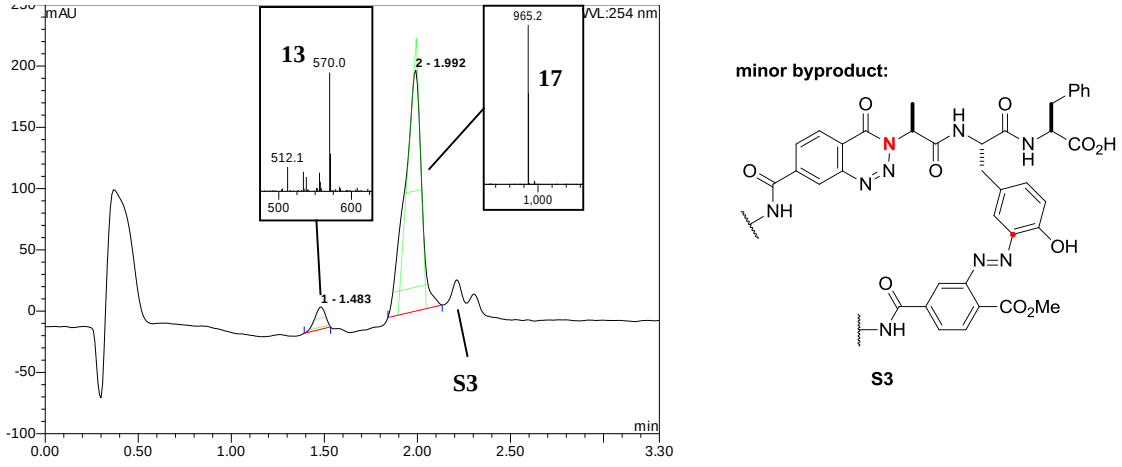


Figure S1. LC-MS trace (UV at 254 nm) of the reaction of **13** with tripeptide **15** after 2 h.

To a solution of aniline **12** (10 mg, 17 μmol) in MeOH/water (1 mL, 1:3) was added *p*-toluenesulfonic acid (20 mg, 102 μmol) followed by sodium nitrite (5 mg, 68 μmol). The solution was stirred for 30 min. In a separate flask tripeptide **15** (7.5 mg, 19 μmol) was dissolved in pH 6.0 buffer (5 mL, 100 mM NaH_2PO_4). The diazonium solution was transferred to this flask and the reaction was allowed to stir at ambient temperature for 2 h while the reaction was monitored by LC-MS (Figure S1). EtOAc was added to this mixture and the aqueous phase was extracted three times with EtOAc. The combined organic layers were dried over $MgSO_4$ and the solvent was removed. The crude product was purified by prep-HPLC to give acid **17** (6.1 mg, 6.3 μmol, 37%).[2]

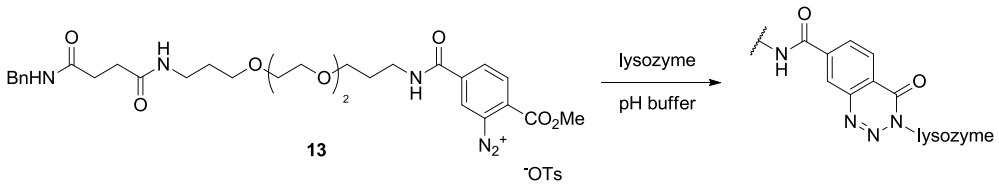
17: R_f 0.00 (9:1, CH_2Cl_2/MeOH); **^{1}H-NMR** (600 MHz, CD_3OD): δ 8.58 (dd, *J* = 1.7, 0.6 Hz, 1H), 8.34 (dd, *J* = 8.3, 0.5 Hz, 1H), 8.26 (dd, *J* = 8.3, 1.7 Hz, 1H), 7.30-7.17 (m, 10H), 6.95-6.92 (m, 2H), 6.52-6.49 (m, 2H), 5.57 (q, *J* = 7.2 Hz, 1H), 4.64 (dd, *J* = 8.3, 5.3 Hz, 1H), 4.56 (dd, *J* = 8.8, 5.5 Hz, 1H), 4.34 (s, 2H), 3.67-3.65 (m, 2H), 3.64-3.61 (m, 6H), 3.57-3.53 (m, 4H), 3.46 (t, *J* = 6.2 Hz, 2H), 3.23-3.18 (m, 3H), 3.03-2.97 (m, 2H), 2.78 (dd, *J* = 14.1, 8.8 Hz, 1H), 2.53-2.50 (m, 2H), 2.49-2.46 (m, 2H), 1.96-1.91 (m, 2H), 1.74 (d, *J* = 7.2 Hz, 3H), 1.70 (t, *J* = 6.3 Hz, 2H); **^{13}C-NMR** (150 MHz, CD_3OD): δ 174.5, 174.5, 174.2, 173.1, 171.6, 167.7, 157.1, 156.4, 145.1, 142.4, 140.0, 138.3, 132.1, 131.20 (2C), 130.4 (2C), 129.5 (2C), 129.5 (2C), 128.7, 128.5 (2C), 128.2, 128.2, 127.8, 126.6, 122.3, 116.1 (2C), 71.5, 71.5, 71.3, 71.2, 70.2, 69.9, 58.5, 56.2, 55.1, 44.1, 39.1, 38.4, 37.9, 37.6, 32.3, 32.3, 30.4, 30.3, 16.3; **IR** ν_{max} (film)/cm^{-1}: 3300, 2927, 1640, 1553, 1517, 1455, 1204, 1137; **HRMS** (ESI) *m/z* calculated for $C_{50}H_{61}N_8O_{12}$ ($[M+H]^+$) 965.4403, found 965.4402.

3.2. Optimization

For optimization of the reported bioconjugation protocol for protein substrates, we first evaluated different buffers as solvents in the reaction of diazonium tosylate **13** with lysozyme[3] (Table S1). The reaction proceeds with low efficiency at pH 5.0 (entries 1 and 5). The best results were obtained using phosphate buffers with pH 6.5 or 7.0 (entries 7 and 8). Most importantly, for all reactions only the desired amine coupling could be observed. We never detected products resulting from the reaction of tyrosine, histidine or tryptophane residues with the diazonium reagent.

[2] The low yield can be attributed to difficulties in purifying the product.
[3] For the subsequent optimization studies, myoglobin was used as model substrate as this protein proved more sensitive to changes in reaction conditions.

Table S1. Reaction of diazonium salt **13** with lysozyme in different buffer solutions.



Entry[a]	buffer[b]	unmod. (%)[b]	+1 (%)	+2 (%)	+3 (%)	+4 (%)
1	pH 5.0 NaOAc	50	38	13	0	0
2	pH 6.0 NaOAc	30	40	22	8	0
3	pH 6.5 NaOAc	27	43	26	4	0
4	pH 7.0 NaOAc	24	39	26	9	2
5	pH 5.0 NaH_2PO_4	71	29	0	0	0
6	pH 6.0 NaH_2PO_4	28	45	26	0	0
7	pH 6.5 NaH_2PO_4	4	32	43	18	2
8	pH 7.0 NaH_2PO_4	5	26	47	22	0

[a] conditions: 500 µM lysozyme, 10 equiv. diazonium salt **13** (5 mM), 2 h; [b] 100 mM salt; [c] determined by ESI-MS analysis through the relative ratio of the corresponding peak intensities; unmod. = unmodified, +1 = singly modified protein, etc.

As outlined in Table S2, the concentration dependency of the model reaction was tested with varying concentrations of diazonium reagent and myoglobin as model substrate. At protein concentrations of 10 µM a larger excess of reagent is needed to achieve full conversion to modified products.

Table S2. Concentration dependency of the diazonium coupling to myoglobin.

Entry[a]	myoglobin concentration	equiv. reagent 13[b]	unmod. (%)[b]	+1 (%)	+2 (%)	+3 (%)	+4 (%)	+5 (%)	+6 (%)	+7 (%)
1	100 µM	5	16	29	26	17	8	3	0	0
2	100 µM	10	1	5	12	19	22	20	13	7
5	10 µM	5	64	30	6	0	0	0	0	0
6	10 µM	10	41	37	17	5	0	0	0	0
7	10 µM	50	0	6	12	18	21	18	14	7

[a] conditions: 100 mM pH 7.0 NaH_2PO_4 buffer (500 µL), 2 h; [b] equivalents of reagent in respect to protein; [c] determined by ESI-MS analysis through the relative ratio of the corresponding peak intensities; unmod. = unmodified, +1 = singly modified protein, etc.

Table S3 shows the conversion of the reaction after different time periods. The optimal result was obtained after 2 hours of reaction time (entry 3). Prolonged reaction did not lead to any significant improvement of the conversion (data not shown). Very good degree of modification can already be obtained after 30 min of reaction time (entry 2).

Table S3. Diazonium bioconjugation to myoglobin with different reaction times.

Entry[a]	**reaction time**	**unmod. (%)**[b]	**+1 (%)**	**+2 (%)**	**+3 (%)**	**+4 (%)**	**+5 (%)**	**+6 (%)**	**+7 (%)**
1	10 min	24	35	24	13	4	0	0	0
2	30 min	7	19	24	22	14	7	4	1
3	1 h	4	16	22	23	18	10	4	2
4	2 h	1	5	12	19	22	20	13	7

[a] conditions: 100 µM myoglobin, 10 equiv. diazonium salt **13** (1 mM), 100 mM pH 7 NaH_2PO_4 buffer (500 µL); [b] determined by ESI-MS analysis through the relative ratio of the corresponding peak intensities; unmod. = unmodified, +1 = singly modified protein, etc.

4. Protein Bioconjugation

General Procedure for the Bioconjugation of Diazonium Salts to Proteins:

A 25 mM stock solution of the diazonium salt was prepared by mixing the aniline substrate (12.5 μmol) with a solution of *p*-TsOH monohydrate (14 mg, 75 μmol) in MeOH (125 μL) (stock solution). Water (250 μL) was added followed by a solution of $NaNO_2$ (3.35 mg, 50 μmol) in water (125 μL) (stock solution). The mixture was allowed to stir for 15 min at ambient temperature before use. The conversion of the diazotization reaction could be easily checked by LC-MS analysis.

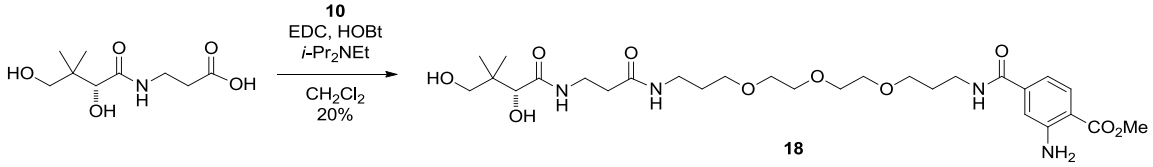
An Eppendorf tube was charged with a solution of protein (0.05 μmol) in pH 7.0 buffer (500 μL, 100 mM NaH_2PO_4) and the freshly prepared solution of diazonium salt (20 μL, 0.5 μmol, 25 mM stock solution) was added. The reaction was left standing for 2 h at ambient temperature. The reaction mixture was then filtered through a Nap-5 column (illustra Nap-5, GE-Healthcare; eluent: H_2O) and the resulting protein solution was analyzed by ESI-MS. ESI samples were prepared by mixing 50 μL of the protein solution with 50 μL of a 0.1 % formic acid stock solution (75% MeCN, 25% H_2O, 0.1 % HCO_2H).

ESI-MS conditions:

The mass spectrometry was performed on an ESI-Q-TOF system (maXis, Bruker Daltonics, Germany) coupled with an Agilent 1200 system (Agilent Ltd., Germany). The MS instrument was operated in wide pass quadrupole mode, for MS experiments, with the TOF data being collected between m/z 100-5000 with low-collision energy of 10 eV. The optimized source conditions were drying gas 8.0 l/h (nitrogen 99,99 % purity) at a temperature of 200°C, nebulizer pressure 1.6 bar, capillary and endplate voltages 500 and 4500 V, respectively, TOF flight tube voltage 9870 V, reflection voltage 1999 V, pusher voltage 1642 V and MCP detector voltage 1554 V. The resolving power of the instrument was around 35‘000 with 2.5 Hz spectra rate. The ESI-TOF mass spectrometer was calibrated routinely for flow injection analysis (FIA) in the positive electrospray ionization mode using Agilent-ESI-TOF tuning mix on the quadratic algorithmic mode.

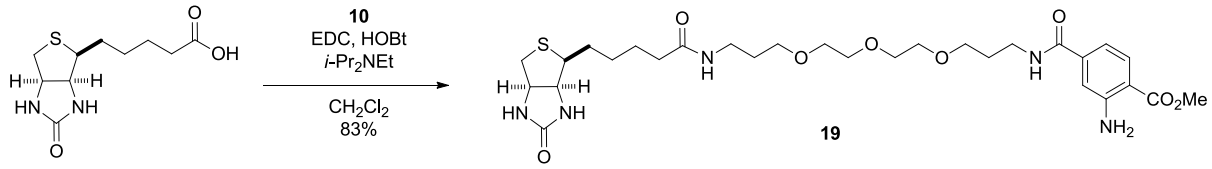
Further data processing was carried out using Data Analysis 4.0 software (Bruker Daltonics, Germany) in combination with the deconvolution algorithm MaxEnt (Maximum Entropy, Spectrum Square Associates Inc.).

Preparation of Natural Product Derivatized Aniline Substrates:



To a solution of calcium pantothenate (50 mg, 0.23 mmol) in CH_2Cl_2 (2 mL) was added amine **10** (118 mg, 0.30 mmol), EDC (53 mg, 0.27 mmol), HOBt (42 mg, 0.27 mmol) and diisopropyl ethylamine (60 μL, 0.34 mmol). The reaction was stirred overnight. The solution was directly subjected to flash column chromatography (CH_2Cl_2/MeOH 10:1 → 5:1) followed by preparative TLC (CH_2Cl_2/MeOH 7:1) to give amide **18** (27 mg, 0.05 mmol, 20%).

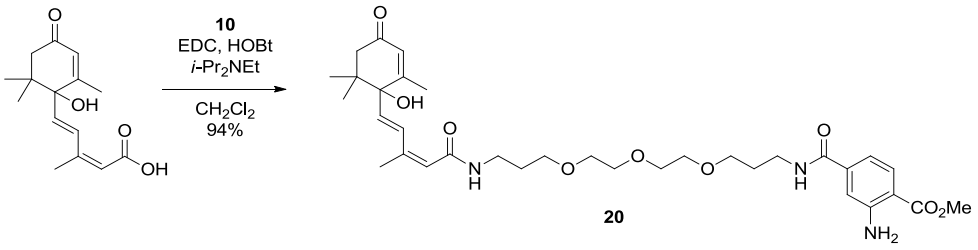
R_f 0.50 (5:1, CH_2Cl_2/MeOH); **$[\alpha]_D^{22.4°}$** +15.08 (*c* 1.35, $CHCl_3$); **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 7.85 (d, *J* = 8.3 Hz, 1H), 7.46-7.39 (m, 2H, N*H*), 7.17 (d, *J* = 1.5 Hz, 1H), 6.96 (dd, *J* = 8.3, 1.6 Hz, 1H), 6.61 (t, *J* = 5.3 Hz, 1H, N*H*), 6.08 (bs-s, 2H, NH_2), 4.41 (bs-s, 1H, O*H*), 3.98 (s, 1H), 3.87 (s, 3H), 3.67-3.56 (m, 9H), 3.56-3.42 (m, 10H), 3.27 (p, *J* = 6.3 Hz, 2H), 2.36 (t, *J* = 6.0 Hz, 2H), 1.87 (p, *J* = 5.7 Hz, 2H), 1.68 (p, *J* = 6.3 Hz, 2H), 0.98 (s, 3H), 0.90 (s, 3H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 173.7, 171.4, 168.1, 166.9, 150.7, 139.5, 131.6, 115.6, 113.9, 112.3, 77.6, 70.9, 70.7, 70.2, 70.1, 69.9, 69.9, 69.8, 51.7, 39.3, 39.0, 37.7, 35.8, 35.3, 28.9, 28.8, 21.6, 20.4; **IR** ν_{max} (film)/cm^{-1}: 3343, 2873, 1696, 1647, 1542, 1439, 1312, 1247, 1192, 1104; **HRMS** (ESI) *m/z* calculated for $C_{28}H_{47}N_4O_{10}$ ($[M+H]^+$) 599.3287, found 599.3286.



To a solution of D-biotin (50 mg, 0.21 mmol) in CH_2Cl_2 (2 mL) was added amine **10** (106 mg, 0.27 mmol), EDC (47 mg, 0.25 mmol), HOBt (38 mg, 0.25 mmol) and diisopropyl ethylamine (54 μL, 0.31 mmol). The reaction was stirred overnight. The solution was directly subjected to flash column chromatography (CH_2Cl_2/MeOH 20:1) to give amide **19** (106 mg, 0.17 mmol, 83%).

R_f 0.14 (9:1, CH_2Cl_2/MeOH); **$[\alpha]_D^{22.3°}$** +22.76 (*c* 1.00, CH_3OH); **^{1}H-NMR** (400 MHz, CD_3OD): δ 7.80 (d, *J* = 8.4 Hz, 1H), 7.15 (d, *J* = 1.6 Hz, 1H), 6.89 (dd, *J* = 8.4, 1.7 Hz, 1H), 4.45 (dd, *J* = 7.7, 4.6 Hz, 1H), 4.25 (dd, *J* = 7.8, 4.4 Hz, 1H), 3.82 (s, 3H), 3.62-3.48 (m, 10H), 3.46-3.39 (m, 4H), 3.22-3.11 (m, 5H), 2.87 (dd, *J* = 12.7, 4.9 Hz, 1H), 2.66 (d, *J* = 12.7 Hz, 1H), 2.15 (t, *J* = 7.4 Hz, 2H), 1.83 (p, *J* = 6.4 Hz, 2H), 1.74-

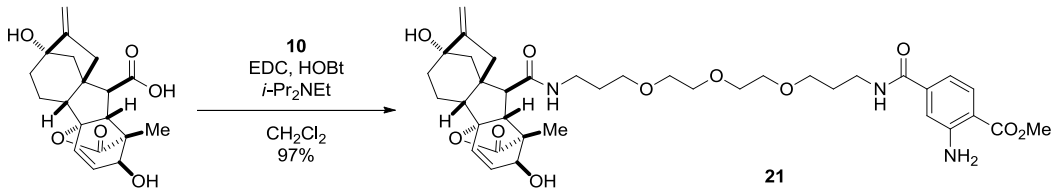
1.47 (m, 6H); **^{13}C-NMR** (100 MHz, CD_3OD): δ 175.8, 169.5, 169.3, 165.9, 152.5, 140.9, 132.5, 116.7, 114.4, 112.8, 71.4, 71.4, 71.2, 71.1, 70.3, 69.9, 63.3, 61.5, 57.0, 52.2, 43.8, 41.1, 38.7, 36.8, 30.4, 30.3, 29.8, 29.5, 26.9; **IR** ν_{max} (film)/cm^{-1}: 3298, 2941, 1695, 1644, 1543, 1438, 1312, 1246, 1109; **HRMS** (ESI) *m/z* calculated for $C_{29}H_{46}N_5O_8S$ ($[M+H]^+$) 624.3062, found 624.3059.



20

To a solution of abscisic acid (50 mg, 0.19 mmol) in CH_2Cl_2 (2 mL) was added amine **10** (98 mg, 0.25 mmol), EDC (44 mg, 0.23 mmol), HOBt (35 mg, 0.23 mmol) and diisopropyl ethylamine (50 μL, 0.28 mmol). The reaction was stirred overnight. The solution was directly subjected to flash column chromatography (CH_2Cl_2/MeOH 20:1) to give amide **20** (115 mg, 0.18 mmol, 94%).

R_f 0.45 (9:1, CH_2Cl_2/MeOH; **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 7.86 (d, *J* = 16.1 Hz, 1H) 7.79 (d, *J* = 8.3 Hz, 1H) 7.31 (t, *J* = 5.1 Hz, 1H, N*H*), 7.14 (d, *J* = 1.5 Hz, 1H), 6.92 (dd, *J* = 8.3, 1.6 Hz, 1H), 6.47 (t, *J* = 5.5 Hz, 1H, N*H*), 6.03 (bs-s, 2H, NH_2), 5.99 (d, *J* = 16.1 Hz, 1H), 5.83 (s, 1H), 5.59 (s, 1H), 3.81 (s, 3H), 3.63-3.54 (m, 8H), 3.51-3.40 (m, 6H), 3.28 (q, *J* = 6.3 Hz, 2H), 3.20 (bs-s, 1H, O*H*), 2.40 (d, *J* = 17.0 Hz, 1H), 2.20 (d, *J* = 17.0 Hz, 1H), 1.89-1.78 (m, 8H), 1.68 (p, *J* = 6.1 Hz, 2H), 1.02 (s, 3H), 0.93 (s, 3H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 198.0, 167.9, 166.6, 166.1, 163.5, 150.5, 144.0, 139.5, 134.6, 131.3, 128.2, 126.4, 121.7, 115.4, 113.8, 112.0, 79.3, 70.6, 70.1, 69.9, 69.8, 69.7, 69.6, 51.5, 49.7, 41.4, 38.9, 37.3, 28.9, 28.6, 24.2, 23.0, 20.8, 18.9; **IR** ν_{max} (film)/cm^{-1}: 3458, 3344, 2949, 2870, 1694, 1243, 1620, 1597, 1534, 1310, 1243, 1101; **HRMS** (ESI) *m/z* calculated for $C_{34}H_{50}N_3O_9$ ($[M+H]^+$) 644.3542, found 644.3536.

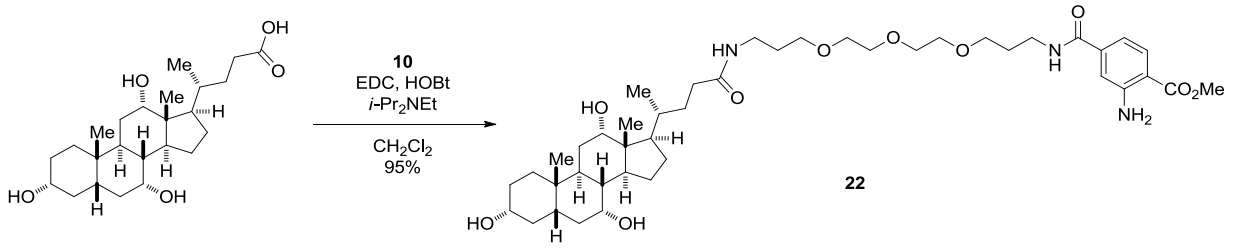


21

To a solution of gibberellic acid (50 mg, 0.14 mmol) in CH_2Cl_2 (2 mL) was added amine **10** (75 mg, 0.19 mmol), EDC (33 mg, 0.17 mmol), HOBt (27 mg, 0.17 mmol) and diisopropyl ethylamine (38 μL, 0.22

mmol). The reaction was stirred overnight. The solution was directly subjected to flash column chromatography (CH_2Cl_2/MeOH 20:1) to give amide **21** (102 mg, 0.14 mmol, 97%).

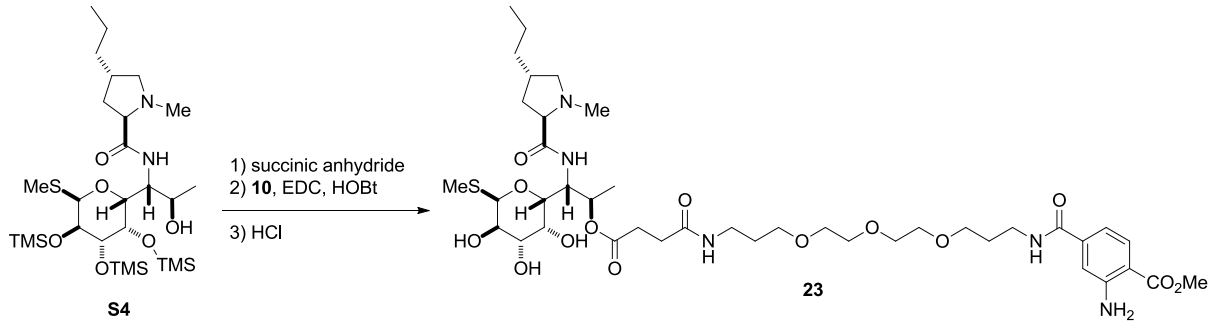
R_f 0.20 (9:1, CH_2Cl_2/MeOH); **$[\alpha]_D^{22.2°}$** +41.24 (*c* 1.00, CH_3OH); **^{1}H-NMR** (400 MHz, CD_3OD): δ 7.88 (d, *J* = 8.4 Hz, 1H) 7.20 (d, *J* = 1.6 Hz, 1H), 6.95 (dd, *J* = 8.4, 1.7 Hz, 1H), 6.39 (d, *J* = 9.4 Hz, 1H), 5.90 (dd, *J* = 9.3, 3.6 Hz, 1H), 5.22 (s, 1H), 4.95 (s, 1H), 4.03 (d, *J* = 3.5 Hz, 1H), 3.90 (s, 3H), 3.70-3.60 (m, 8H), 3.59-3.47 (m, 6H), 3.37-3.27 (m, 3H), 2.65 (d, *J* = 10.2 Hz, 1H), 2.30-2.18 (m, 2H), 2.06-1.86 (m, 6H), 1.85-1.68 (m, 5H), 1.42-1.38 (m, 1H), 1.25 (s, 3H); **^{13}C-NMR** (100 MHz, CD_3OD): δ 181.3, 173.4, 169.6, 169.3, 158.2, 152.5, 140.9, 134.1, 133.2, 132.5, 116.7, 114.3, 112.9, 107.3, 92.8, 79.4, 78.7, 71.5, 71.4, 71.2, 70.6, 70.3, 69.7, 55.0, 54.2, 53.5, 52.1, 52.0, 51.6, 46.0, 44.5, 39.7, 38.8, 37.8, 30.5, 30.3, 18.0, 15.0; **IR** ν_{max} (film)/cm^{-1}: 3356, 2936, 2874, 1756, 1641, 1621, 1542, 1439, 1312, 1246, 1101, 1046; **HRMS** (ESI) *m/z* calculated for $C_{38}H_{52}N_3O_{11}$ ($[M+H]^+$) 726.3596, found 726.3592.



To a solution of cholic acid (50 mg, 0.12 mmol) in CH_2Cl_2 (2 mL) was added amine **10** (63 mg, 0.16 mmol), EDC (28 mg, 0.15 mmol), HOBt (23 mg, 0.15 mg) and diisopropyl ethylamine (32 μL, 0.18 mmol). The reaction was stirred overnight. The reaction was quenched by addition of sat. $NaHCO_3$. The phases were separated and the aqueous phase was extracted three times with CH_2Cl_2. The combined organic layers were washed with brine and dried over $MgSO_4$. The solvent was removed and the residue was purified by flash column chromatography (CH_2Cl_2/MeOH 20:1) to give amide **22** (92 mg, 0.12 mmol, 95%).

R_f 0.29 (9:1, CH_2Cl_2/MeOH); **$[\alpha]_D^{23.7°}$** +9.24 (*c* 2.40, $CHCl_3$); **^{1}H-NMR** (400 MHz, d_6-DMSO): δ 8.40 (bs-t, *J* = 5.5 Hz, 1H, N*H*), 7.75-7.70 (m, 2H), 7.21 (d, *J* = 1.6 Hz, 1H), 6.90 (dd, *J* = 8.4, 1.6 Hz, 1H), 6.75 (bs-s, 2H, NH_2), 4.15-4.05 (bs-m, 2H, O*H*), 3.99 (bs-d, *J* = 2.7 Hz, 1H, O*H*), 3.80 (s, 3H), 3.77 (bs-s, 1H), 3.60 (bs-s, 1H), 3.54-3.43 (m, 8H), 3.39-3.24 (m, 6H), 3.21-3.17 (m, 2H), 3.05 (q, *J* = 6.3 Hz, 2H), 2.26-1.91 (m, 4H), 1.82-1.56 (m, 11H), 1.48-1.11 (m, 10H), 0.99-0.89 (m, 4H), 0.88-0.78 (m, 4H), 0.57 (s, 3H); **^{13}C-NMR** (100 MHz, d_6-DMSO): δ 172.5, 167.4, 165.9, 151.0, 139.8, 130.7, 115.8, 112.8, 110.2, 71.0, 70.4, 69.8, 69.7, 69.6, 69.5, 68.3, 68.1, 66.2, 51.6, 48.6, 46.2, 45.7, 41.5, 41.4, 39.5, 36.7, 35.7, 35.3, 35.1, 34.9, 34.4, 32.6, 31.8, 30.4, 29.4, 29.3, 28.6, 27.3, 26.2, 22.8, 22.6, 17.1, 12.3; **IR** ν_{max} (film)/cm^{-1}: 3364, 2936,

2870, 1686, 1621, 1553, 1440, 1247, 1203, 1138; **HRMS** (MALDI) *m/z* calculated for $C_{43}H_{70}N_3O_{10}$ ($[M+H]^+$) 788.5056, found 788.5055.



To a solution of 2,3,4-tri-*O*-trimethylsilyl lincomycin (**S4**)[4] (200 mg, 0.32 mmol) in THF (2 mL) was added sodium hydride (13 mg, 0.32 mmol, 60% dispersion in mineral oil). After 15 min succinic anhydride (39 mg, 0.39 mmol) was added and the mixture was heated to 60 °C overnight. The reaction was quenched by addition of MeOH and the solvent was removed. The residue was directly subjected to flash column chromatography (CH_2Cl_2/MeOH 10:1 → 5:1) to give the corresponding acid (113 mg, 0.16 mmol, 49%).

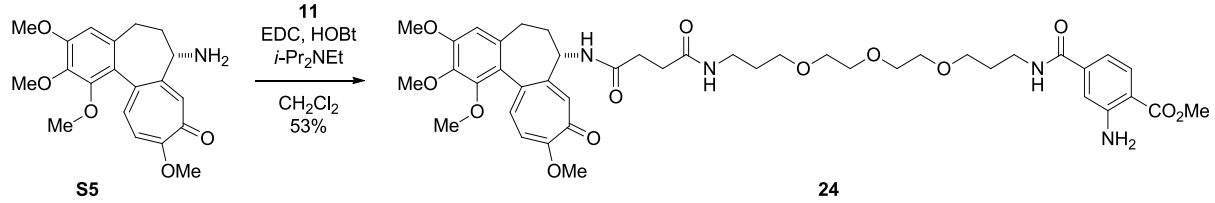
To a solution of this acid (48 mg, 0.07 mmol) in CH_2Cl_2 (2 mL) was added amine **10** (34 mg, 0.09 mmol), EDC (15 mg, 0.08 mmol), HOBt (12 mg, 0.08 mmol) and diisopropyl ethylamine (17 µL, 0.10 mmol). The mixture was allowed to stir over night and the solvent was removed. The residue was purified by flash column chromatography (CH_2Cl_2/MeOH 20:1) to give the corresponding amide (53 mg, 0.05 mmol, 72%).

To a solution of this amide (53 mg, 0.05 mmol) in CH_2Cl_2 (2 mL) was added HCl in MeOH (96 µL, 0.05 mmol, 0.5 M in MeOH). After 15 min TLC and LC-MS analysis indicated full conversion to the product. The solvent was removed and the residue was subjected to flash column chromatography (CH_2Cl_2/MeOH 10:1 → 5:1) to give aniline **23** (43 mg, 0.05 mmol, 100%).

R_f 0.26 (9:1, CH_2Cl_2/MeOH); **$[\alpha]_D^{22.4°}$** +66.16 (*c* 2.15, $CHCl_3$); **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 7.98 (d, *J* = 6.0 Hz, 1H, N*H*), 7.84 (d, *J* = 8.3 Hz, 1H), 7.33 (t, *J* = 4.6 Hz, 1H, N*H*), 7.16 (d, *J* = 1.3 Hz, 1H), 6.96 (dd, *J* = 8.3, 1.6 Hz, 1H), 6.56 (bs-s, 1H, N*H*), 6.05 (bs-s, 2H, NH_2), 5.34-5.25 (m, 2H), 4.90 (bs-s, 1H, O*H*), 4.35 (t, *J* = 8.2 Hz, 1H), 4.12 (dd, *J* = 9.8, 5.5 Hz, 1H), 3.89 (d, *J* = 10.2 Hz, 1H), 3.85 (s, 3H), 3.70-3.57 (m, 10H), 3.56-3.43 (m, 8H), 3.28 (q, *J* = 6.2 Hz, 2H), 3.23-2.88 (bs-m, 4H), 2.62 (dt, *J* = 15.5, 7.8 Hz, 1H), 2.54-2.35 (m, 4H), 2.16-2.00 (m, 6H), 1.95-1.81 (m, 4H), 1.70 (p, *J* = 6.1 Hz, 2H), 1.31-1.22 (m, 7H), 0.91-0.84 (m, 3H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 177.1, 172.1, 171.2, 168.1, 166.8, 150.7, 139.6, 131.5, 115.5, 113.9, 112.1, 88.5, 71.0, 70.6, 70.3, 70.2, 70.1, 69.9, 69.8, 69.7, 69.2, 68.9, 68.4, 68.2, 62.6,

[4] Prepared according to: PCT Int. Appl., WO 2008146919.

51.6, 50.8, 41.5, 38.9, 37.7, 37.6, 37.5, 35.7, 30.8, 29.7, 28.9, 28.7, 21.5, 14.2, 14.1, 14.1; **IR** ν_{max} (film)/cm^{-1}: 3340, 2923, 2872, 1733, 1695, 1648, 1620, 1595, 1530, 1438, 1311, 1245, 1189, 1096, 1051; **HRMS** (ESI) *m/z* calculated for $C_{41}H_{68}N_5O_{14}S$ ([M+H]$^+$) 886.4478, found 886.4462.



To a solution of deacetoxy colchicine (**S5**)[5] (30 mg, 84 μmol) in CH_2Cl_2 (2 mL) was added acid **11** (50 mg, 0.10 mmol), EDC (19 mg, 0.10 mmol), HOBt (15 mg, 0.10 mmol) and diisopropyl ethylamine (22 μL, 0.13 mmol). The mixture was allowed to stir overnight and the solvent was removed. The residue was purified by flash column chromatography (CH_2Cl_2/MeOH 20:1 → 10:1) followed by preparative TLC (CH_2Cl_2/MeOH 10:1) to give amide **24** (37 mg, 44 μmol, 53%).

R$_f$ 0.32 (9:1, CH_2Cl_2/MeOH); **[α]$_D^{24.1°}$** -9.12 (*c* 1.20, $CHCl_3$); **^{1}H-NMR** (400 MHz, $CDCl_3$): δ 7.82 (d, *J* = 8.3 Hz, 1H), 7.65 (d, *J* = 6.7 Hz, 1H, N*H*), 7.43 (t, *J* = 5.1 Hz, 1H, N*H*), 7.38 (s, 1H), 7.27-7.22 (m, 1H), 7.17 (d, *J* = 1.5 Hz, 1H), 6.94 (dd, *J* = 8.3, 1.6 Hz, 1H), 6.78 (d, *J* = 10.9 Hz, 1H), 6.62 (t, *J* = 5.4 Hz, 1H, N*H*), 6.50 (s, 1H), 6.05 (bs-s, 2H, NH_2), 4.54 (dt, *J* = 12.7, 6.6 Hz, 1H), 3.94 (s, 3H), 3.91 (s, 3H), 3.88 (s, 3H), 3.85 (s, 3H), 3.65-3.55 (m, 11H), 3.54-3.42 (m, 6H), 3.33-3.27 (m, 2H), 2.55-2.32 (m, 6H), 2.20 (tt, *J* = 12.8, 6.5 Hz, 1H), 1.89-1.78 (m, 3H), 1.69 (p, *J* = 6.1 Hz, 2H); **^{13}C-NMR** (100 MHz, $CDCl_3$): δ 179.4, 172.1, 171.9, 168.0, 166.8, 163.9, 153.4, 151.2, 151.1, 150.7, 141.5, 139.6, 136.3, 134.9, 134.2, 131.4, 130.8, 125.6, 115.6, 113.8, 112.1, 112.0, 107.3, 70.6, 70.3, 70.1, 69.9, 69.9, 69.6, 61.4, 61.3, 56.2, 56.1, 52.2, 51.6, 38.8, 37.8, 36.5, 31.6, 31.4, 29.8, 28.8 (2C); **IR** ν_{max} (film)/cm^{-1}: 3457, 3303, 2936, 2868, 1648, 1617, 1589, 1546, 1488, 1322, 1249, 1095, 1018; **HRMS** (ESI) *m/z* calculated for $C_{43}H_{57}N_4O_{13}$ ([M+H]$^+$) 837.3917, found 837.3911.

[5] Prepared according to: Zhang, X.; Zhang, J.; Tong, L.; Luo, Y.; Su, M.; Zang, Y.; Li, J.; Lu, W.; Chen, Y. *Bioorg. Med. Chem.* **2013**, *21*, 3240-3244.

5. ESI-MS Spectra of Protein Bioconjugates

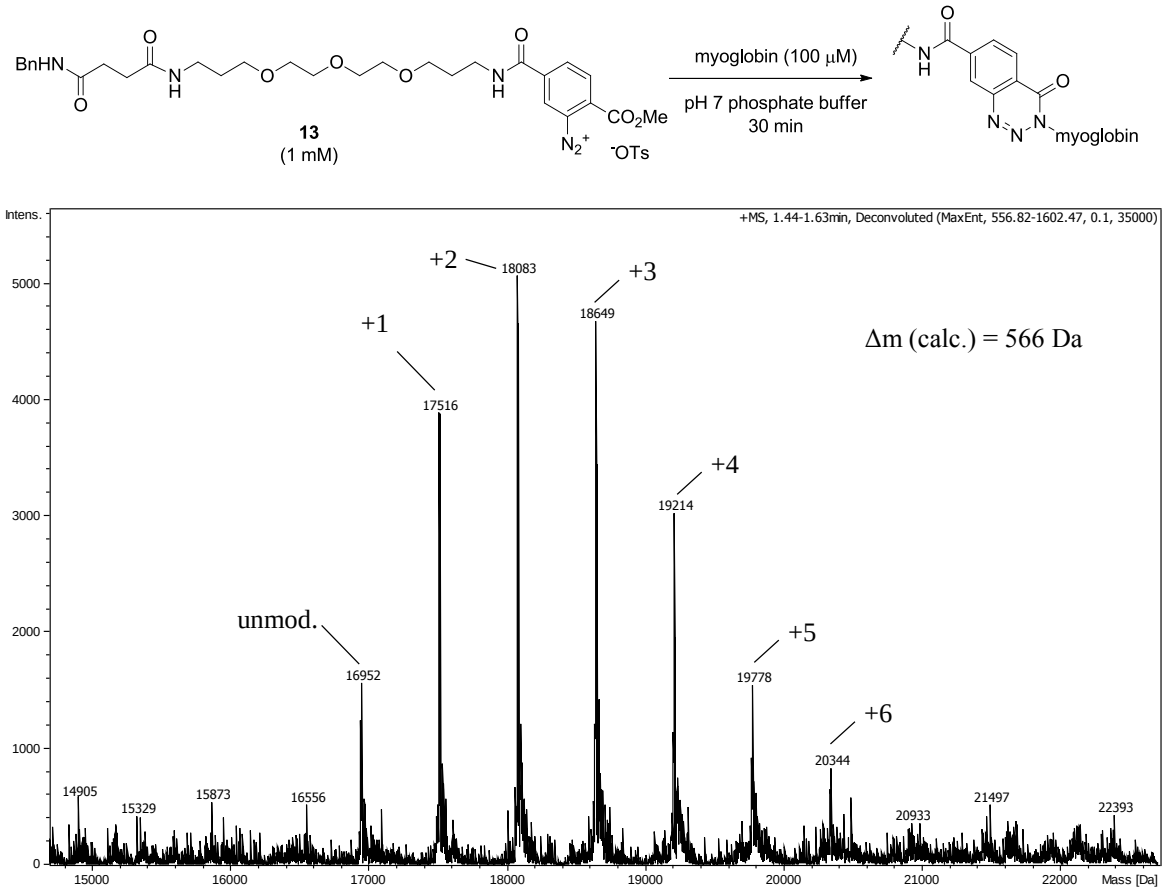


Figure S2. ESI-MS spectrum (deconvoluted) after 30 min for the bioconjugation of **13** (1 mM) to myoglobin (100 μM) (Figure 1 in main text).

Table S4. ESI-MS peak intensities (deconvoluted) after 30 min for the conjugation of **13** (1 mM) to myoglobin (100 μM) (Figure 1 in main text).

peak mass [Da]	modification	intensity	relative ratio [%]
16952	+0	1556	7
17516	+1	3872	19
18083	+2	5044	24
18649	+3	4656	22
19214	+4	3010	14
19778	+5	1539	7
20344	+6	827	4
20912	+7	308	1

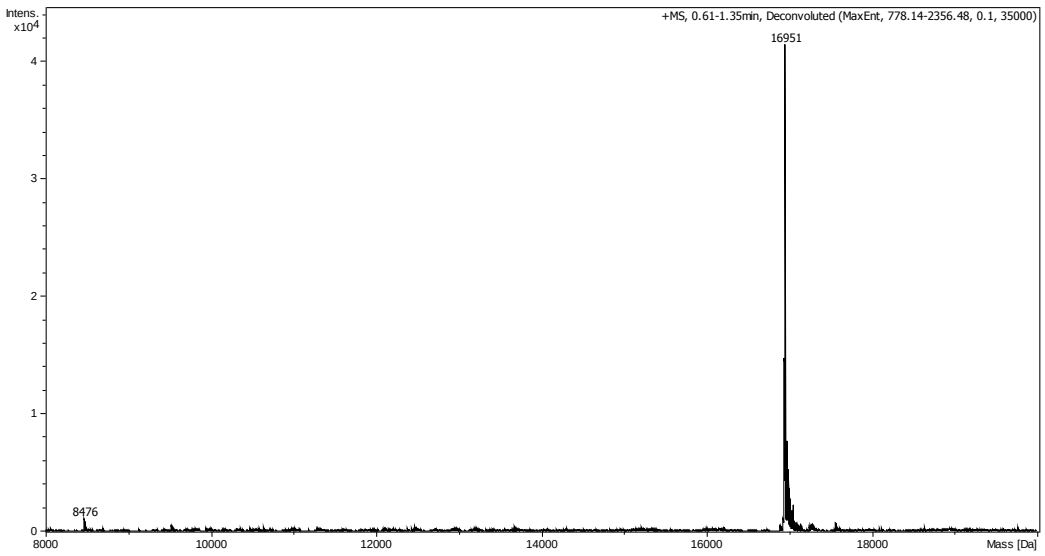


Figure S3. ESI-MS spectrum (deconvoluted) of unmodified myoblogin (equine skeletal muscle; Sigma-Aldrich M0630).

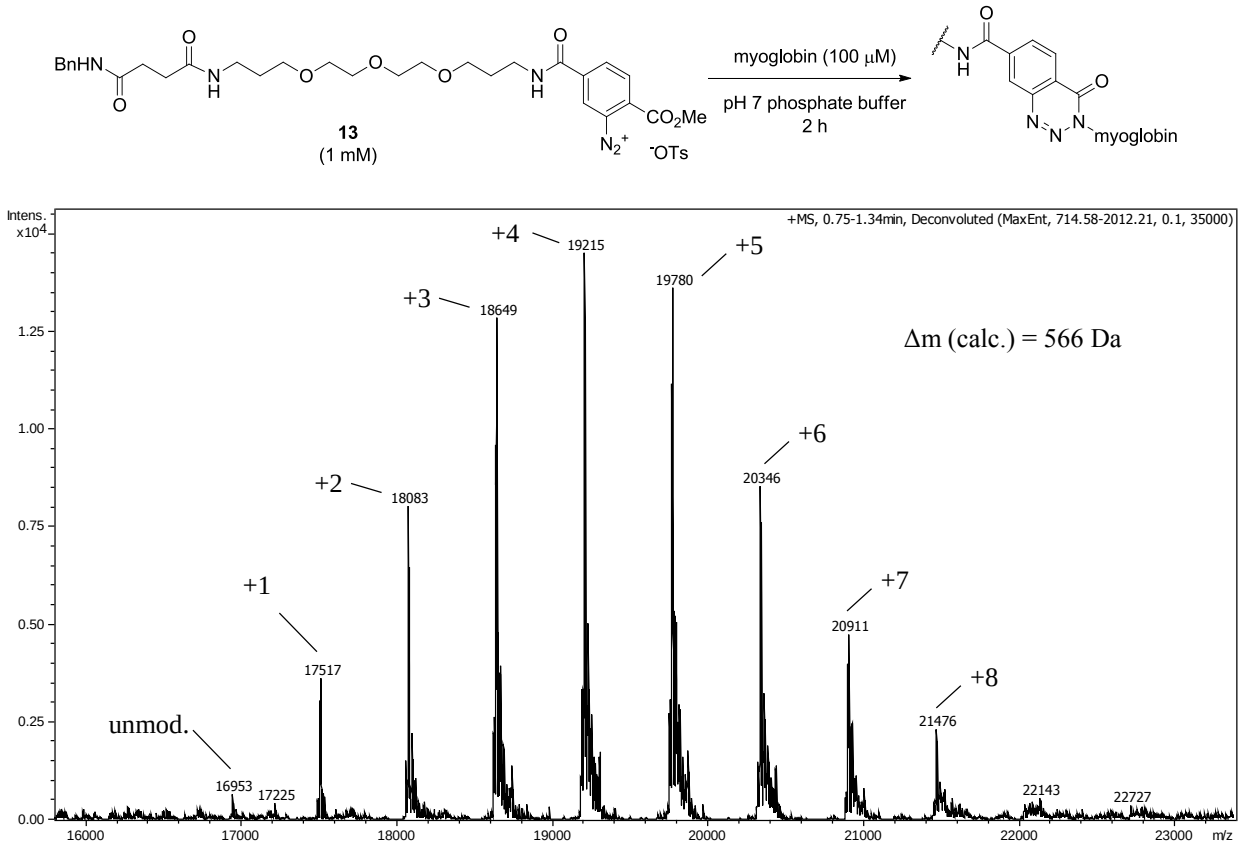


Figure S4. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **13** (1 mM) to myoglobin (100 μM).

Table S5. ESI-MS peak intensities (deconvoluted) after 2 hours for the conjugation of **13** (1 mM) to myoglobin (100 μM).

peak mass [Da]	modification	intensity	relative ratio [%]
16953	+0	634	1
17517	+1	3597	5
18083	+2	8002	12
18649	+3	12785	19
19215	+4	14470	22
19780	+5	13551	20
20346	+6	8498	13
20911	+7	4723	7
21476	+8	2301	3

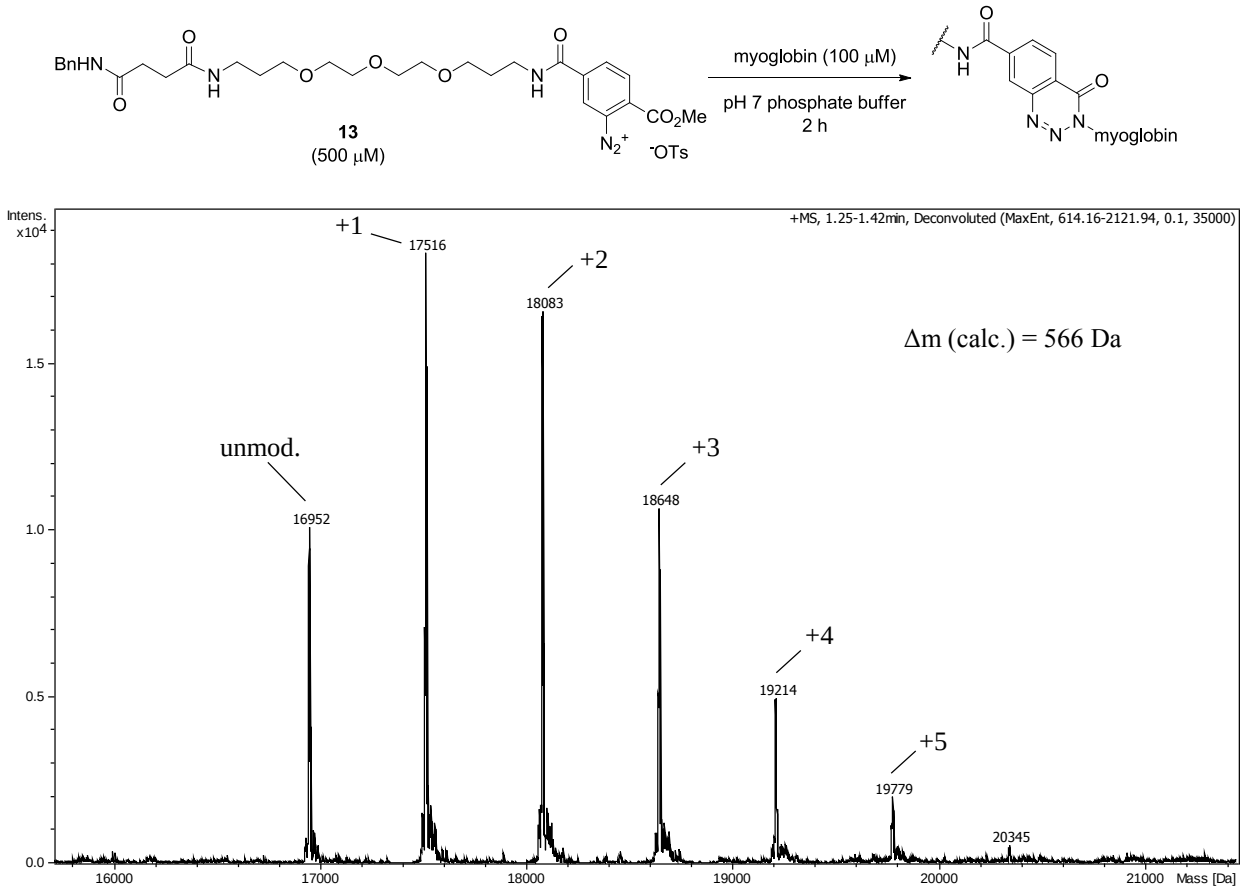


Figure S5. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **13** (500 μM) to myoglobin (100 μM) (main text, Table 1, entry 1).

Table S6. ESI-MS peak intensities (deconvoluted) after 2 hours for the conjugation of **13** (500 μM) to myoglobin (100 μM) (main text, Table 1, entry 1).

peak mass [Da]	modification	intensity	relative ratio [%]
16952	+0	10035	16
17516	+1	18243	29
18083	+2	16483	26
18648	+3	10588	17
19214	+4	4925	8
19779	+5	1984	3

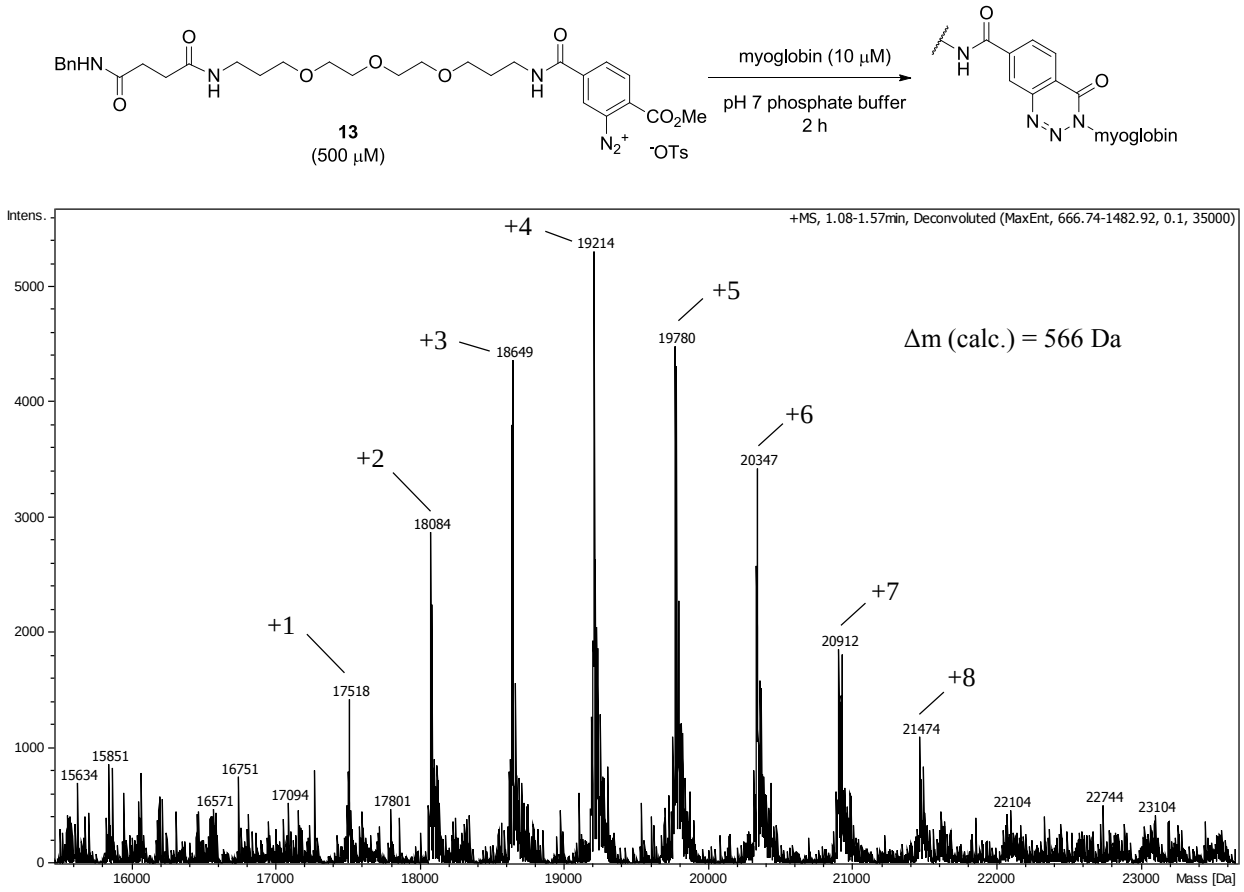


Figure S6. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **13** (500 μM) to myoglobin (10 μM) (main text, Table 1, entry 2).

Table S7. ESI-MS peak intensities (deconvoluted) after 2 hours for the conjugation of **13** (500 μM) to myoglobin (10 μM) (main text, Table 1, entry 2).

peak mass [Da]	modification	intensity	relative ratio [%]
16952	+0	137	0
17518	+1	1212	6
18084	+2	2861	12
18649	+3	4349	18
19214	+4	5279	21
19780	+5	4376	18
20347	+6	3411	14
20912	+7	1847	7
21474	+8	1094	4

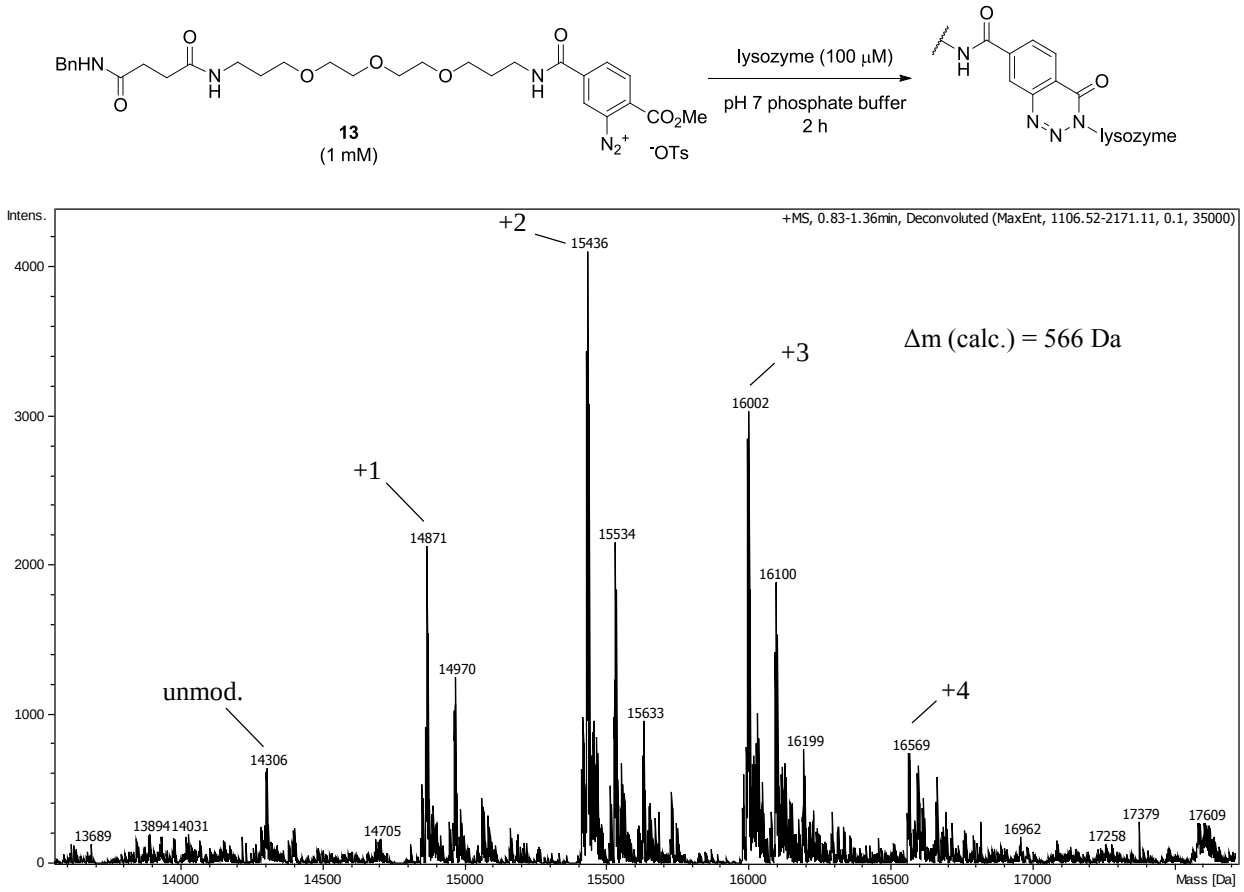


Figure S7. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **13** (1 mM) to lysozyme (100 μM) (main text, Table 1, entry 3).

Table S8. ESI-MS peak intensities (deconvoluted) after 2 hours for the conjugation of **13** (1 mM) to myoglobin (100 μM) (main text, Table 1, entry 3).

peak mass [Da]	modification	intensity	relative ratio [%]
14306	+0	634	6
14871	+1	2117	20
15436	+2	4087	39
16002	+3	3021	29
16569	+4	734	7

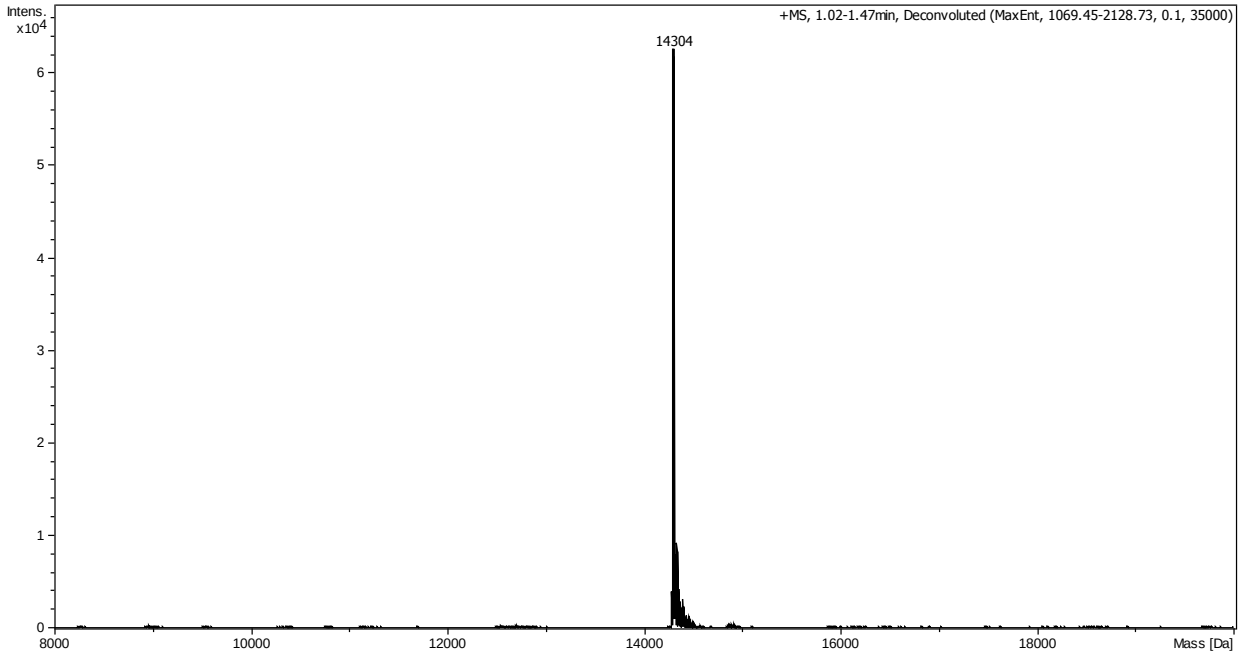


Figure S8. ESI-MS spectrum (deconvoluted) of unmodified lysozyme (egg white; Sigma-Aldrich 62970).

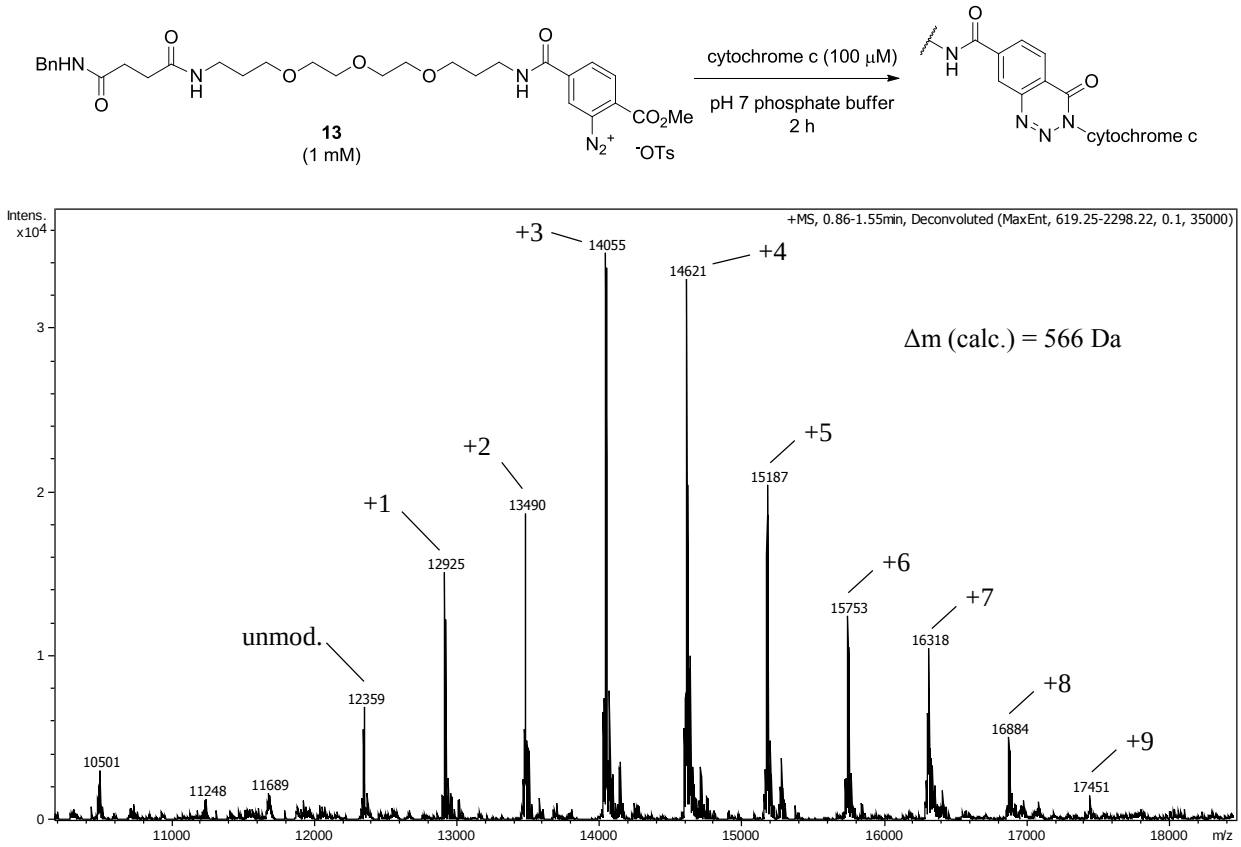


Figure S9. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **13** (1 mM) to cytochrome c (100 μM) (main text, Table 1, entry 4).

Table S9. ESI-MS peak intensities (deconvoluted) after 2 hours for the conjugation of **13** (1 mM) to cytochrome c (100 μM) (main text, Table 1, entry 4).

peak mass [Da]	modification	intensity	relative ratio [%]
12359	+0	6863	4
12925	+1	15055	10
13490	+2	18581	12
14055	+3	34469	22
14621	+4	32893	21
15187	+5	20377	13
15753	+6	12403	8
16318	+7	10374	7
16884	+8	4977	3
17451	+9	1469	1

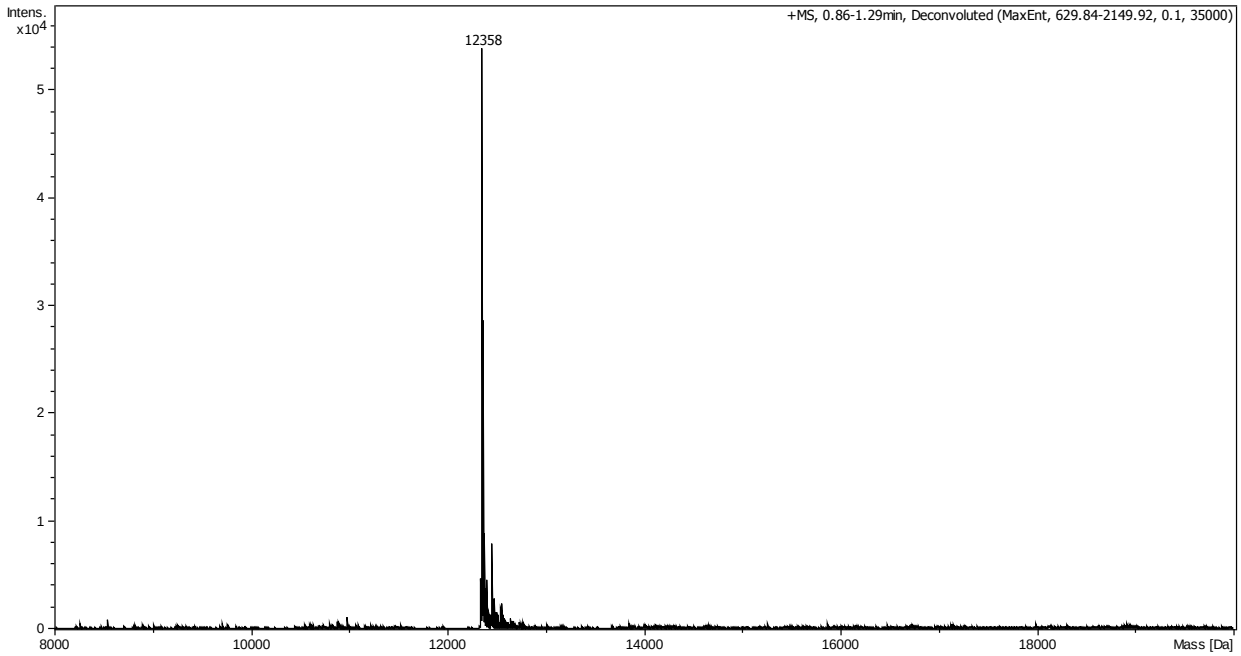


Figure S10. ESI-MS spectrum (deconvoluted) of unmodified cytochrome c (equine heart; Sigma-Aldrich C2506).

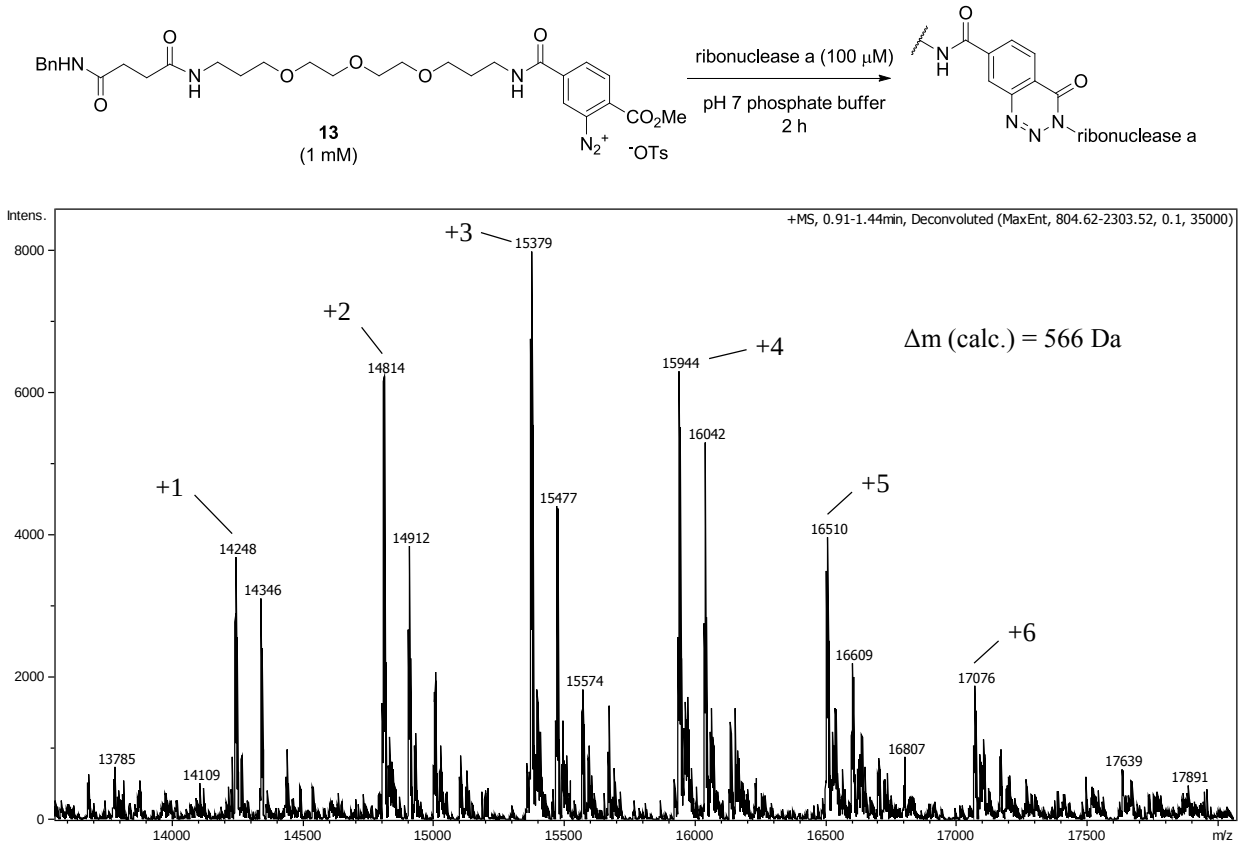


Figure S11. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **13** (1 mM) to ribonuclease a (100 μM) (main text, Table 1, entry 5).

Table S10. ESI-MS peak intensities (deconvoluted) after 2 hours for the conjugation of **13** (1 mM) to ribonuclease a (100 μM) (main text, Table 1, entry 5).

peak mass [Da]	modification	intensity	relative ratio [%]
13781	+0	-	0
14248	+1	3670	12
14814	+2	6206	21
15379	+3	7952	27
15944	+4	6277	21
16510	+5	3951	13
17076	+6	1866	6

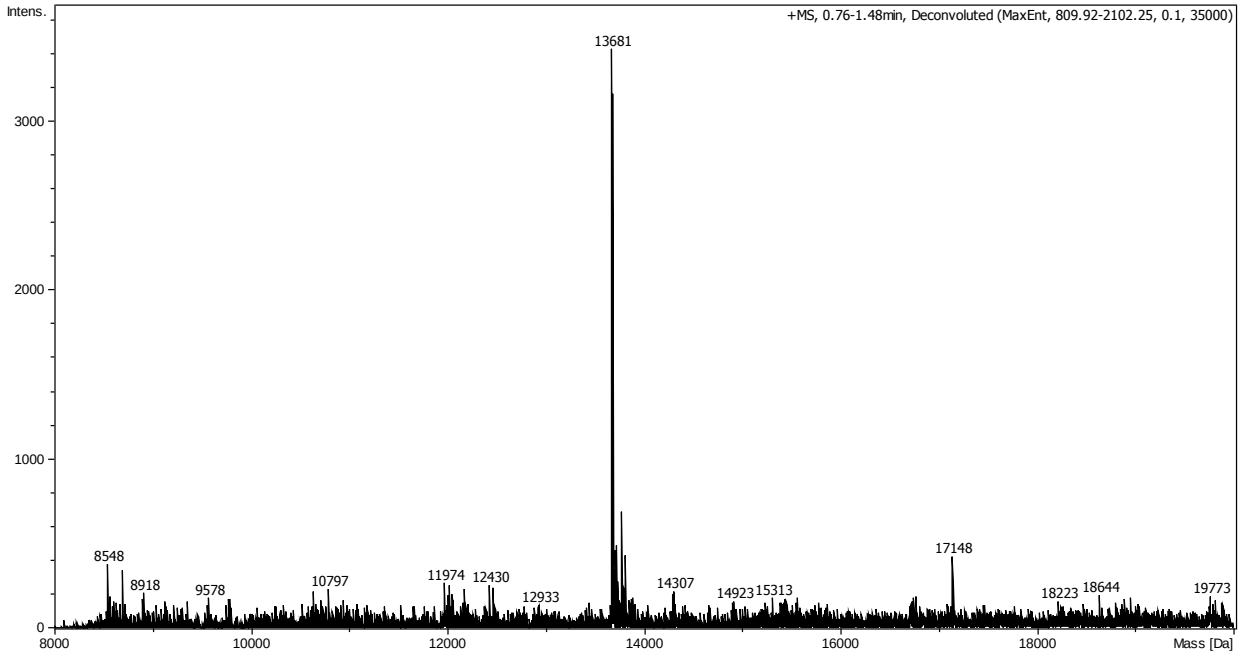


Figure S12. ESI-MS spectrum (deconvoluted) of unmodified ribonuclease a (bovine pancreas; Sigma-Aldrich 83831).

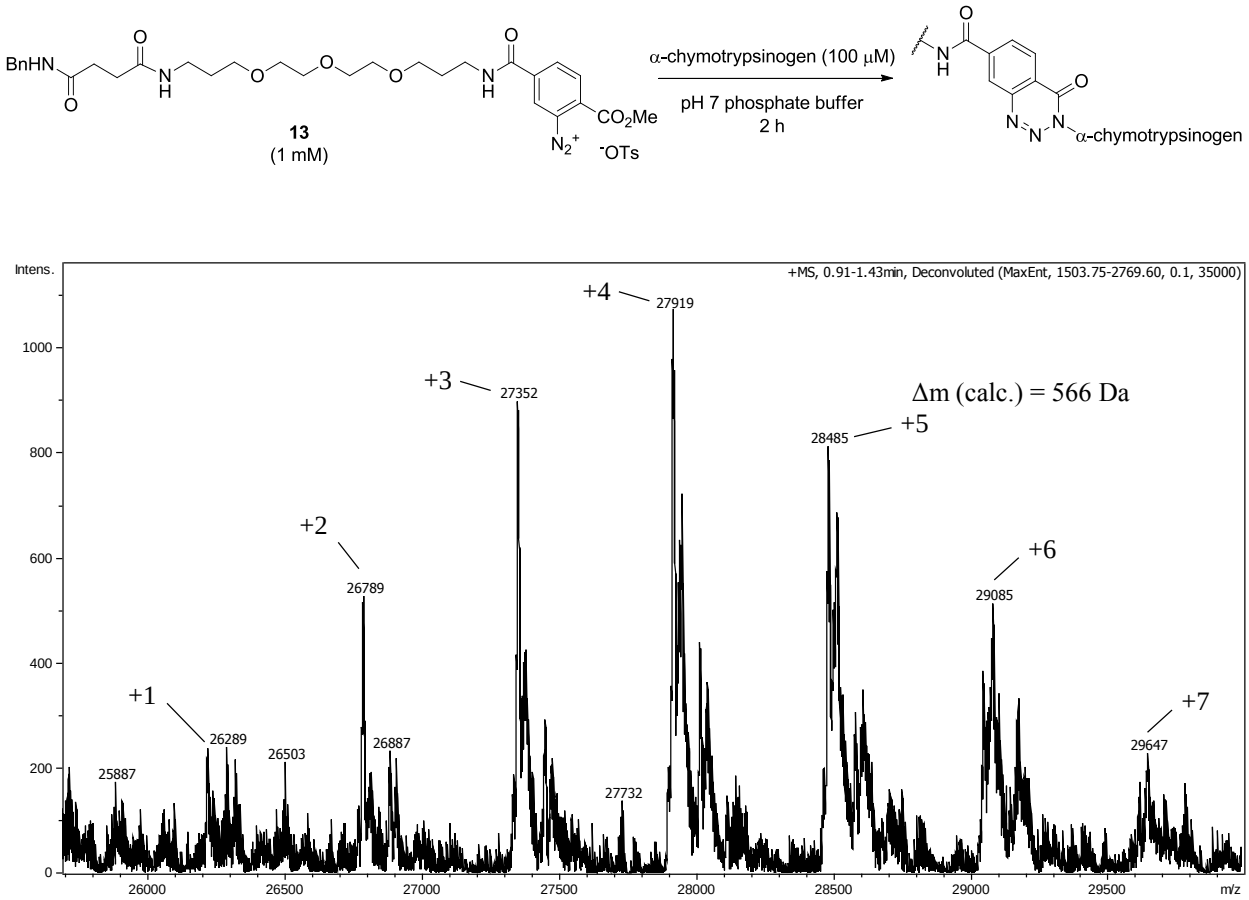


Figure S13. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **13** (1 mM) to α-chymotrypsinogen (100 μM) (main text, Table 1, entry 6).

Table S11. ESI-MS peak intensities (deconvoluted) after 2 hours for the conjugation of **13** (1 mM) to α-chymotrypsinogen (100 μM) (main text, Table 1, entry 6).

peak mass [Da]	modification	intensity	relative ratio [%]
25675	+0	-	0
26239	+1	156	4
26789	+2	525	13
27352	+3	894	21
27919	+4	1066	25
28485	+5	810	19
29085	+6	511	12
29647	+7	228	5

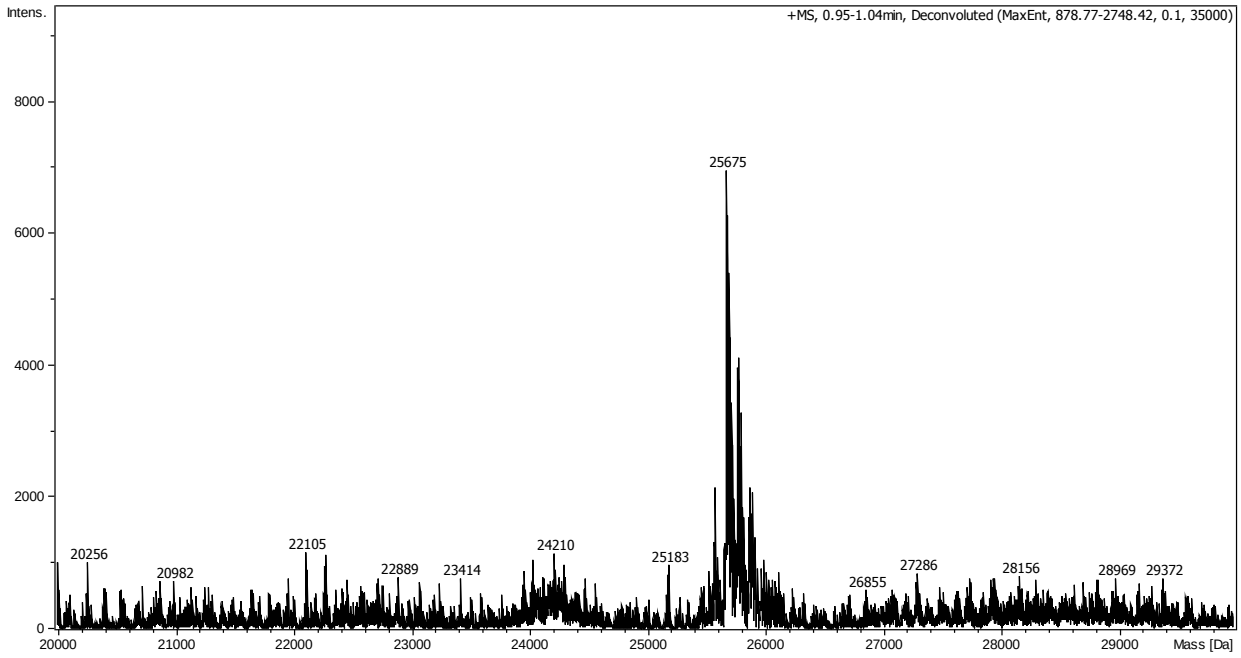


Figure S14. ESI-MS spectrum (deconvoluted) of unmodified α-chymotrypsinogen (Axon Lab A3975).

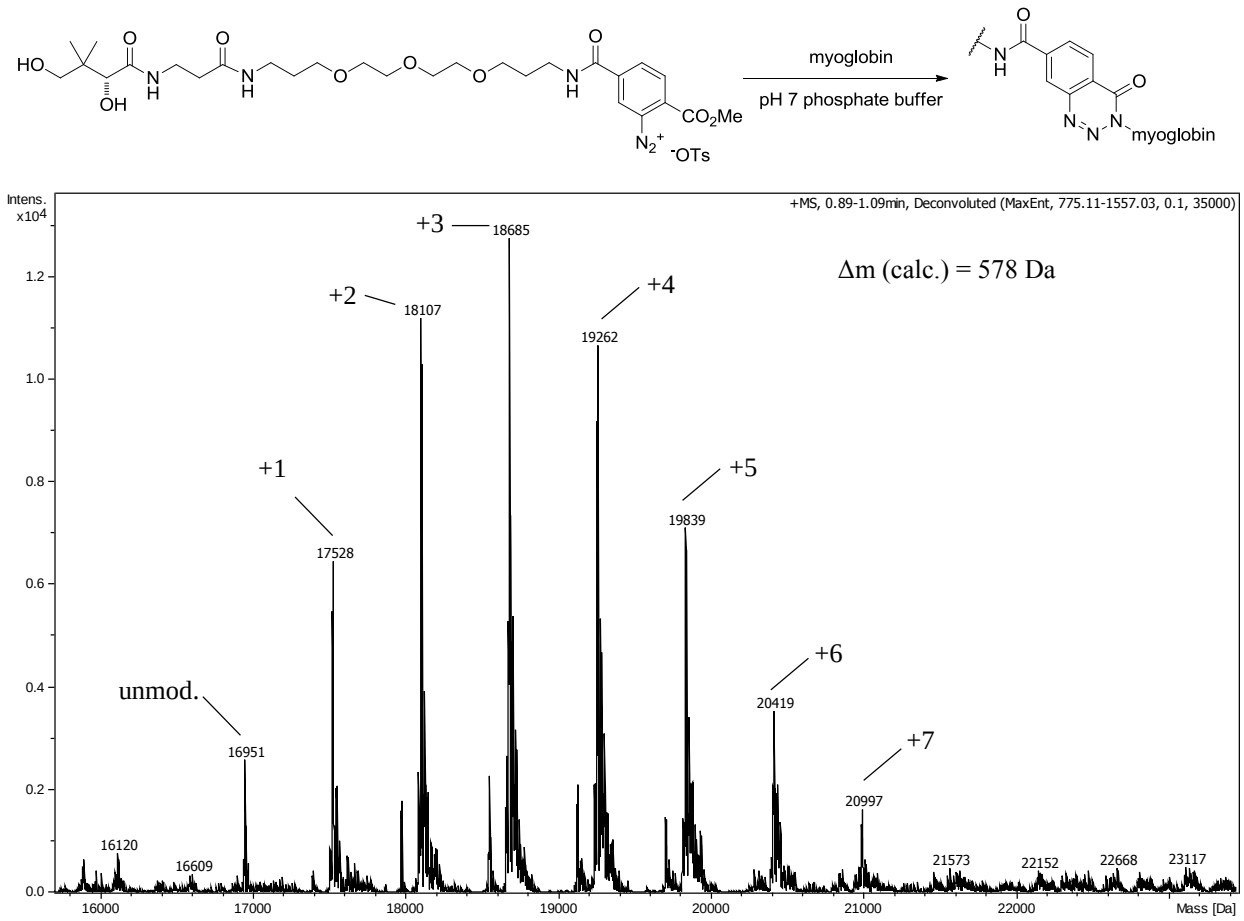


Figure S15. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **18** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 1).

Table S12. ESI-MS intensities (deconvoluted) after 2 h for the bioconjugation of **18** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 1).

peak mass [Da]	modification	intensity	relative ratio [%]
16951	+0	2566	5
17528	+1	6433	11
18107	+2	11148	20
18685	+3	12694	23
19262	+4	10624	19
19839	+5	7072	13
20419	+6	3513	6
20997	+7	1602	3

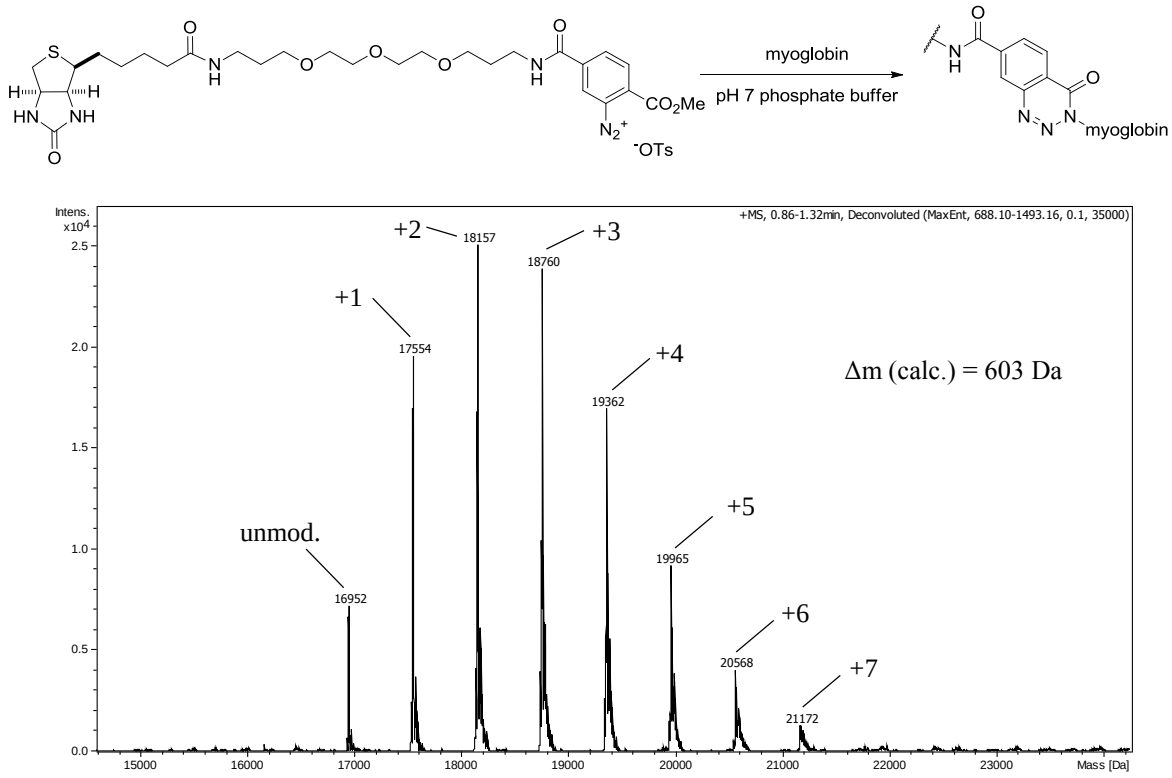
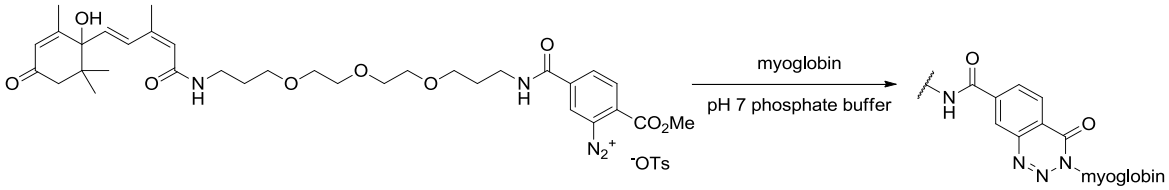


Figure S16. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **19** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 2).

Table S13. ESI-MS intensities (deconvoluted) after 2 h for the bioconjugation of **19** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 2).

peak mass [Da]	modification	intensity	relative ratio [%]
16952	+0	7159	7
17554	+1	19452	18
18157	+2	24962	23
18760	+3	23758	22
19362	+4	16917	16
19965	+5	9168	9
20568	+6	3983	4
21172	+7	1249	1



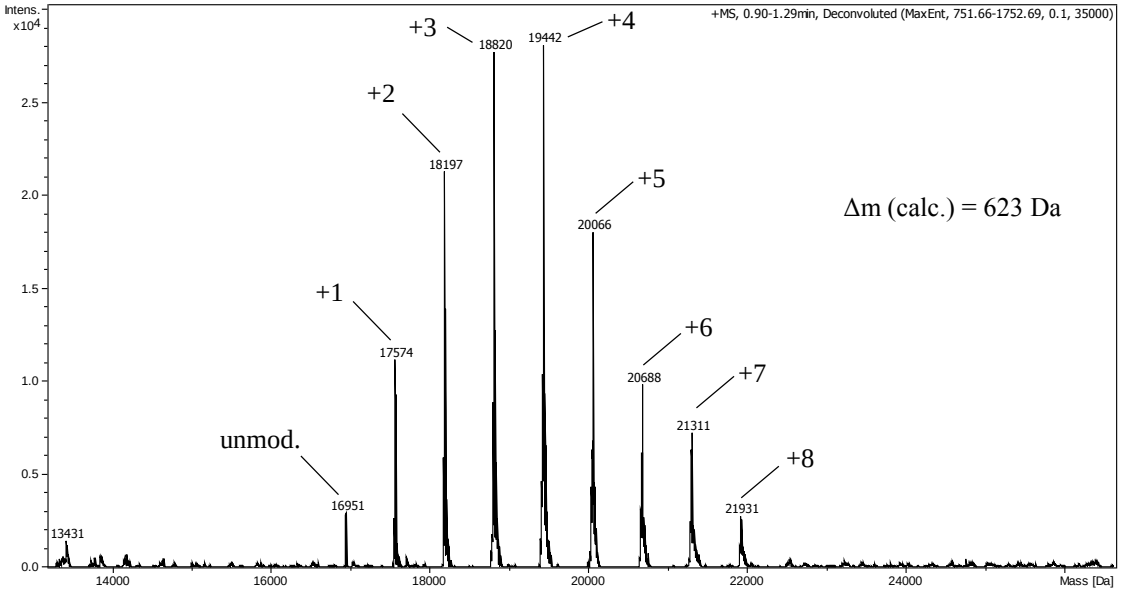


Figure S17. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **20** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 3).

Table S14. ESI-MS intensities (deconvoluted) after 2 h for the bioconjugation of **20** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 3).

peak mass [Da]	modification	intensity	relative ratio [%]
16951	+0	2898	2
17574	+1	11112	9
18197	+2	21196	16
18820	+3	27594	21
19442	+4	27993	22
20066	+5	17955	14
20688	+6	9769	8
21311	+7	7181	6
21931	+8	2729	2

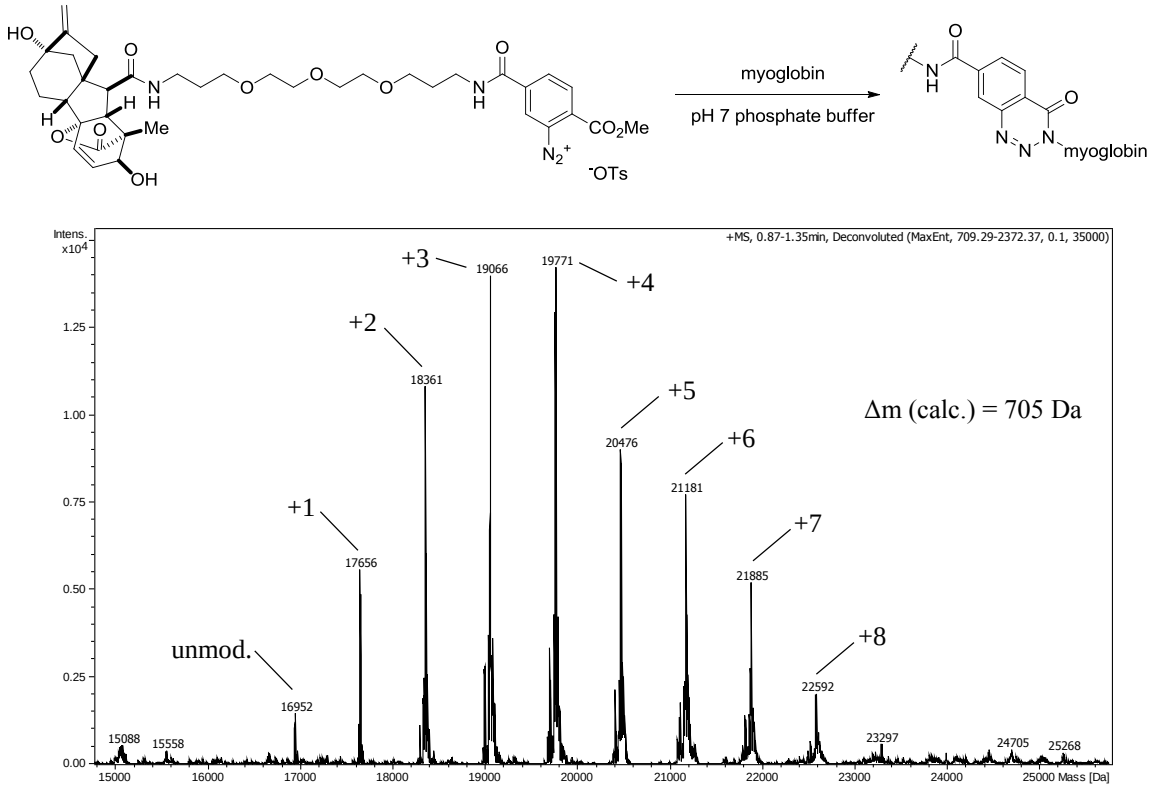


Figure S18. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **21** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 4).

Table S15. ESI-MS intensities (deconvoluted) after 2 h for the bioconjugation of **21** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 4).

peak mass [Da]	modification	intensity	relative ratio [%]
16952	+0	1448	2
17656	+1	5537	8
18361	+2	10765	16
19066	+3	13914	20
19771	+4	14165	20
20476	+5	8976	13
21181	+6	7689	11
21885	+7	5157	7
22592	+8	1984	3

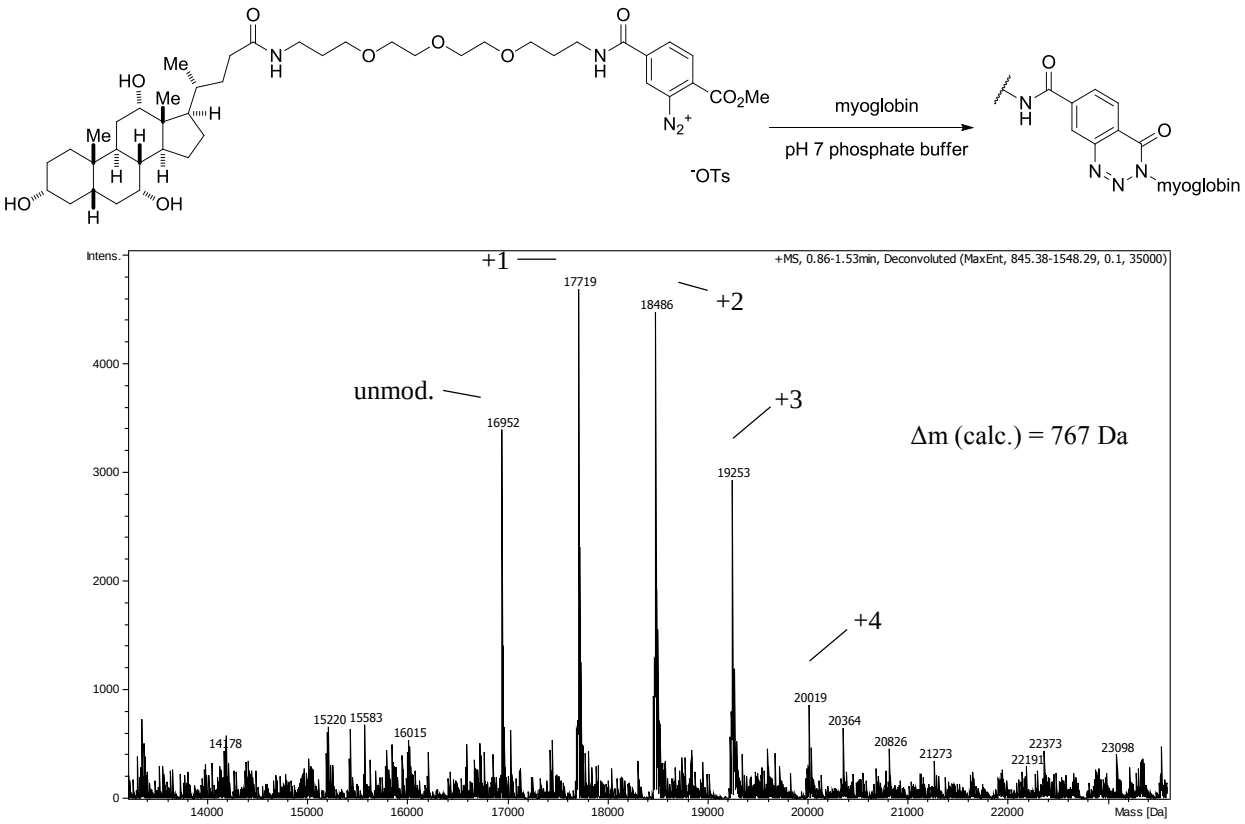


Figure S19. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **22** (2 mM) to myoglobin (100 μM) (main text, Table 2, entry 5).

Table S16. ESI-MS intensities (deconvoluted) after 2 h for the bioconjugation of **22** (2 mM) to myoglobin (100 μM) (main text, Table 2, entry 5).

peak mass [Da]	modification	intensity	relative ratio [%]
16952	+0	3377	20
17719	+1	4668	28
18486	+2	4459	27
19253	+3	2918	19
20019	+4	858	5
20785	+5	245	1

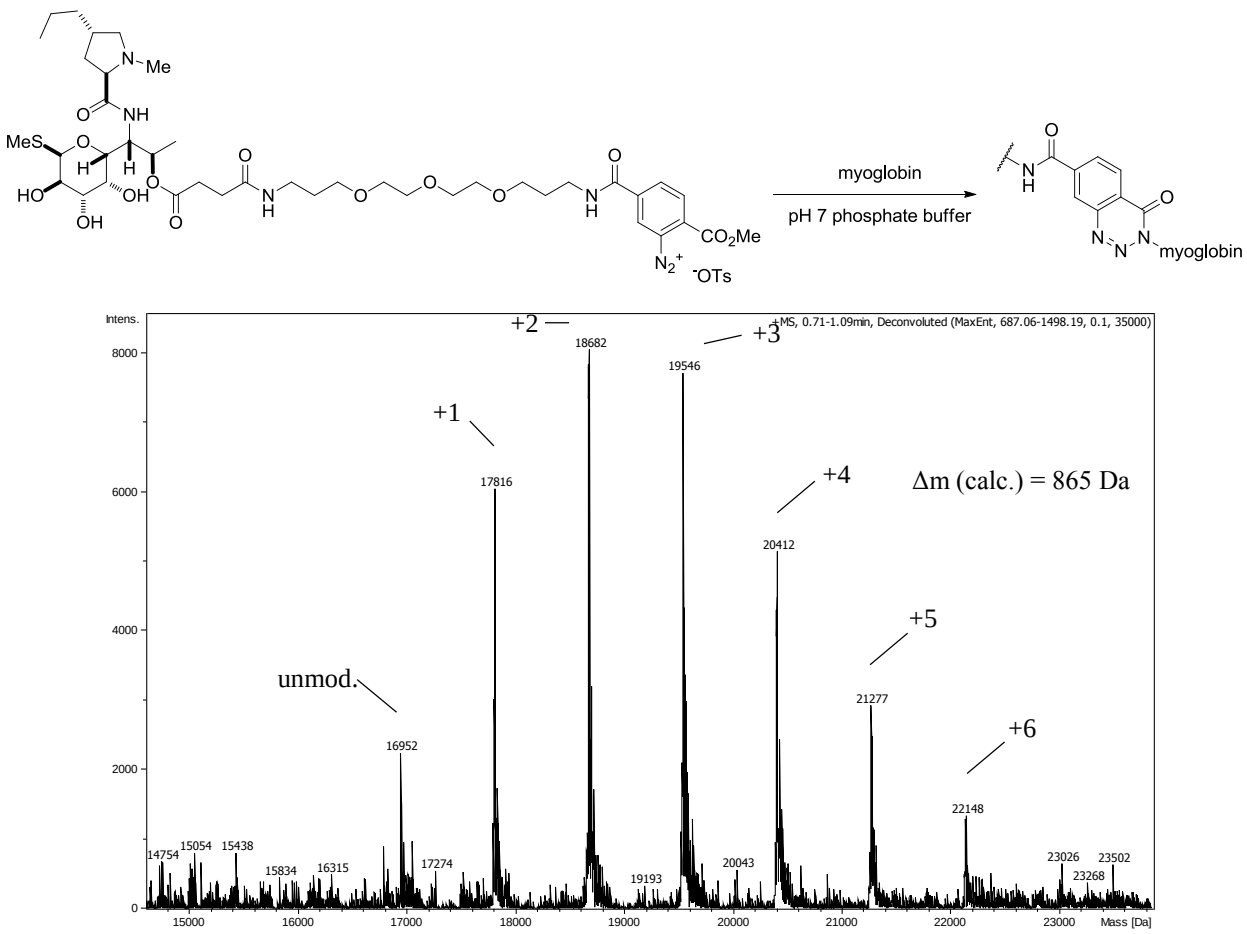


Figure S20. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **23** (2 mM) to myoglobin (100 μM) (main text, Table 2, entry 6).

Table S17. ESI-MS intensities (deconvoluted) after 2 h for the bioconjugation of **23** (2 mM) to myoglobin (100 μM) (main text, Table 2, entry 6).

peak mass [Da]	modification	intensity	relative ratio [%]
16952	+0	2219	7
17816	+1	6023	18
18682	+2	8015	24
19546	+3	7675	23
20412	+4	5114	15
21277	+5	2915	9
22148	+6	1324	4

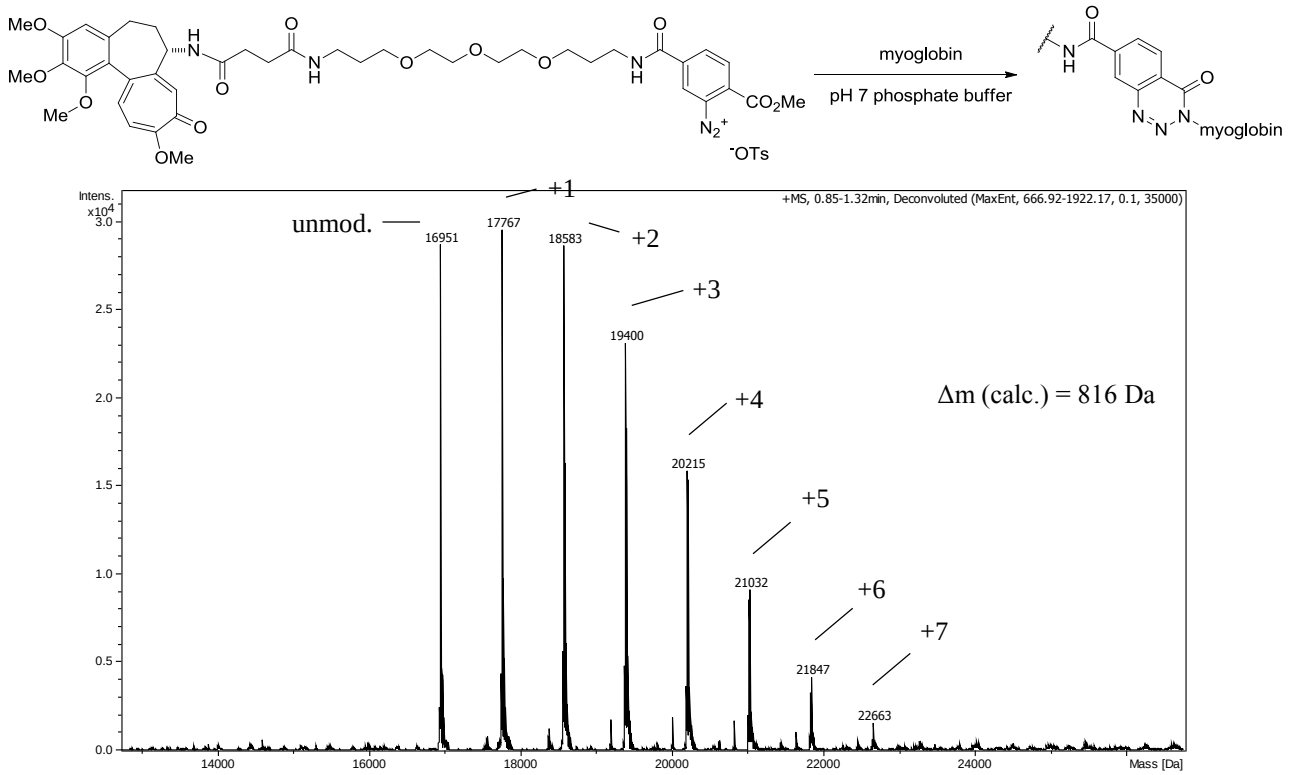
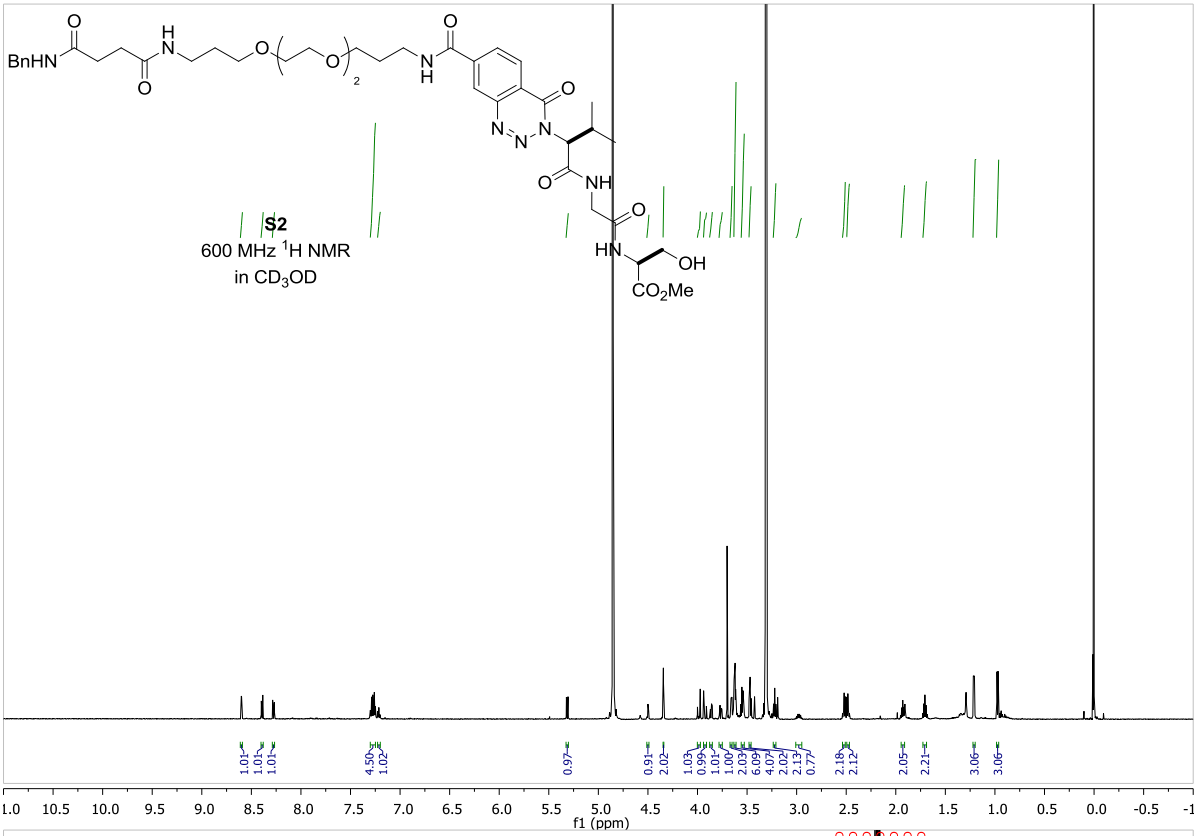


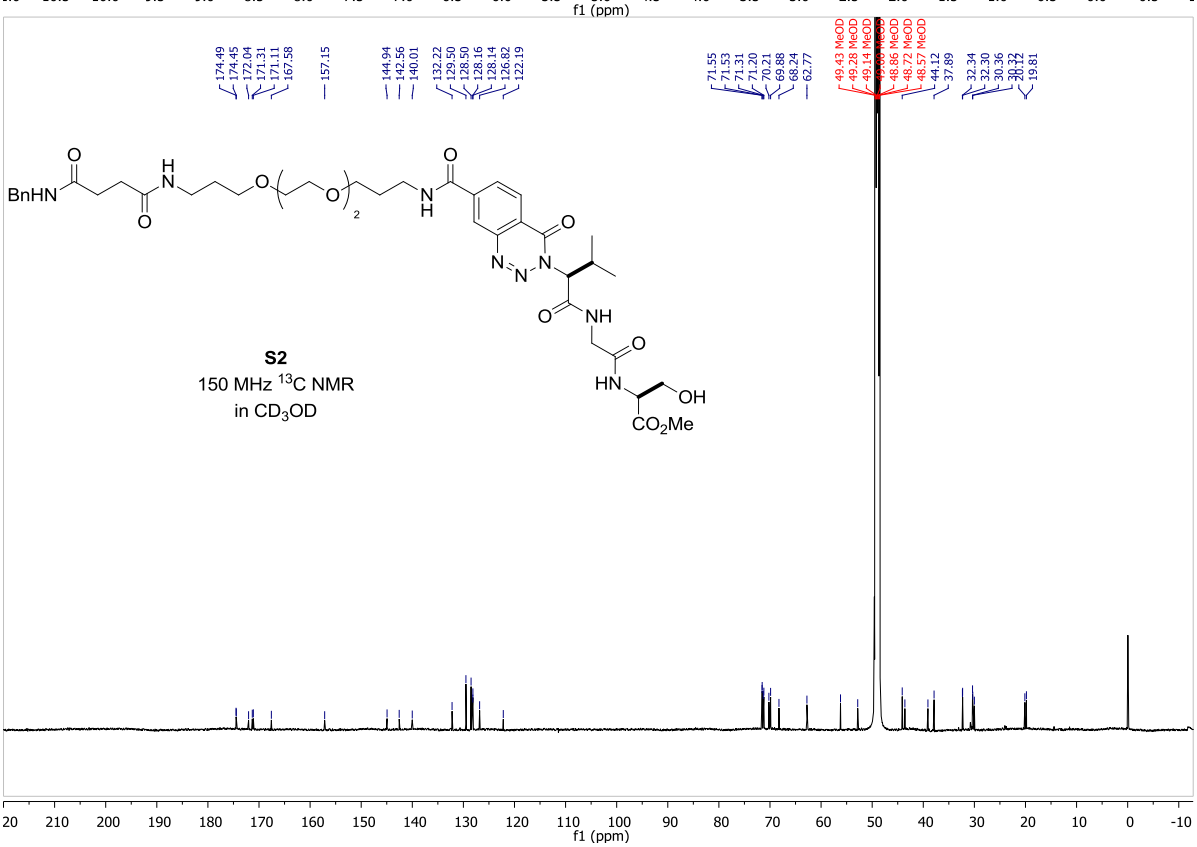
Figure S21. ESI-MS spectrum (deconvoluted) after 2 h for the bioconjugation of **24** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 7).

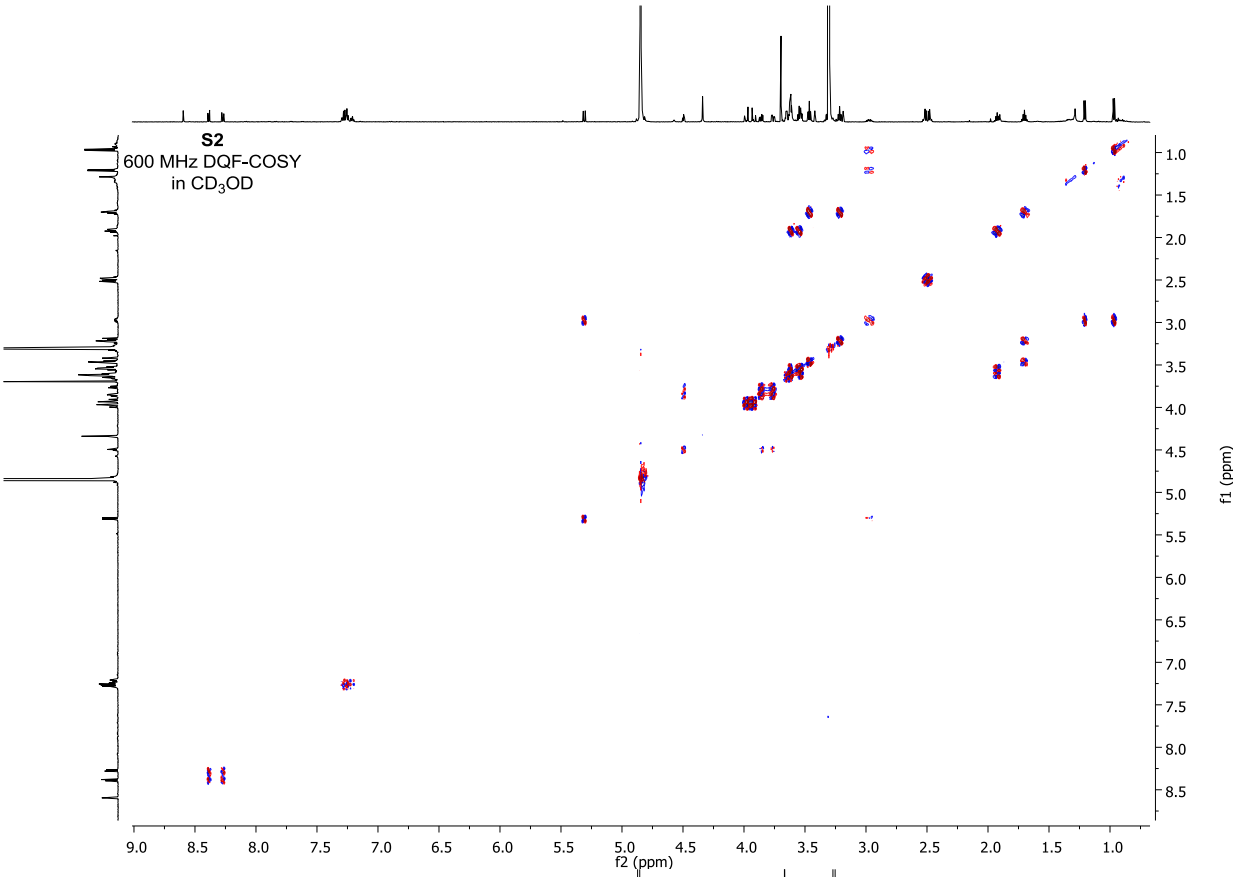
Table S18. ESI-MS intensities (deconvoluted) after 2 h for the bioconjugation of **24** (1 mM) to myoglobin (100 μM) (main text, Table 2, entry 7).

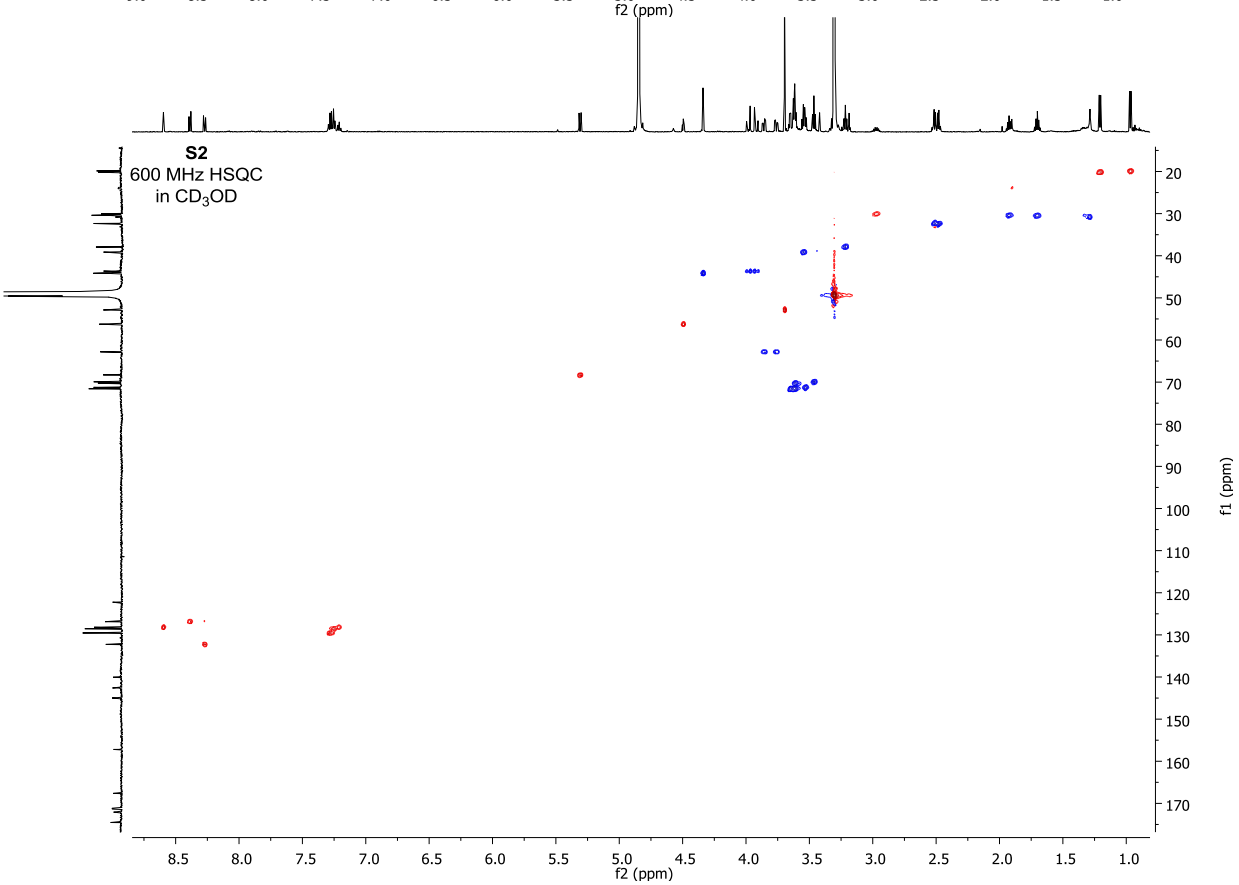
peak mass [Da]	modification	intensity	relative ratio [%]
16951	+0	28622	21
17767	+1	29403	21
18583	+2	28530	21
19400	+3	23022	15
20215	+4	15807	12
21032	+5	9081	7
21847	+6	4140	3
22663	+7	1541	1

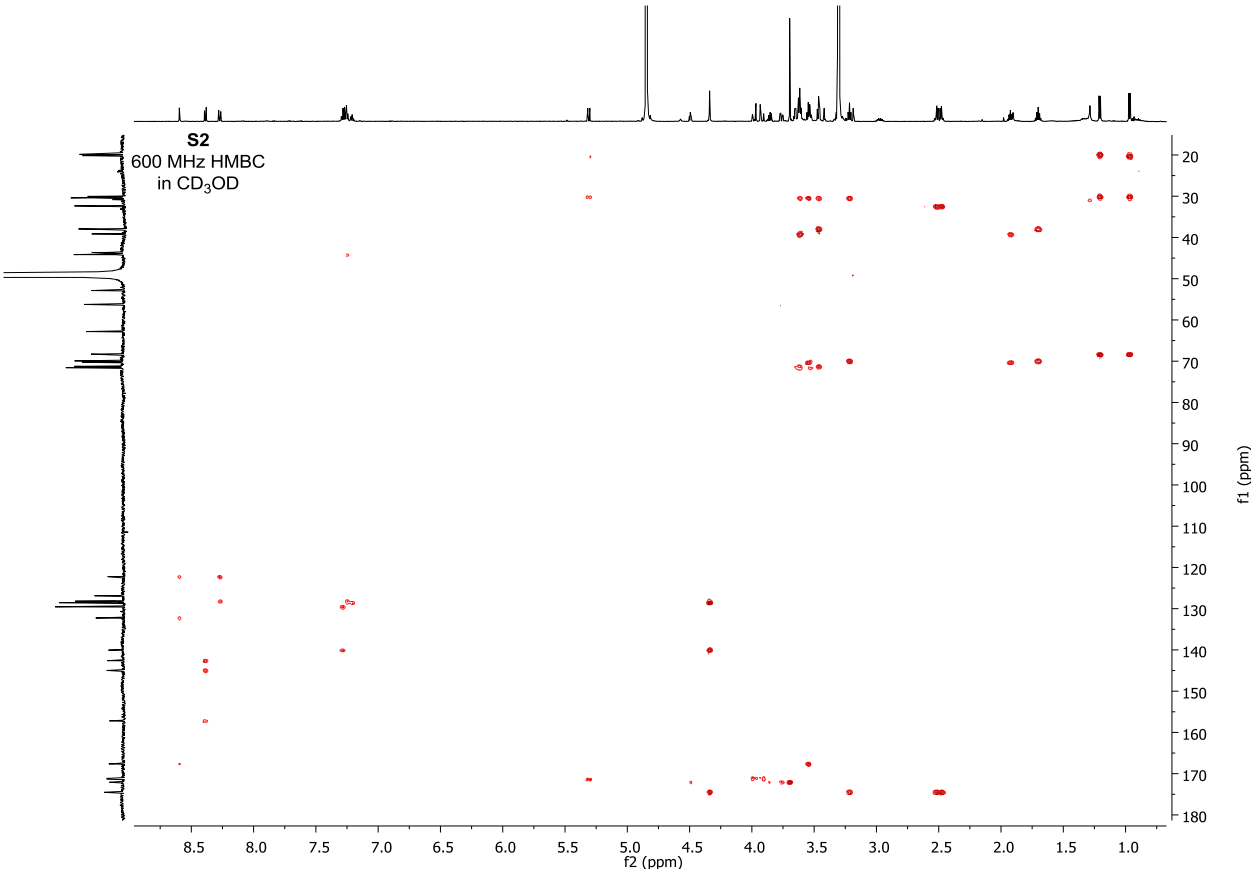
6. NMR Spectra

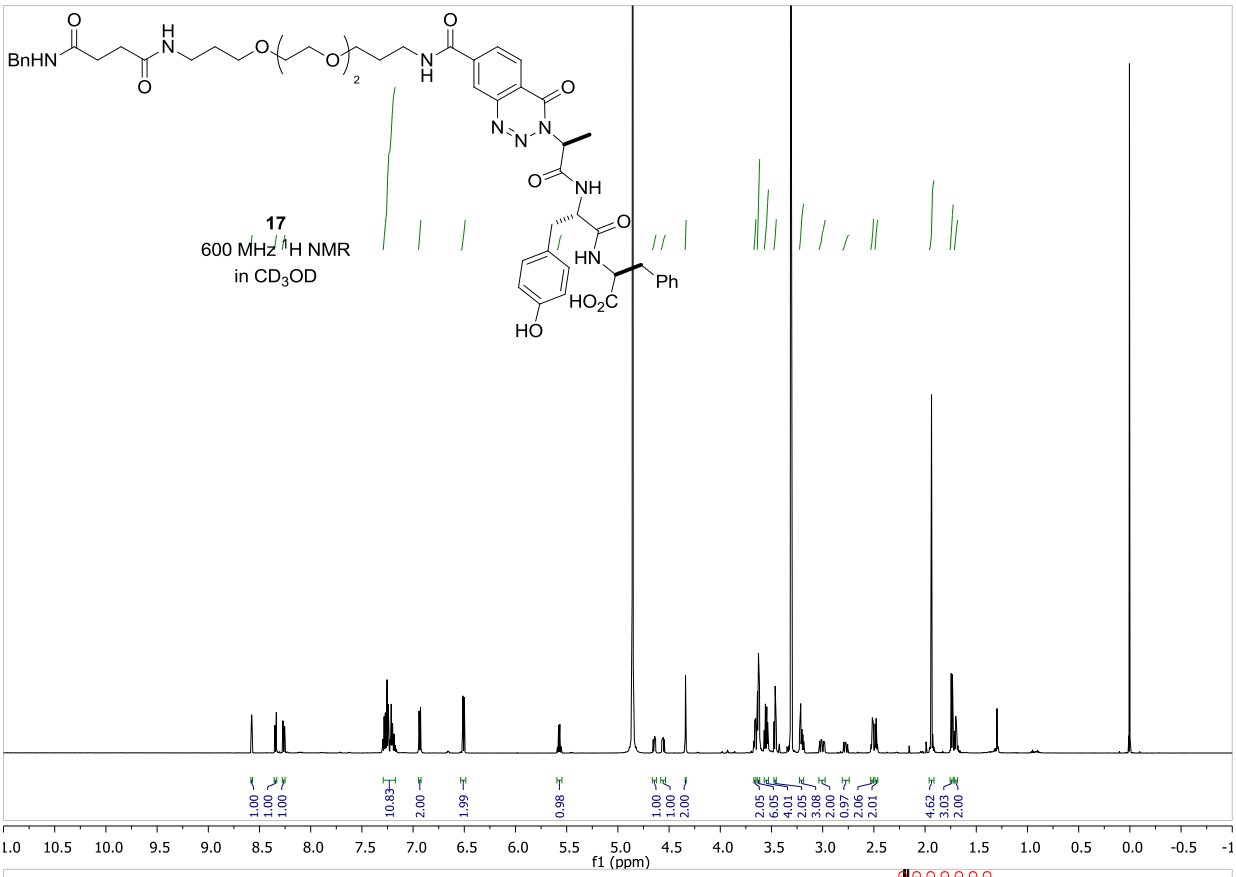


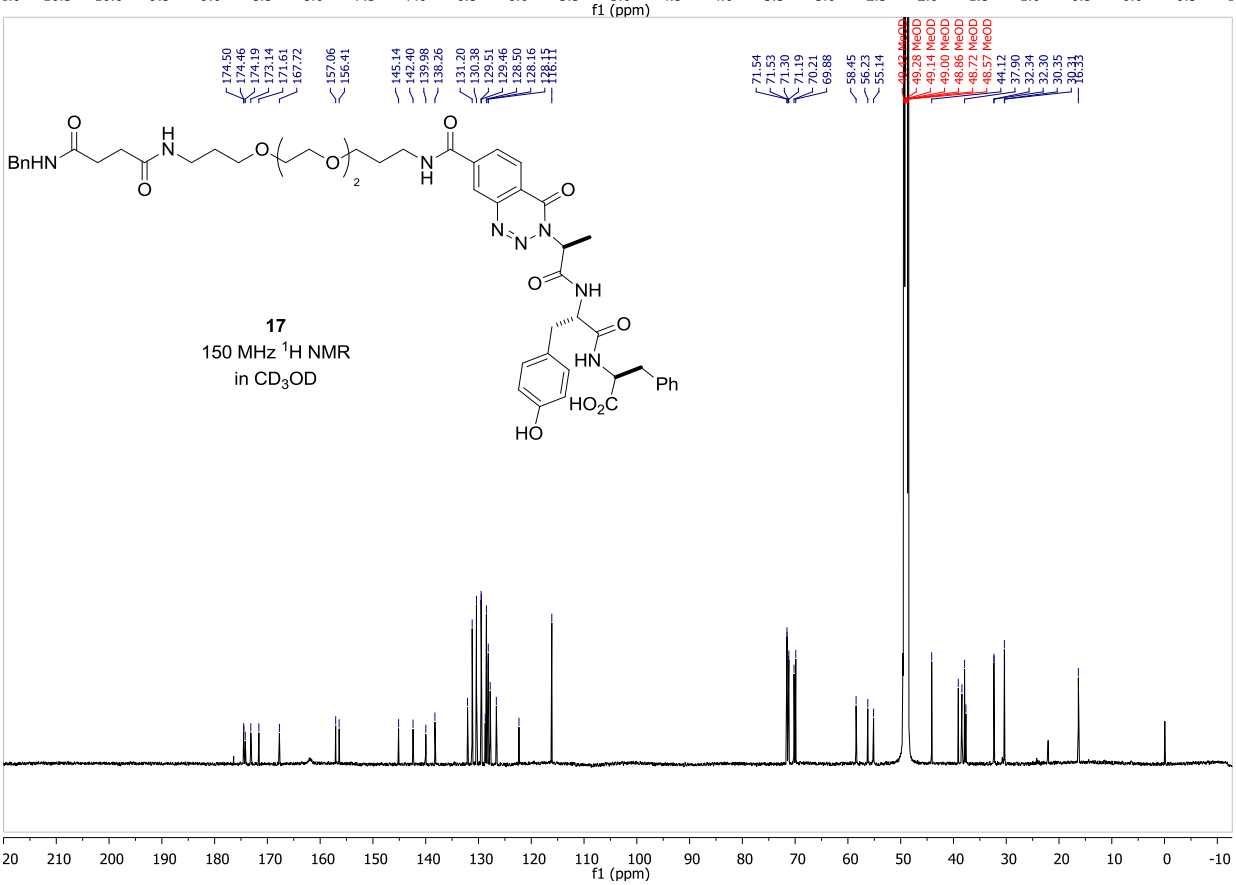


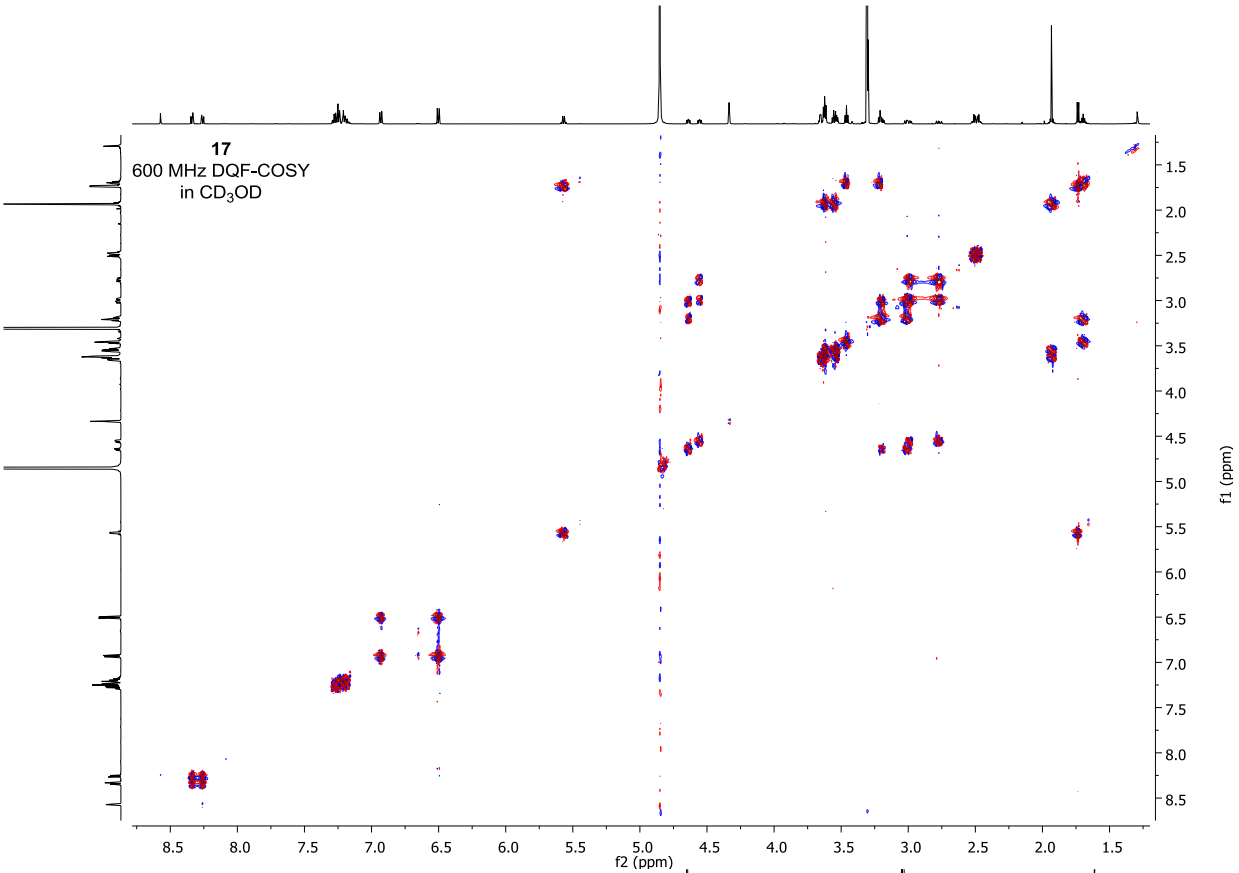

S2
600 MHz DQF-COSY
in CD3OD
f2 (ppm)
f1 (ppm)

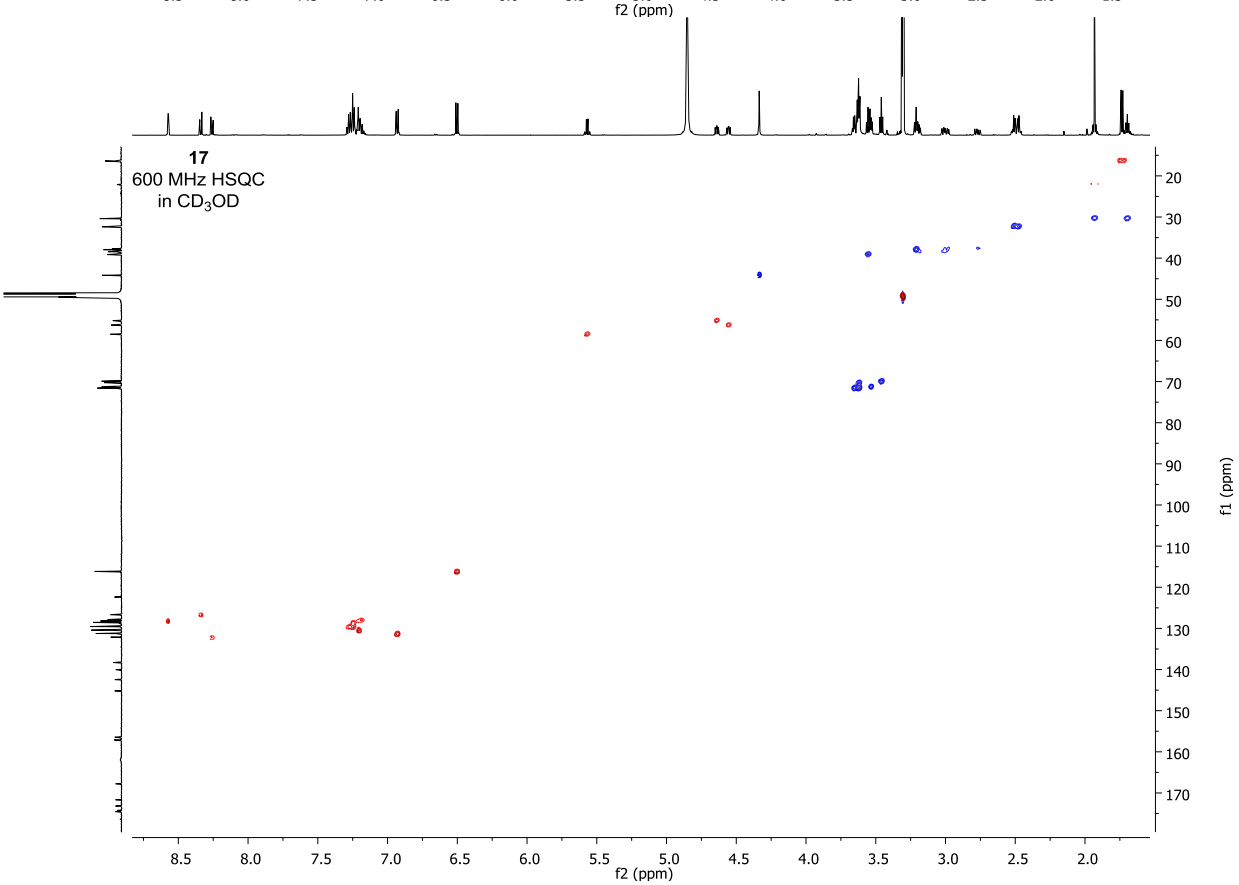

S2
600 MHz HSQC
in CD3OD
f2 (ppm)
f1 (ppm)

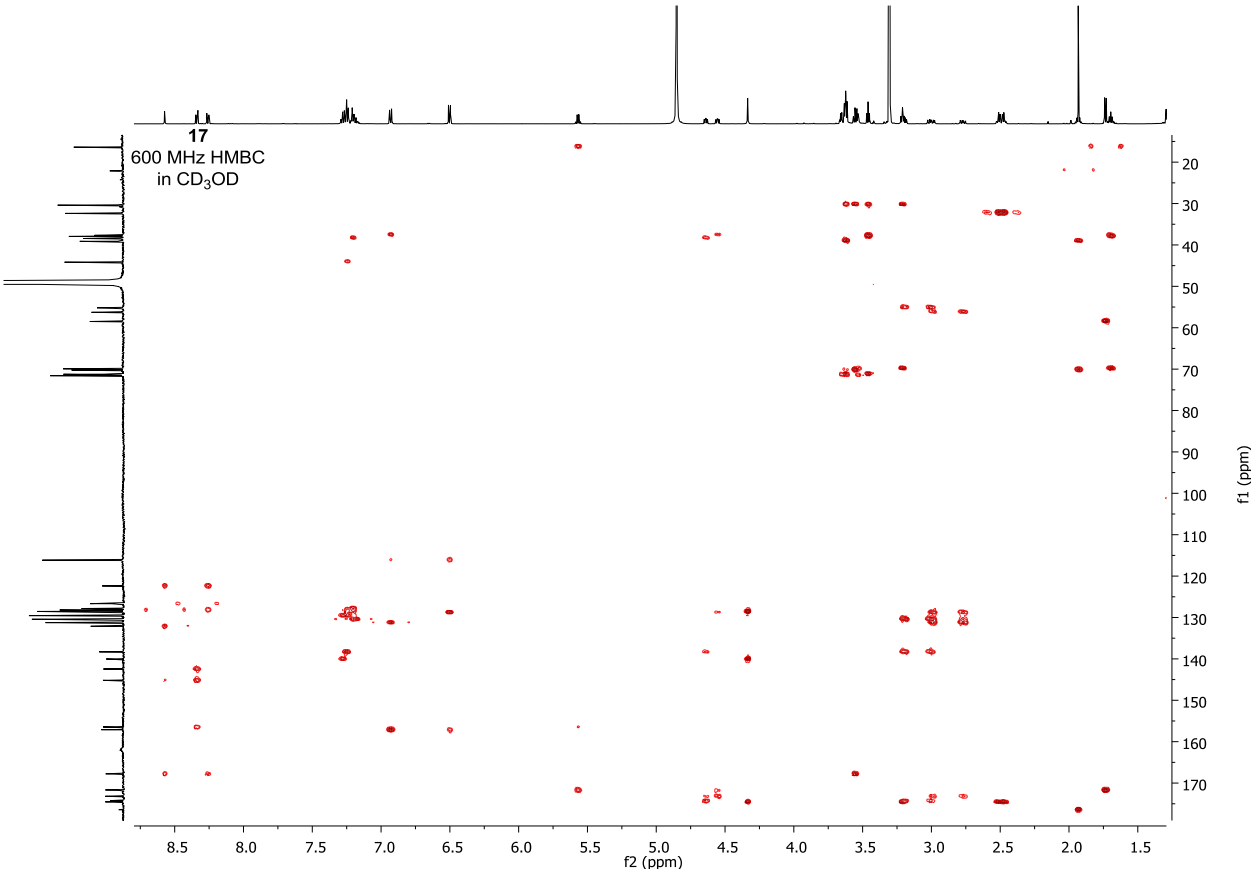

S2
600 MHz HMBC
in CD3OD
f2 (ppm)
f1 (ppm)

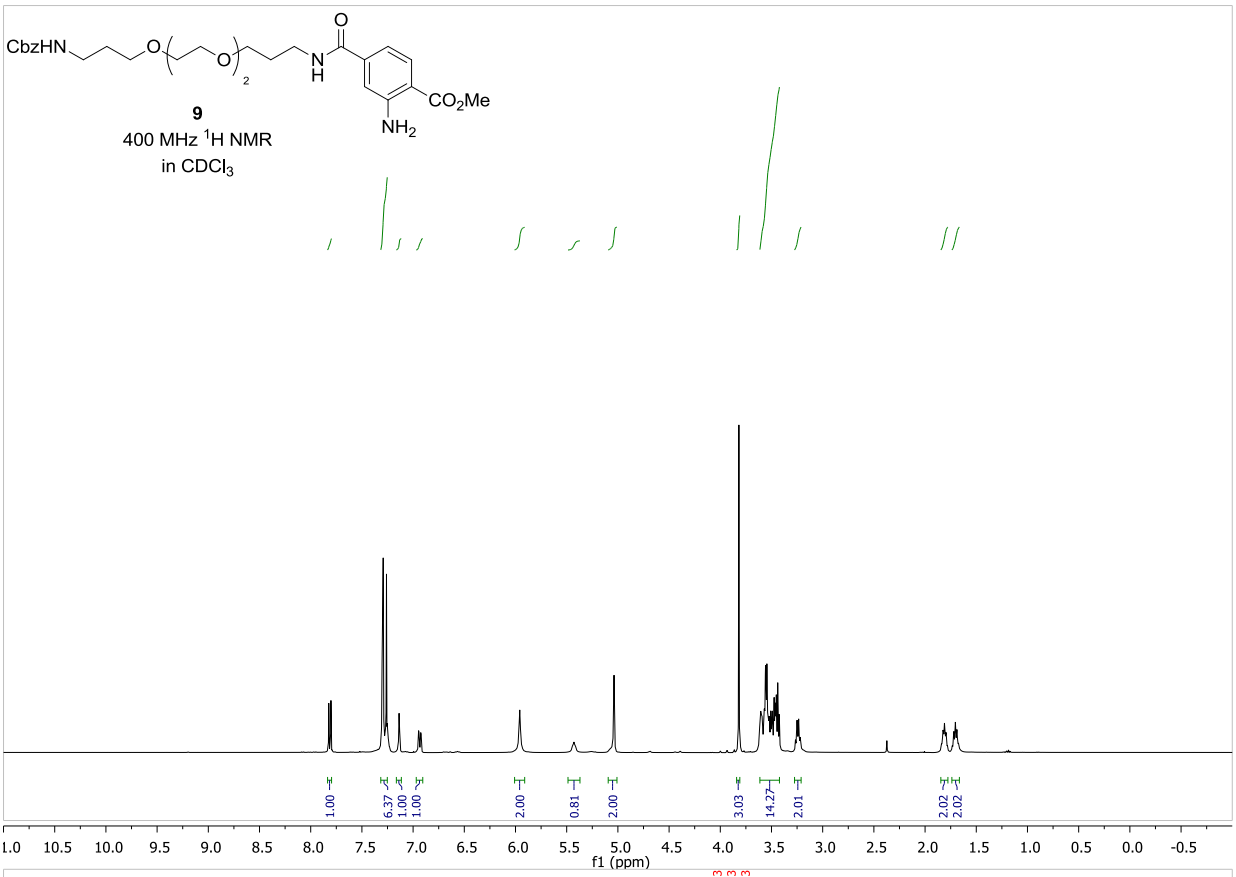

17
600 MHz ^{1}H NMR
in CD_3OD
f1 (ppm)

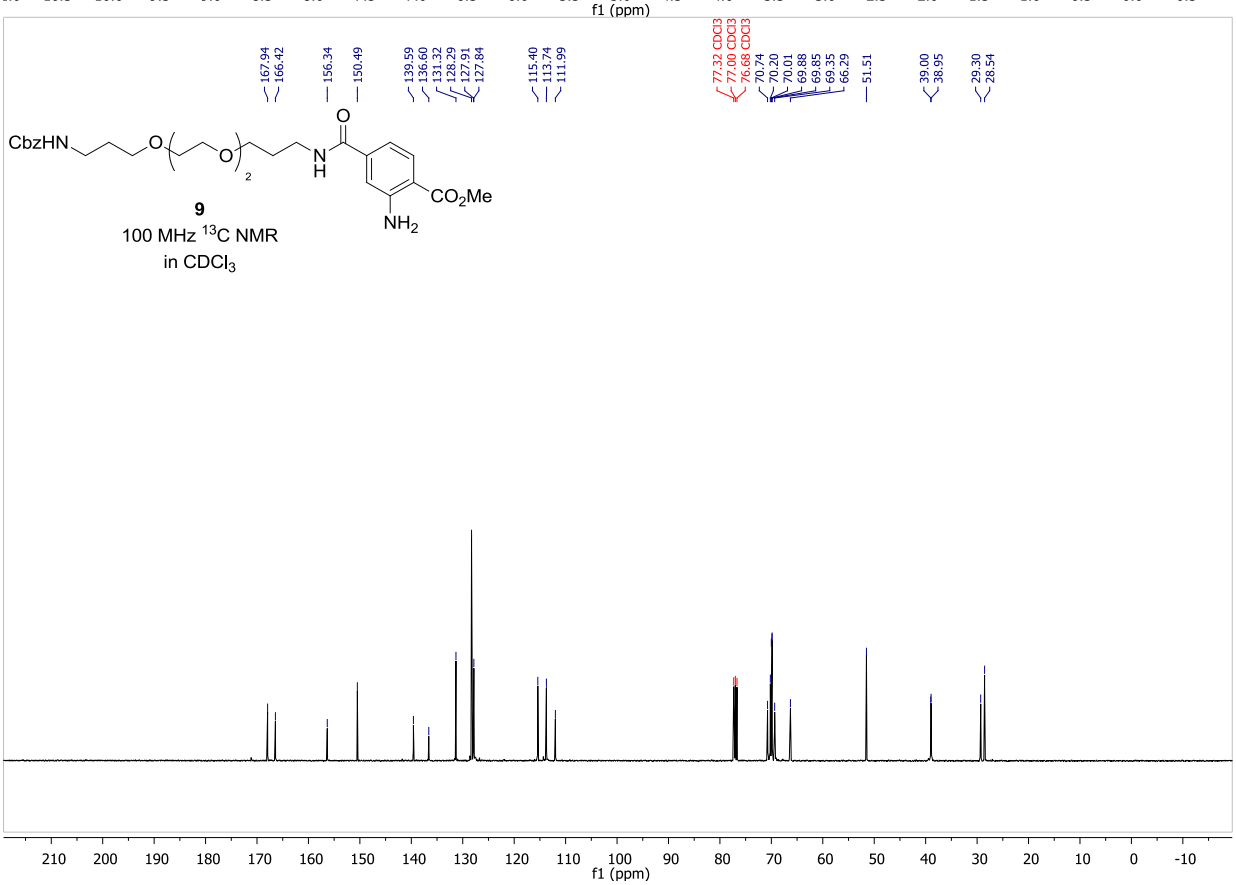

17
150 MHz ^{1}H NMR
in CD_3OD
f1 (ppm)

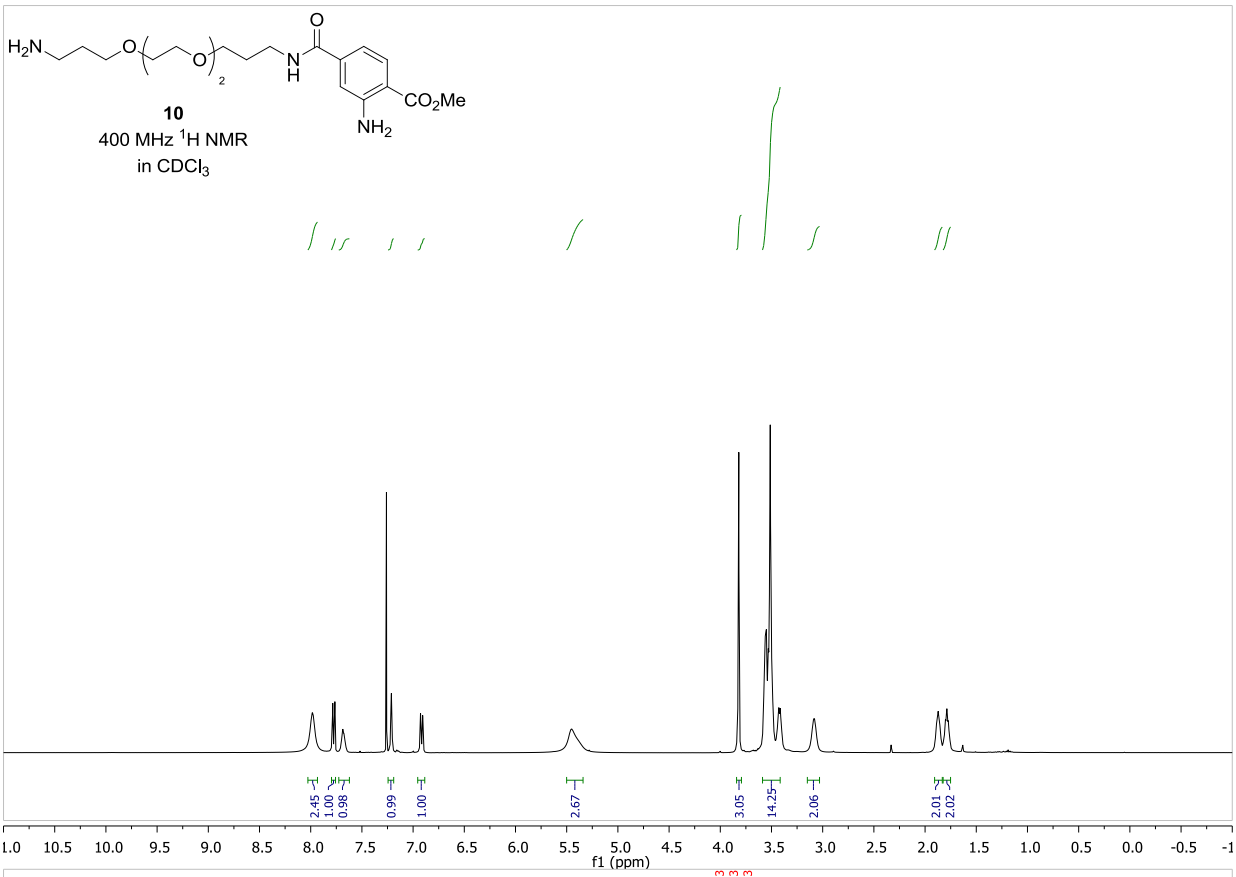

17
600 MHz DQF-COSY
in CD3OD
f2 (ppm)
f1 (ppm)

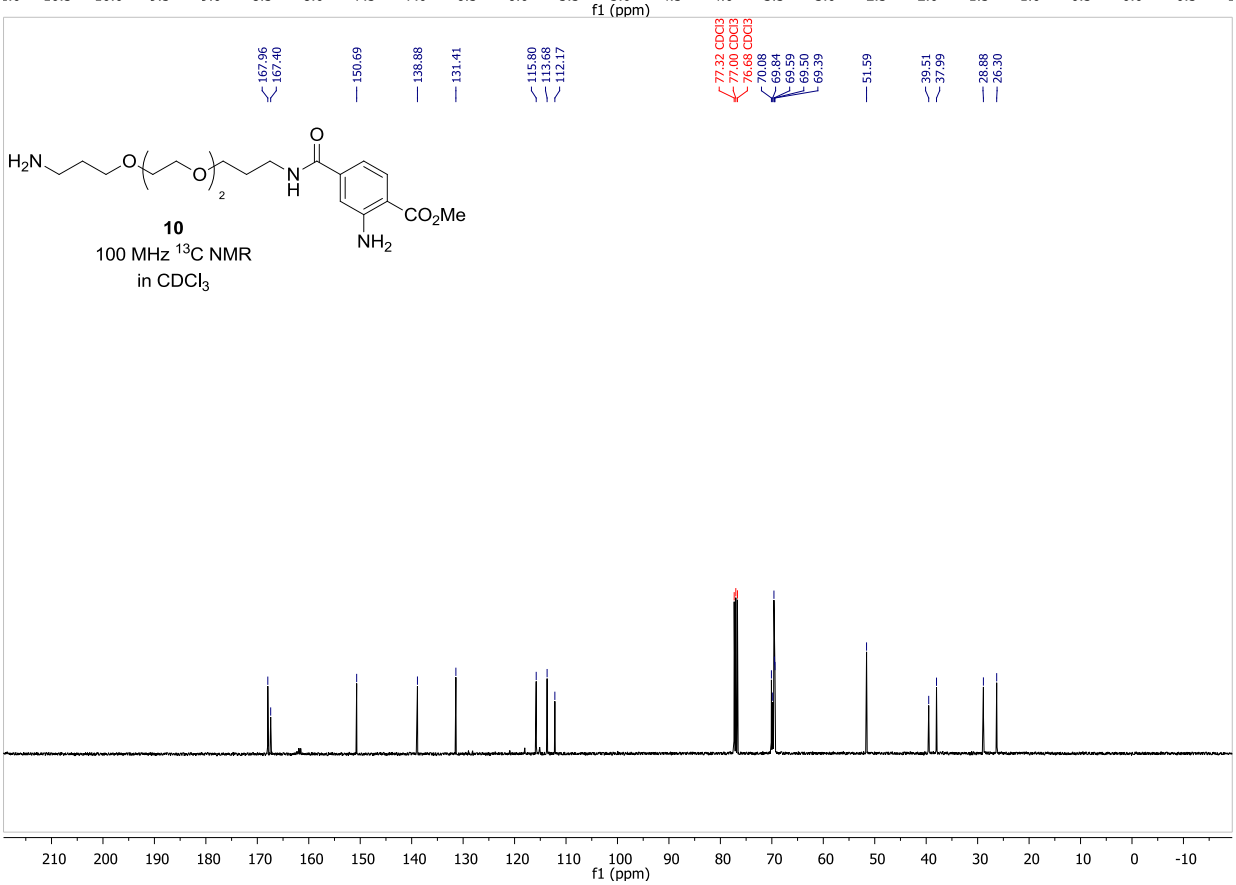

17
600 MHz HSQC
in CD3OD
f2 (ppm)
f1 (ppm)

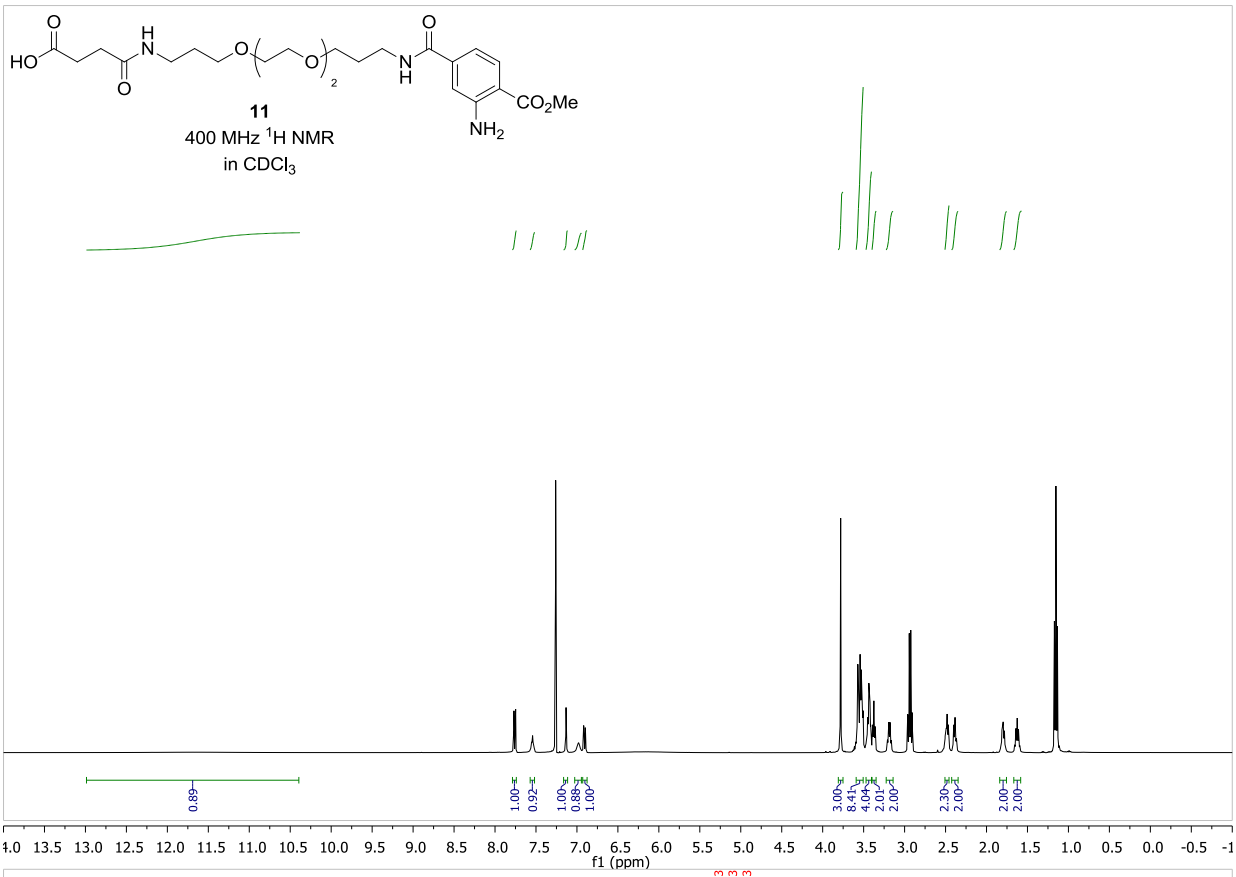

17
600 MHz HMBC
in CD3OD
f2 (ppm)
f1 (ppm)

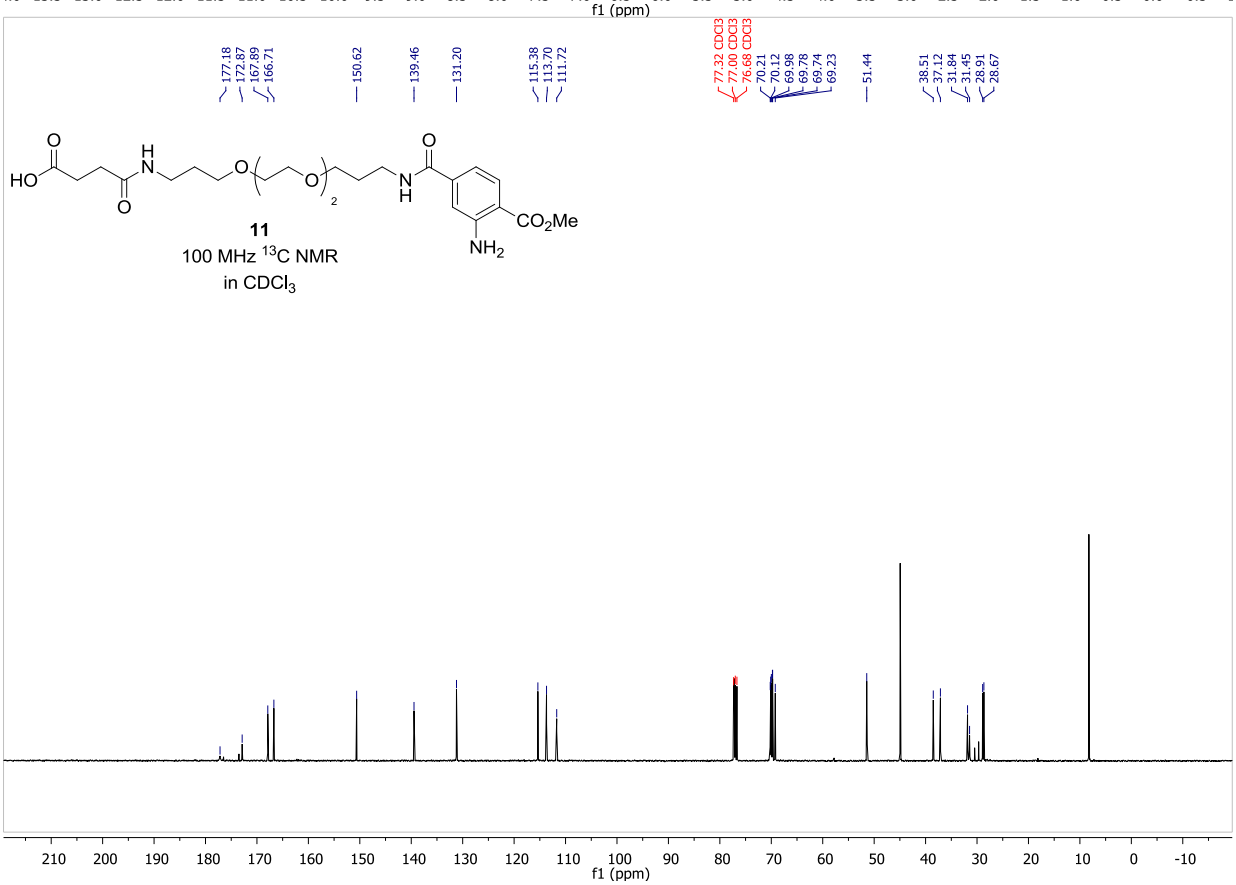

CbzHN
O
N
H
O
O
2
CO2Me
NH2
9
400 MHz 1H NMR
in CDCl3
1.00
6.37
1.00
1.00
2.00
0.81
2.00
3.03
14.27
2.01
2.02
2.02
f1 (ppm)

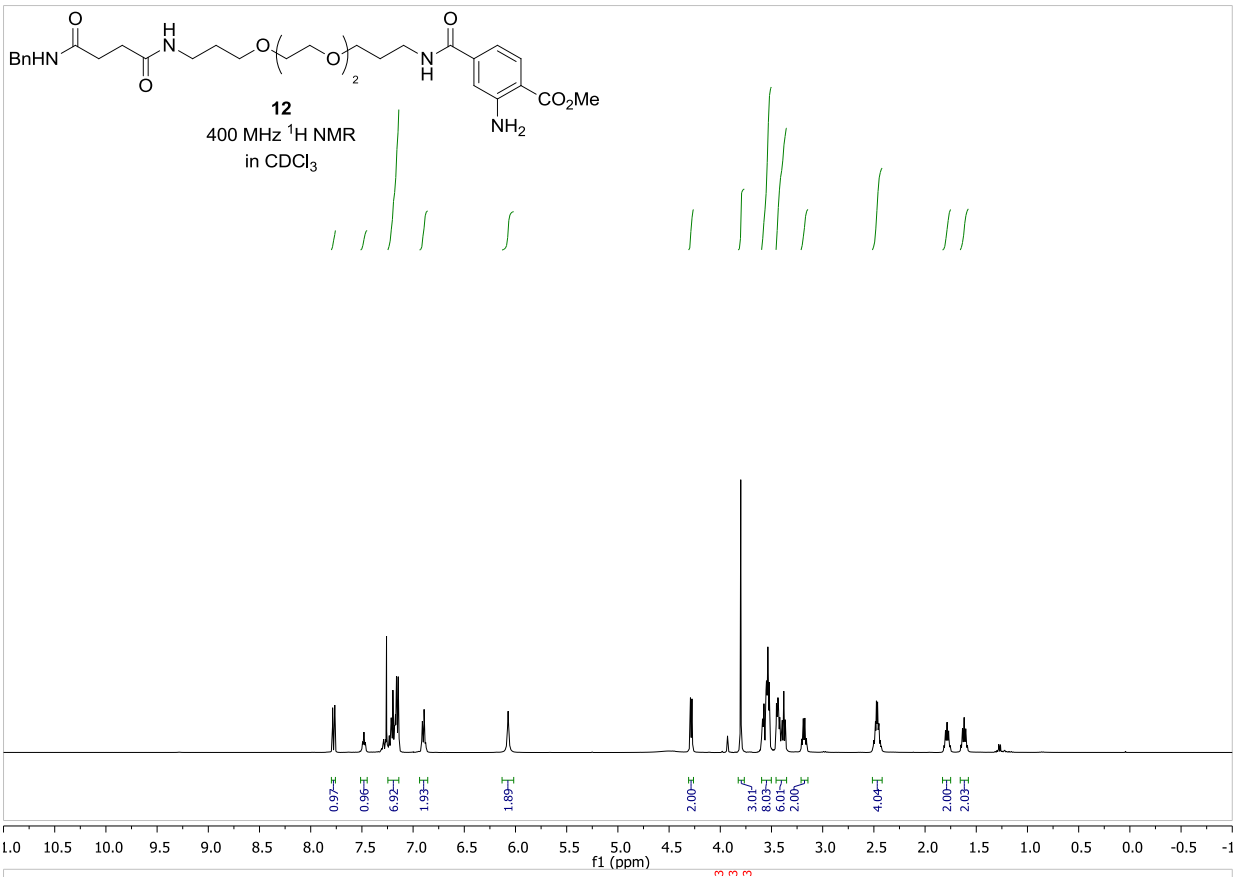

167.94
166.42
156.34
150.49
139.59
136.60
131.32
128.29
127.91
127.84
115.40
113.74
111.99
77.32 CDCl3
77.00 CDCl3
76.68 CDCl3
70.74
70.20
70.01
69.88
69.85
69.35
66.29
51.51
39.00
38.95
29.30
28.54
CbzHN
O
N
H
O
O
2
CO2Me
NH2
9
100 MHz 13C NMR
in CDCl3
f1 (ppm)

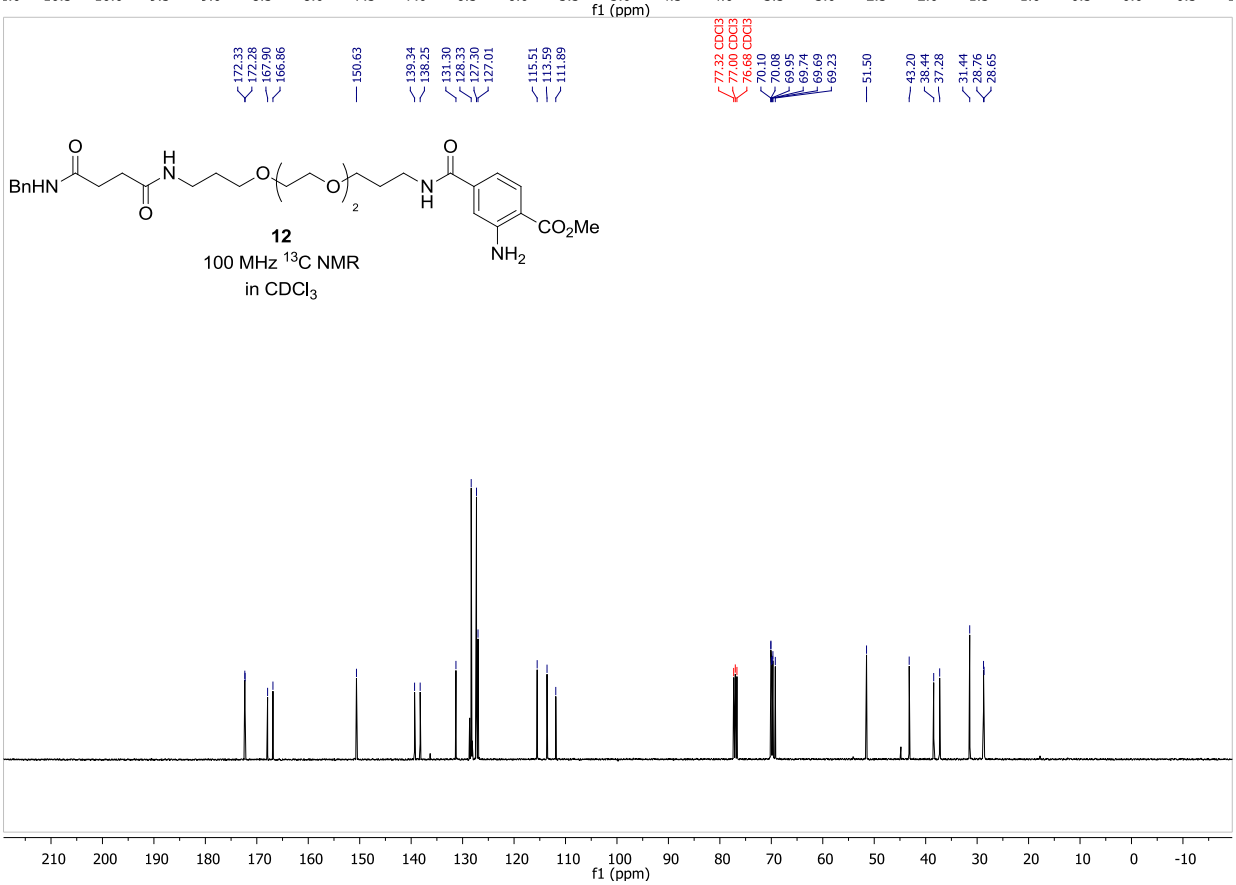

10
400 MHz 1H NMR
in CDCl3
2.45
1.00
0.98
0.99
1.00
2.67
3.05
14.25
2.06
2.01
2.02
f1 (ppm)

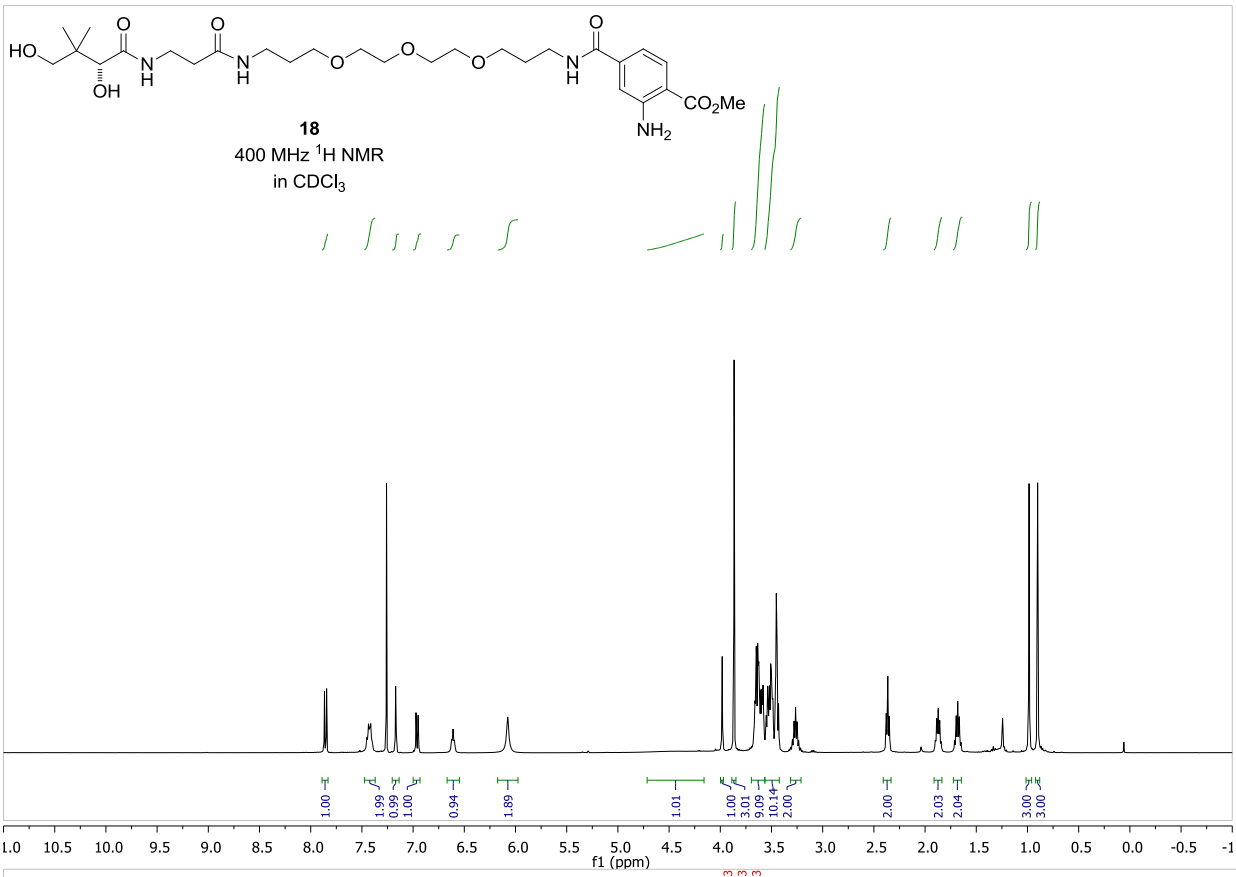

167.96
167.40
150.69
138.88
131.41
115.80
113.68
112.17
77.32 CDCl3
77.00 CDCl3
76.68 CDCl3
70.08
69.84
69.59
69.50
69.39
51.59
39.51
37.99
28.88
26.30
10
100 MHz 13C NMR
in CDCl3
f1 (ppm)

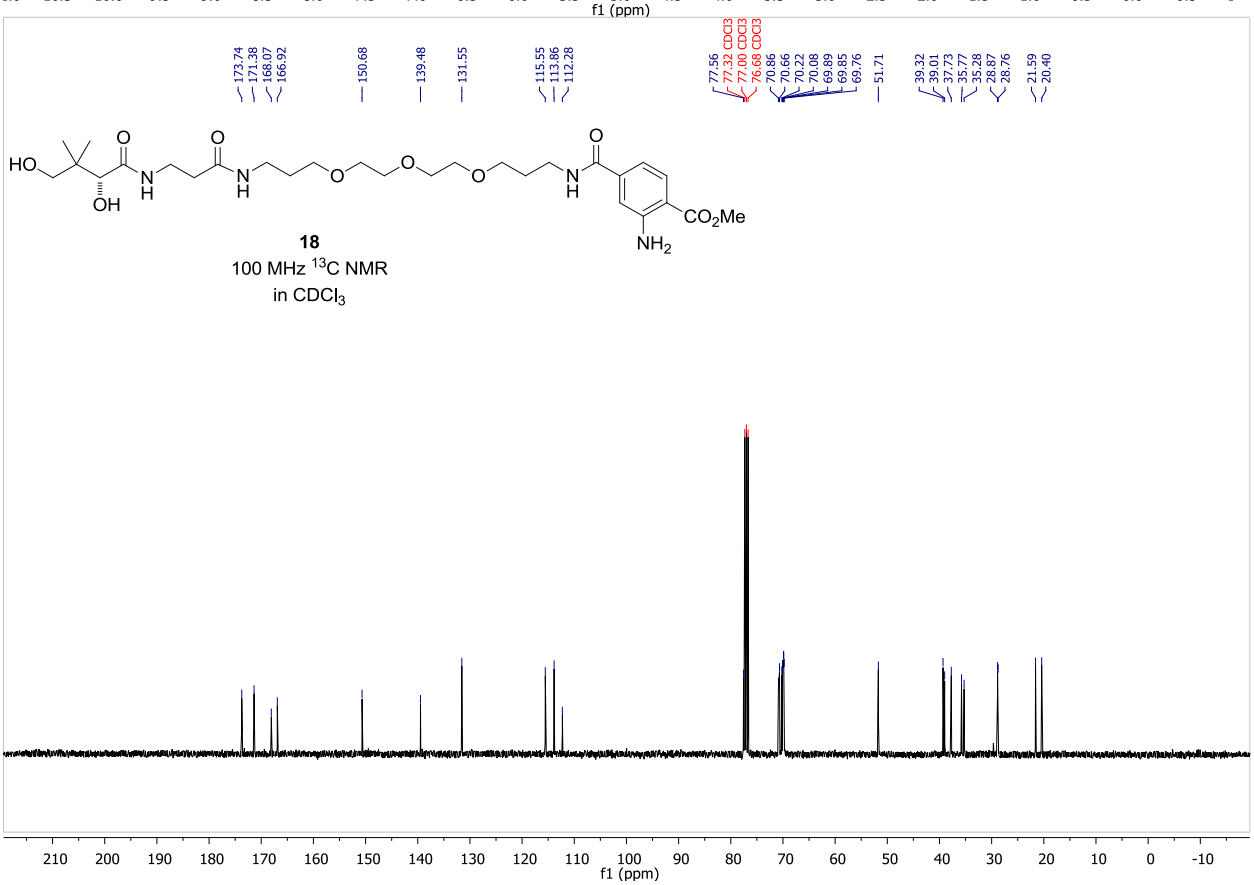

11
400 MHz ^{1}H NMR
in $CDCl_3$
f1 (ppm)

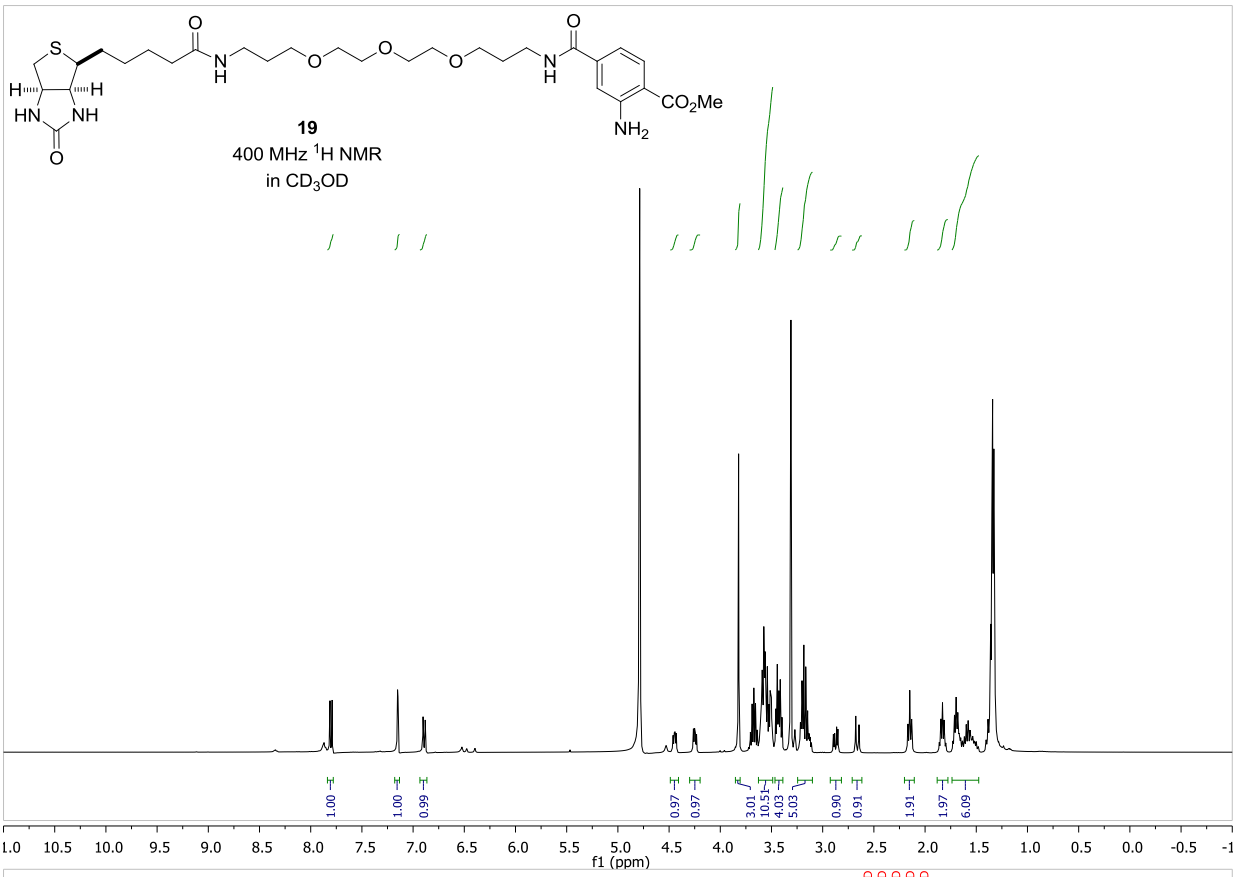

11
100 MHz ^{13}C NMR
in $CDCl_3$
f1 (ppm)

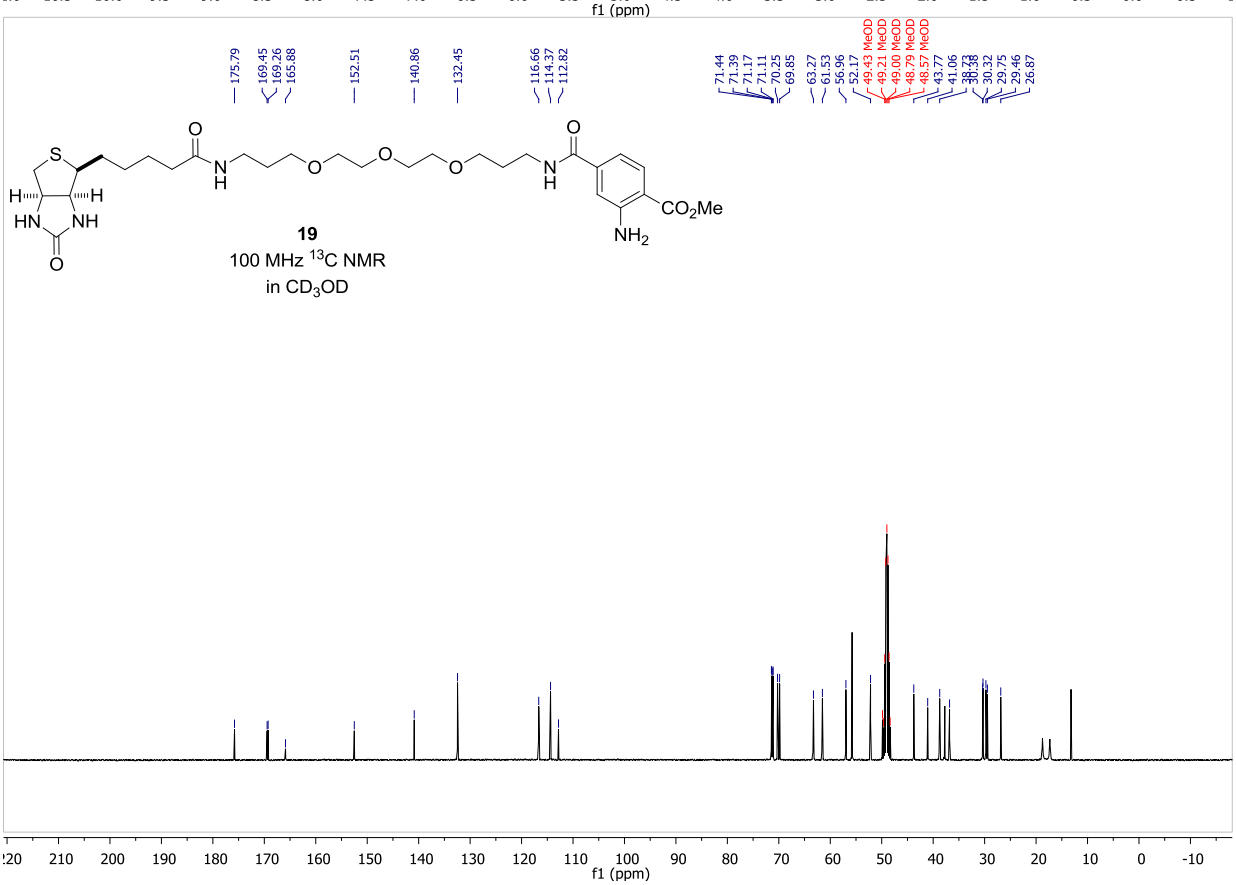

BnHN
O
H
N
O
O
O
O
2
N
H
O
CO2Me
NH2
12
400 MHz 1H NMR
in CDCl3
0.97
0.96
6.92
1.93
1.89
2.00
3.01
8.03
6.01
2.00
4.04
2.00
2.03
f1 (ppm)

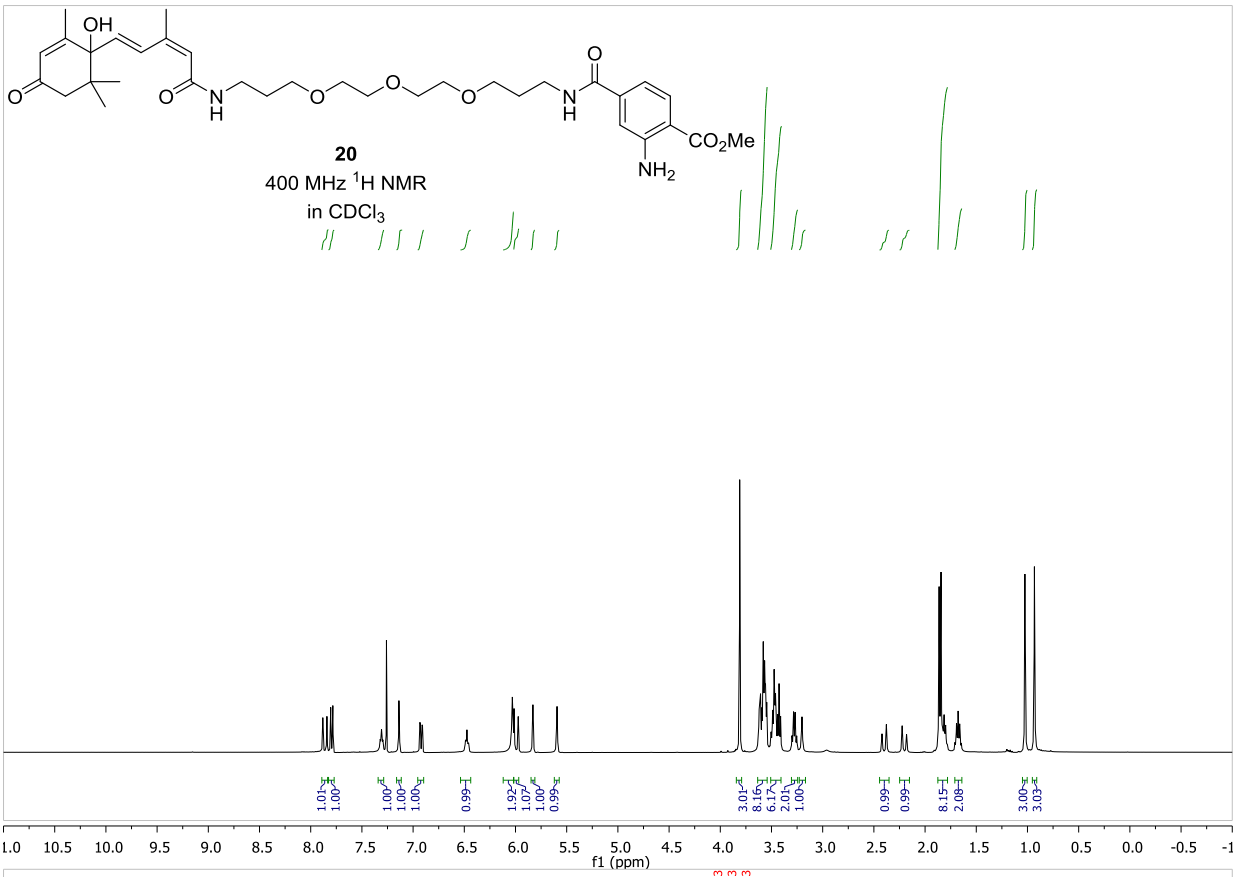

172.33
172.28
167.90
166.86
150.63
139.34
138.25
131.30
128.33
127.30
127.01
115.51
113.59
111.89
77.32 CDCl3
77.00 CDCl3
76.68 CDCl3
70.10
70.08
69.95
69.74
69.69
69.23
51.50
43.20
38.44
37.28
31.44
28.76
28.65
BnHN
O
H
N
O
O
O
O
2
N
H
O
CO2Me
NH2
12
100 MHz 13C NMR
in CDCl3
f1 (ppm)

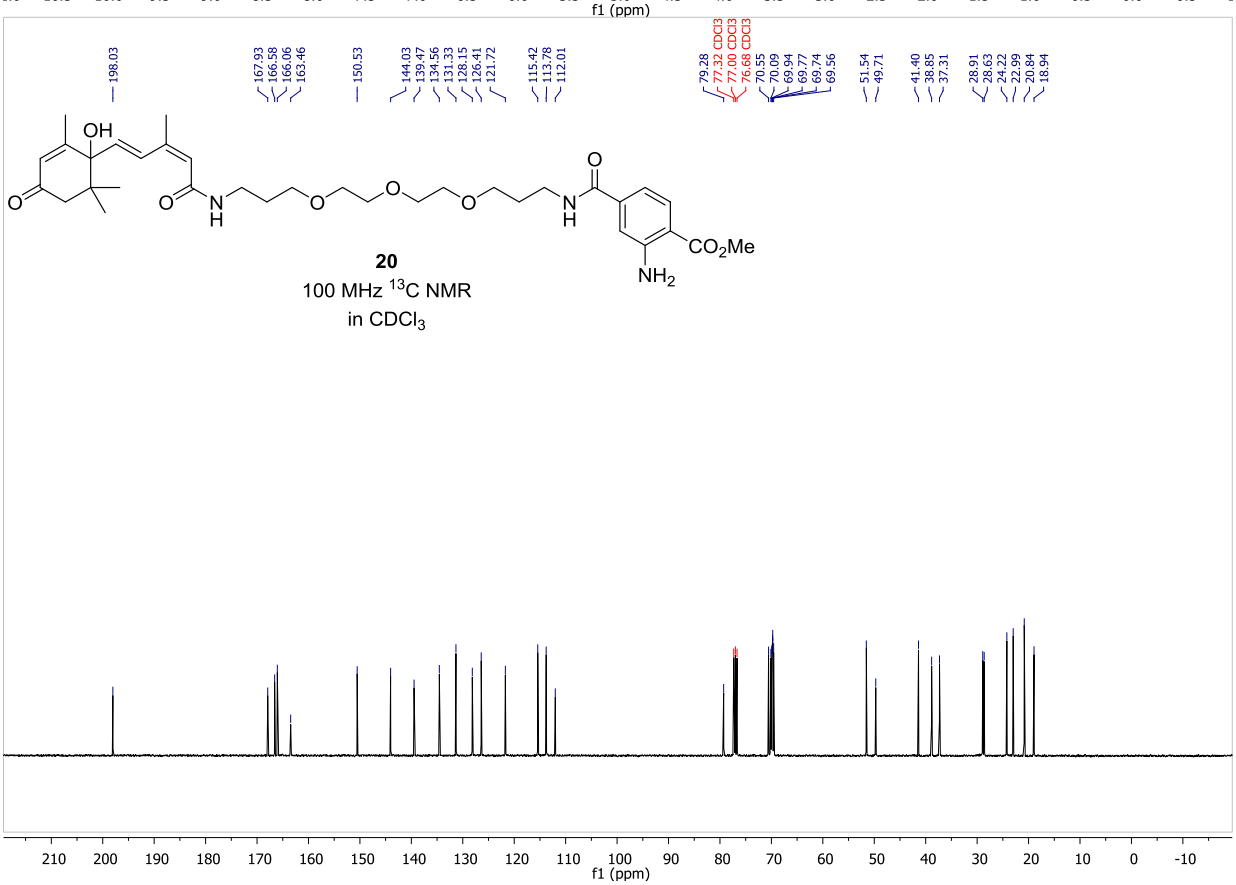

HO
OH
O
N
H
O
N
H
O
O
O
N
H
O
CO2Me
NH2
18
400 MHz 1H NMR
in CDCl3
1.00
1.99
0.99
1.00
0.94
1.89
1.01
1.00
3.01
9.09
10.14
2.00
2.00
2.03
2.04
3.00
3.00
f1 (ppm)

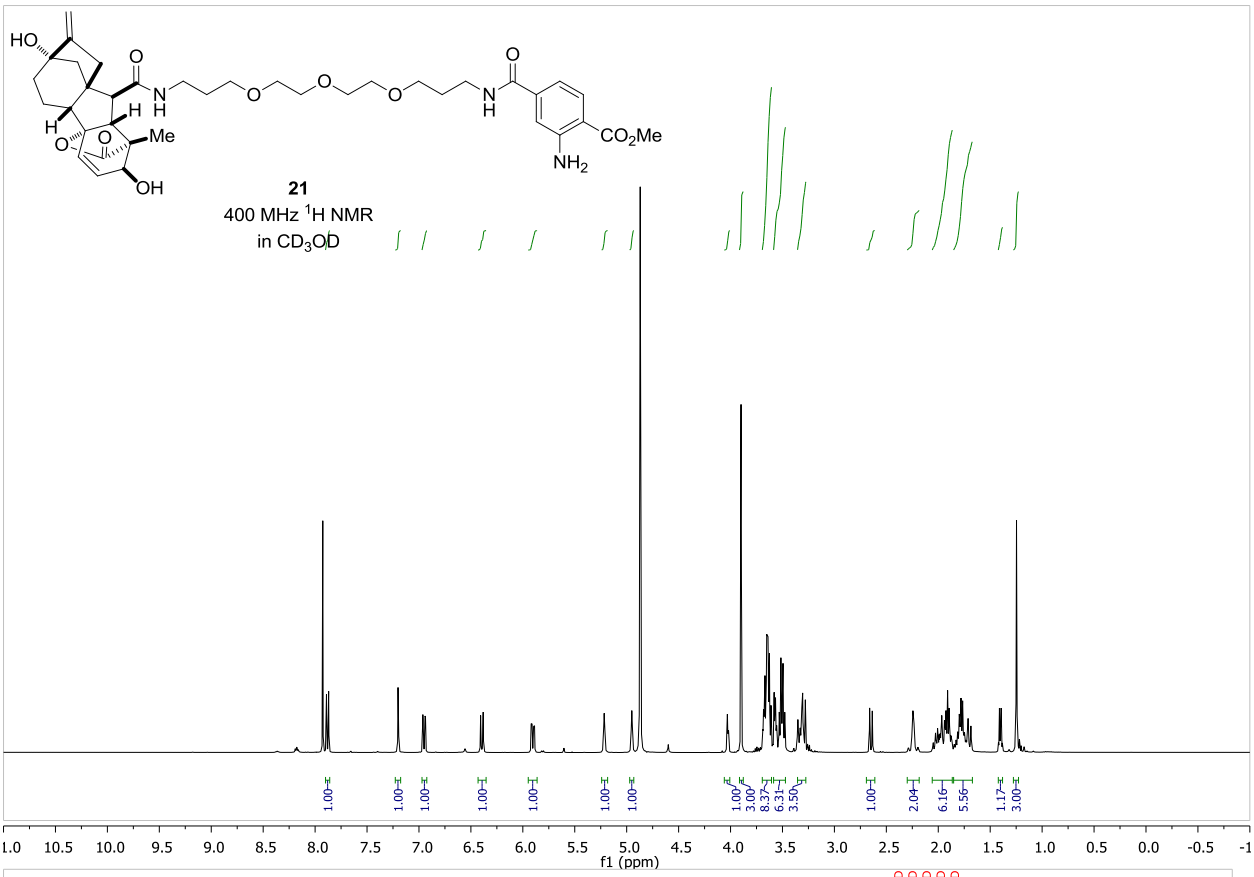

173.74
171.38
168.07
166.92
150.68
139.48
131.55
115.55
113.86
112.28
77.56 CDCl3
77.32 CDCl3
77.00 CDCl3
76.68 CDCl3
70.86
70.66
70.22
70.08
69.89
69.85
69.76
51.71
39.32
39.01
37.73
35.77
35.28
28.87
28.76
21.59
20.40
HO
OH
O
N
H
O
N
H
O
O
O
N
H
O
CO2Me
NH2
18
100 MHz 13C NMR
in CDCl3
f1 (ppm)

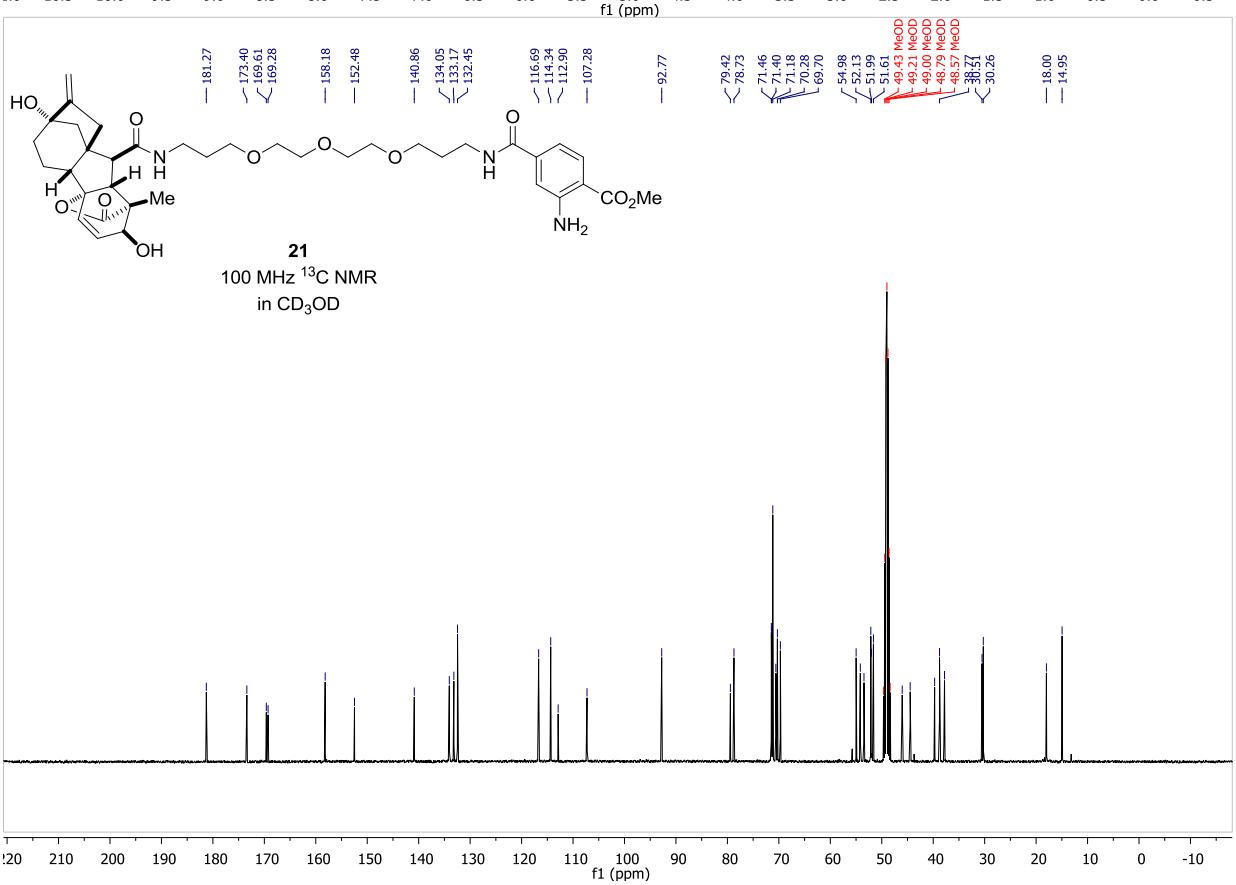

19
400 MHz 1H NMR
in CD3OD
1.00
1.00
0.99
0.97
0.97
3.01
10.51
4.03
5.03
0.90
0.91
1.91
1.97
6.09
f1 (ppm)

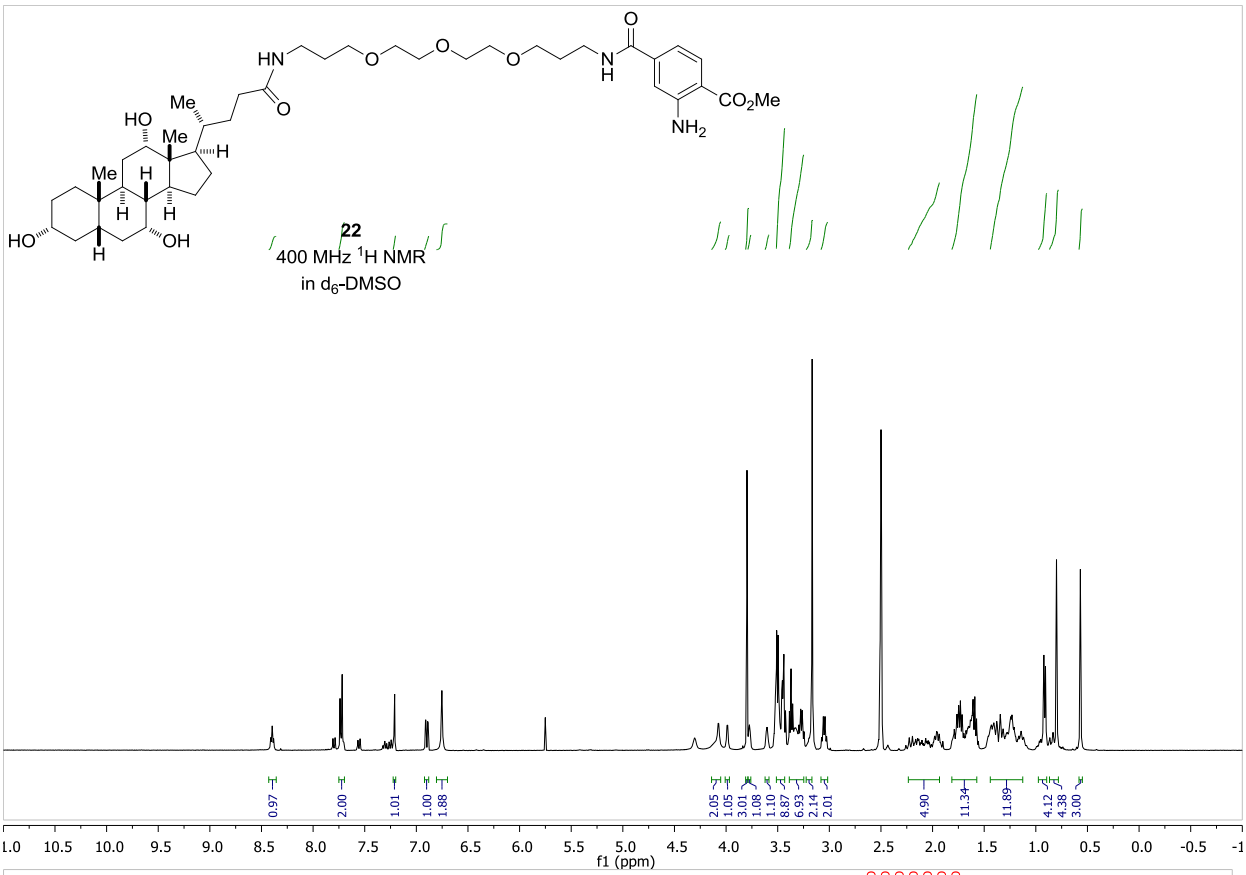

175.79
169.45
169.26
165.88
152.51
140.86
132.45
116.66
114.37
112.82
71.44
71.39
71.17
71.11
70.25
69.85
63.27
61.53
56.96
52.17
49.43 MeOD
49.21 MeOD
49.00 MeOD
48.79 MeOD
48.57 MeOD
43.77
41.06
30.32
29.75
29.46
26.87
19
100 MHz 13C NMR
in CD3OD
f1 (ppm)

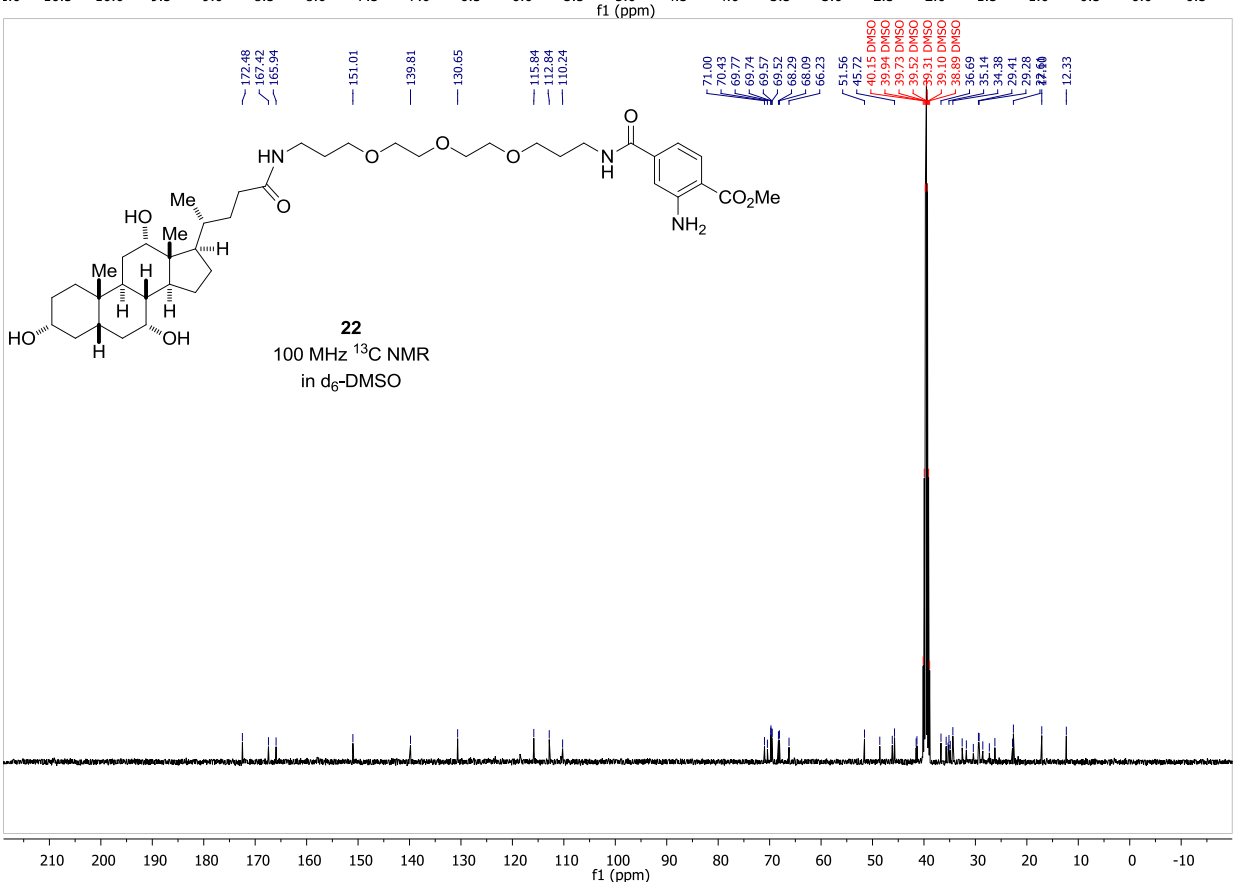

OH
O
O
N
H
O
O
O
O
N
H
O
CO2Me
NH2
20
400 MHz 1H NMR
in CDCl3
1.01
1.00
1.00
1.00
1.00
0.99
1.92
1.07
1.00
0.99
3.01
8.16
6.17
2.01
1.00
0.99
0.99
8.15
2.08
3.00
3.03
1.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1
f1 (ppm)

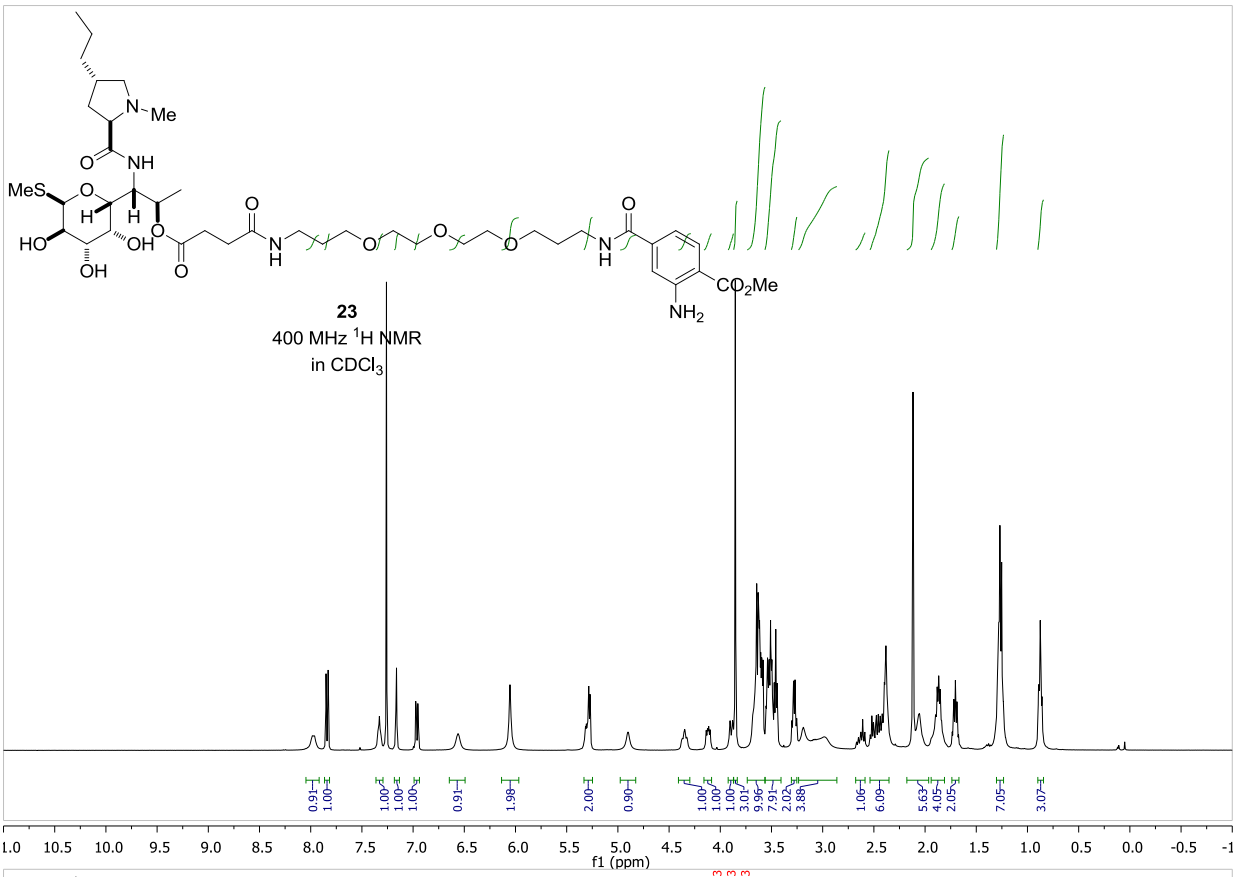

198.03
167.93
166.58
166.06
163.46
150.53
144.03
139.47
134.56
131.33
128.15
126.41
121.72
115.42
113.78
112.01
79.28
77.32 CDCl3
77.00 CDCl3
76.68 CDCl3
70.55
70.09
69.94
69.77
69.74
69.56
51.54
49.71
41.40
38.85
37.31
28.91
28.63
24.22
22.99
20.84
18.94
OH
O
O
N
H
O
O
O
O
N
H
O
CO2Me
NH2
20
100 MHz 13C NMR
in CDCl3
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

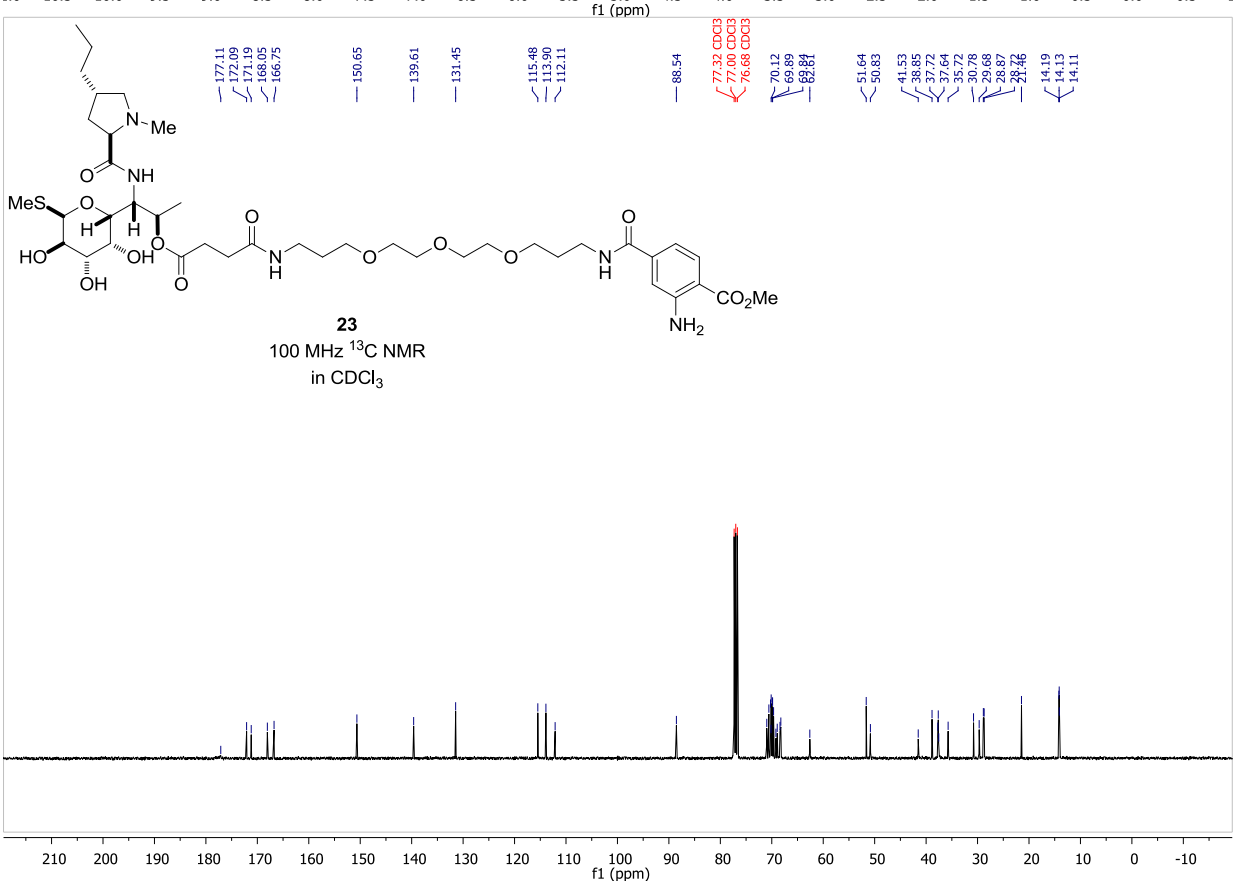

HO
O
N
H
H
O
O
Me
OH
O
N
H
O
O
O
CO2Me
NH2
21
400 MHz 1H NMR
in CD3OD
1.00
1.00
1.00
1.00
1.00
1.00
1.00
1.00
3.00
8.37
6.31
3.50
1.00
2.04
6.16
5.56
1.17
3.00
1.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1
f1 (ppm)

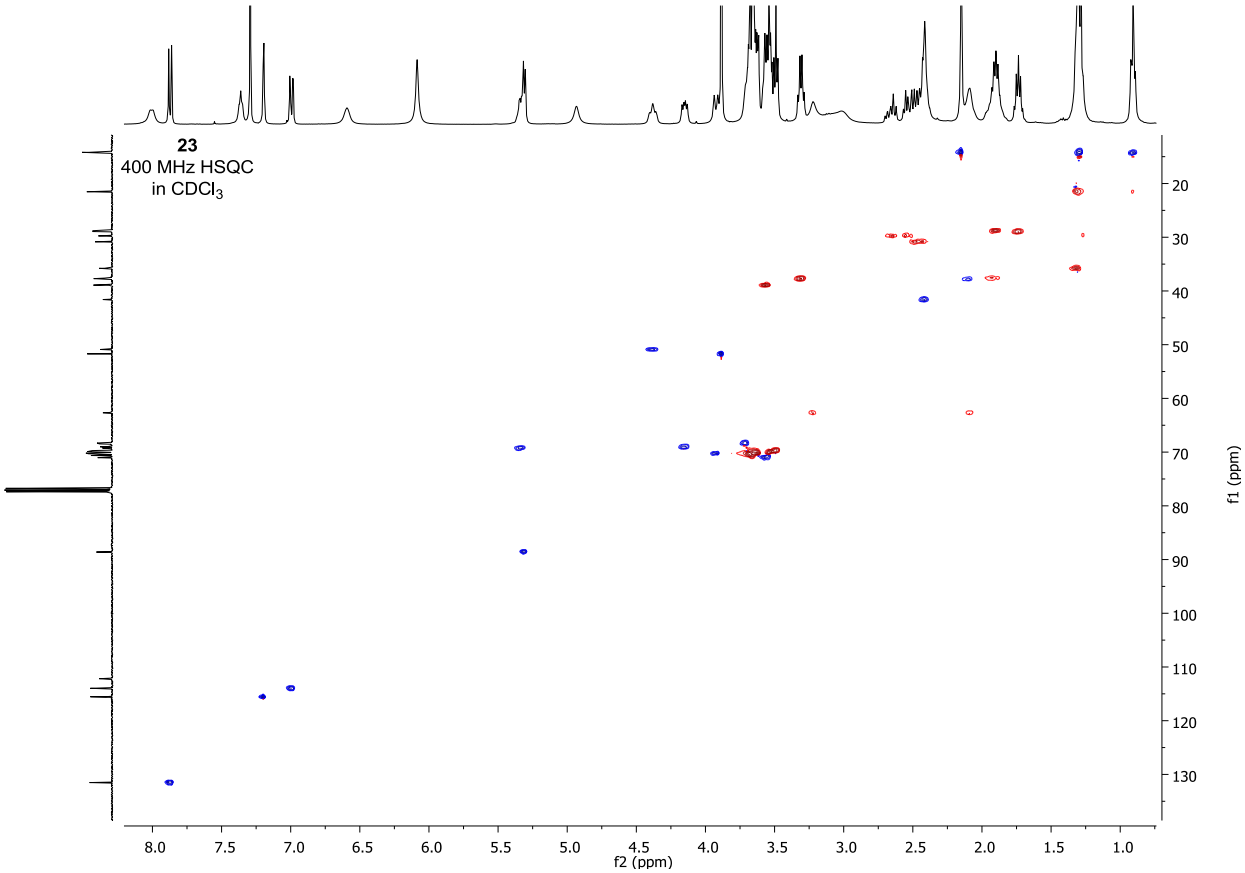

181.27
173.40
169.61
169.28
158.18
152.48
140.86
134.05
133.17
132.45
116.69
114.34
112.90
107.28
92.77
79.42
78.73
71.46
71.40
71.18
70.28
69.70
54.98
52.13
51.99
51.61
49.43 MeOD
49.21 MeOD
49.00 MeOD
48.79 MeOD
48.57 MeOD
38.27
30.51
30.26
18.00
14.95
HO
O
N
H
H
O
O
Me
OH
O
N
H
O
O
O
CO2Me
NH2
21
100 MHz 13C NMR
in CD3OD
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

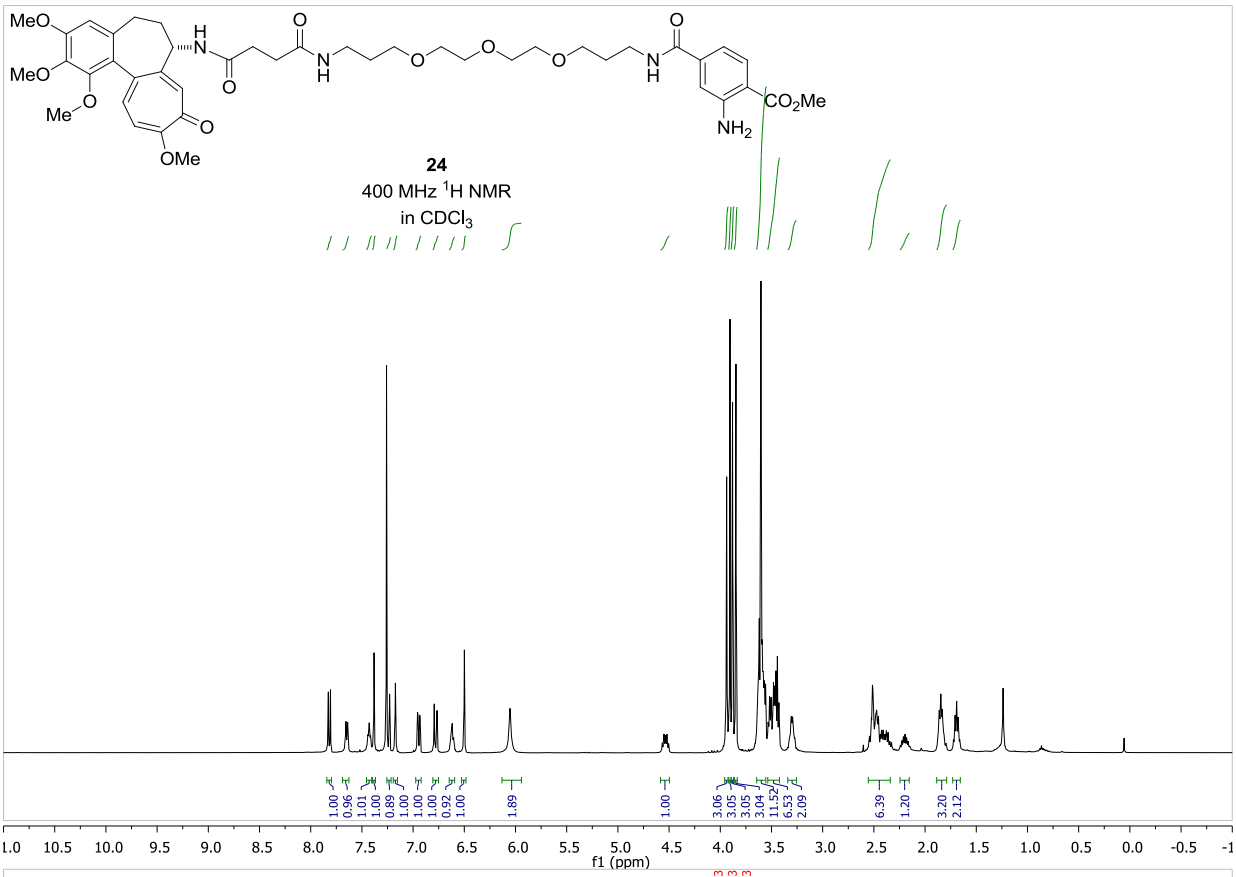

22
400 MHz ^{1}H NMR
in d_6-DMSO
f1 (ppm)

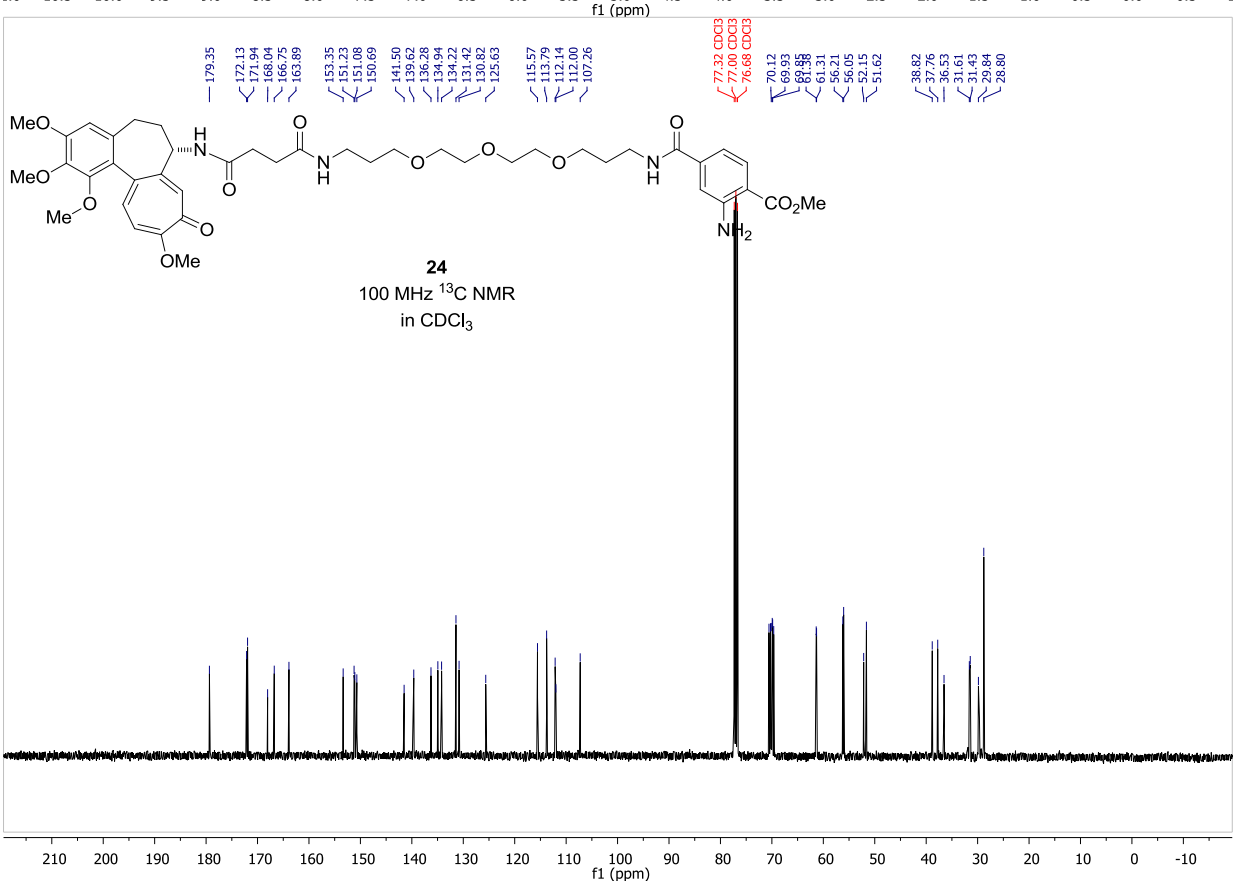

22
100 MHz ^{13}C NMR
in d_6-DMSO
f1 (ppm)


23
400 MHz ^{1}H NMR
in $CDCl_3$
f1 (ppm)

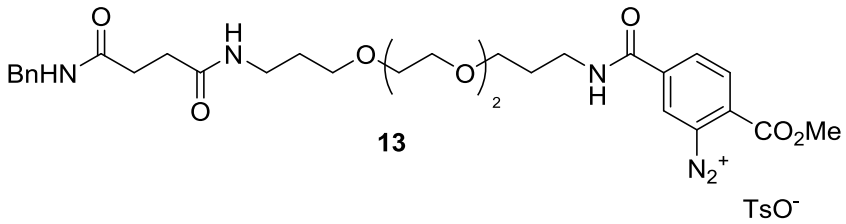

23
100 MHz ^{13}C NMR
in $CDCl_3$
f1 (ppm)


23
400 MHz HSQC
in CDCl3
f2 (ppm)
f1 (ppm)

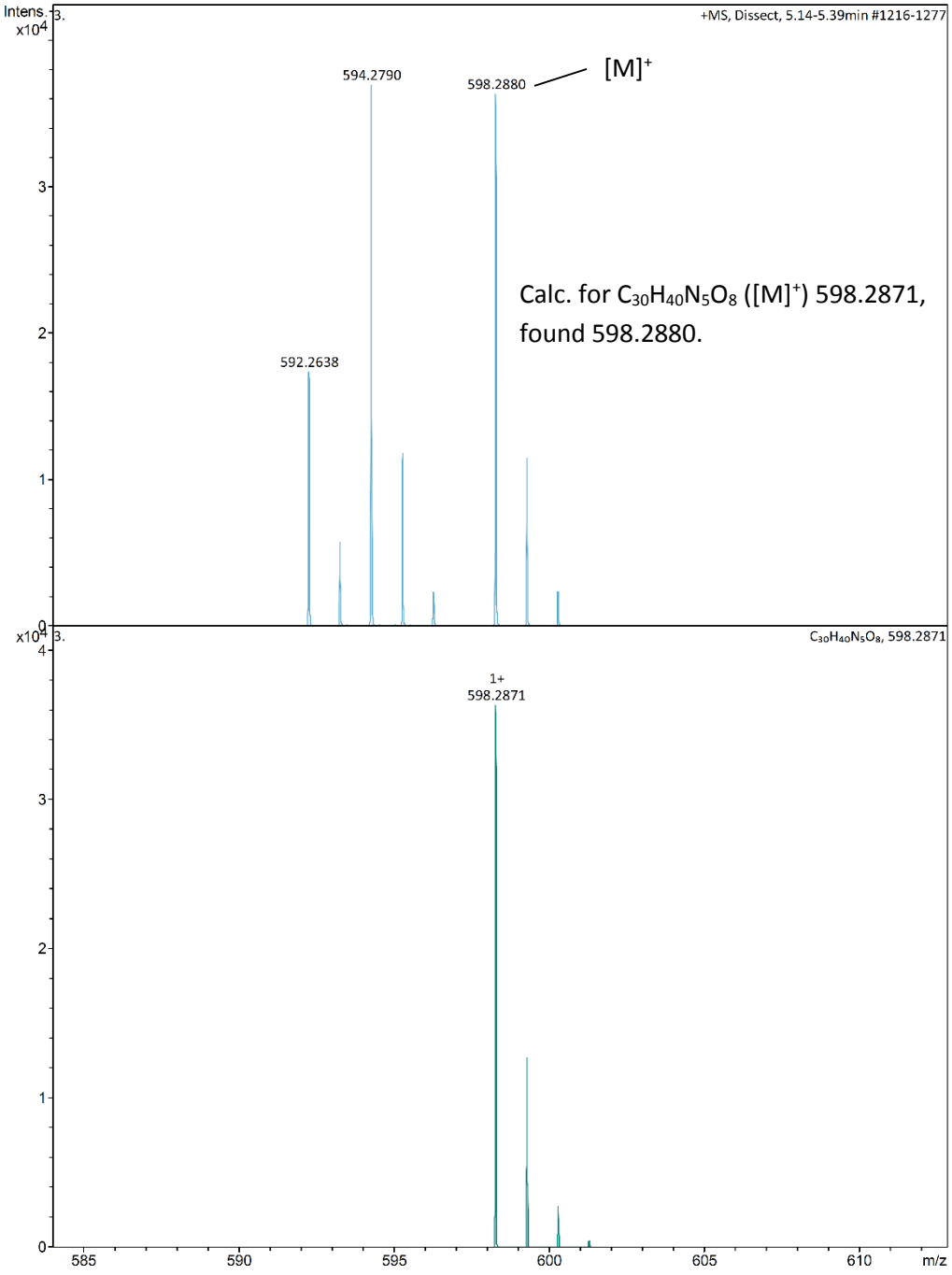

MeO
MeO
Me
OMe
H
N
O
N
H
O
O
O
O
O
N
H
O
CO2Me
NH2
24
400 MHz 1H NMR
in CDCl3
1.00
0.96
1.01
1.00
0.89
1.00
1.00
1.00
0.92
1.00
1.89
1.00
3.06
3.05
3.05
3.04
11.52
6.53
2.09
6.39
1.20
3.20
2.12
f1 (ppm)

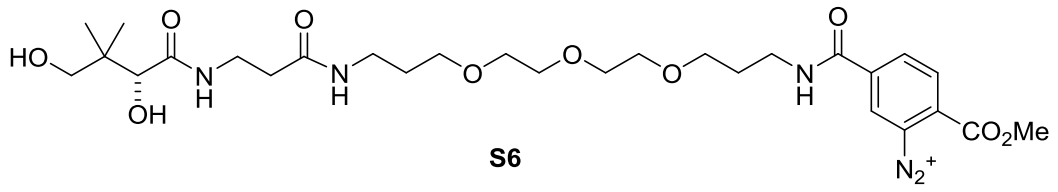

179.35
172.13
171.94
168.04
166.75
163.89
153.35
151.23
151.08
150.69
141.50
139.62
136.28
134.94
134.22
131.42
130.82
125.63
115.57
113.79
112.14
112.00
107.26
77.32 CDCl3
77.00 CDCl3
76.68 CDCl3
70.12
69.93
69.86
69.38
61.31
56.21
56.05
52.15
51.62
38.82
37.76
36.53
36.61
31.43
29.84
28.80
MeO
MeO
Me
OMe
H
N
O
N
H
O
O
O
O
O
N
H
O
CO2Me
NH2
24
100 MHz 13C NMR
in CDCl3
f1 (ppm)

7. MS Spectra of Diazonium Salts

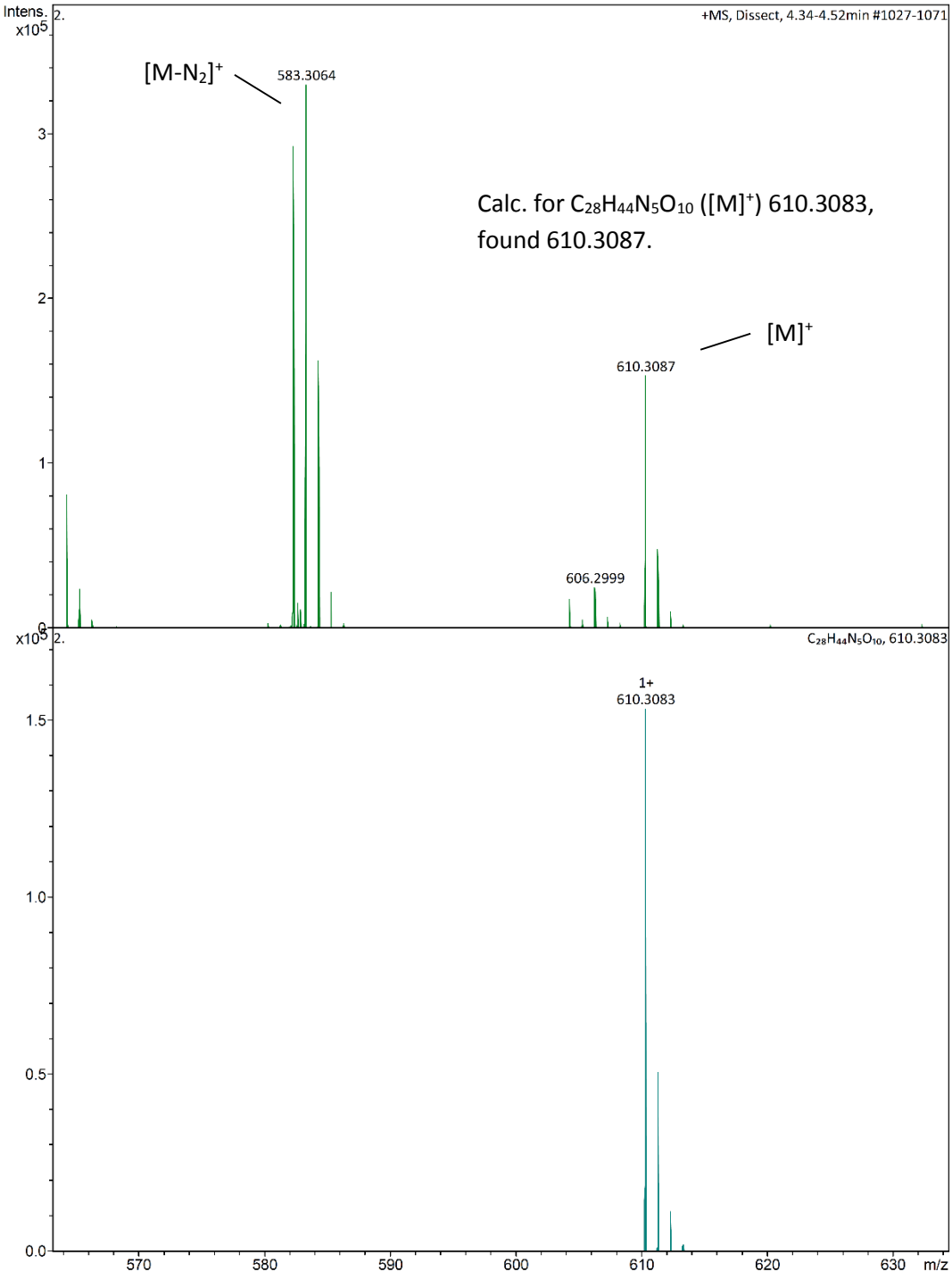


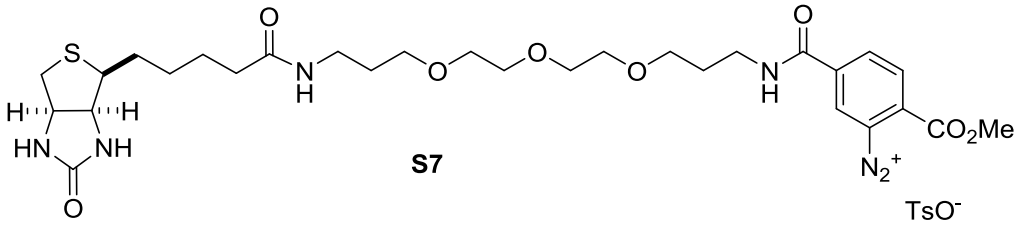
13



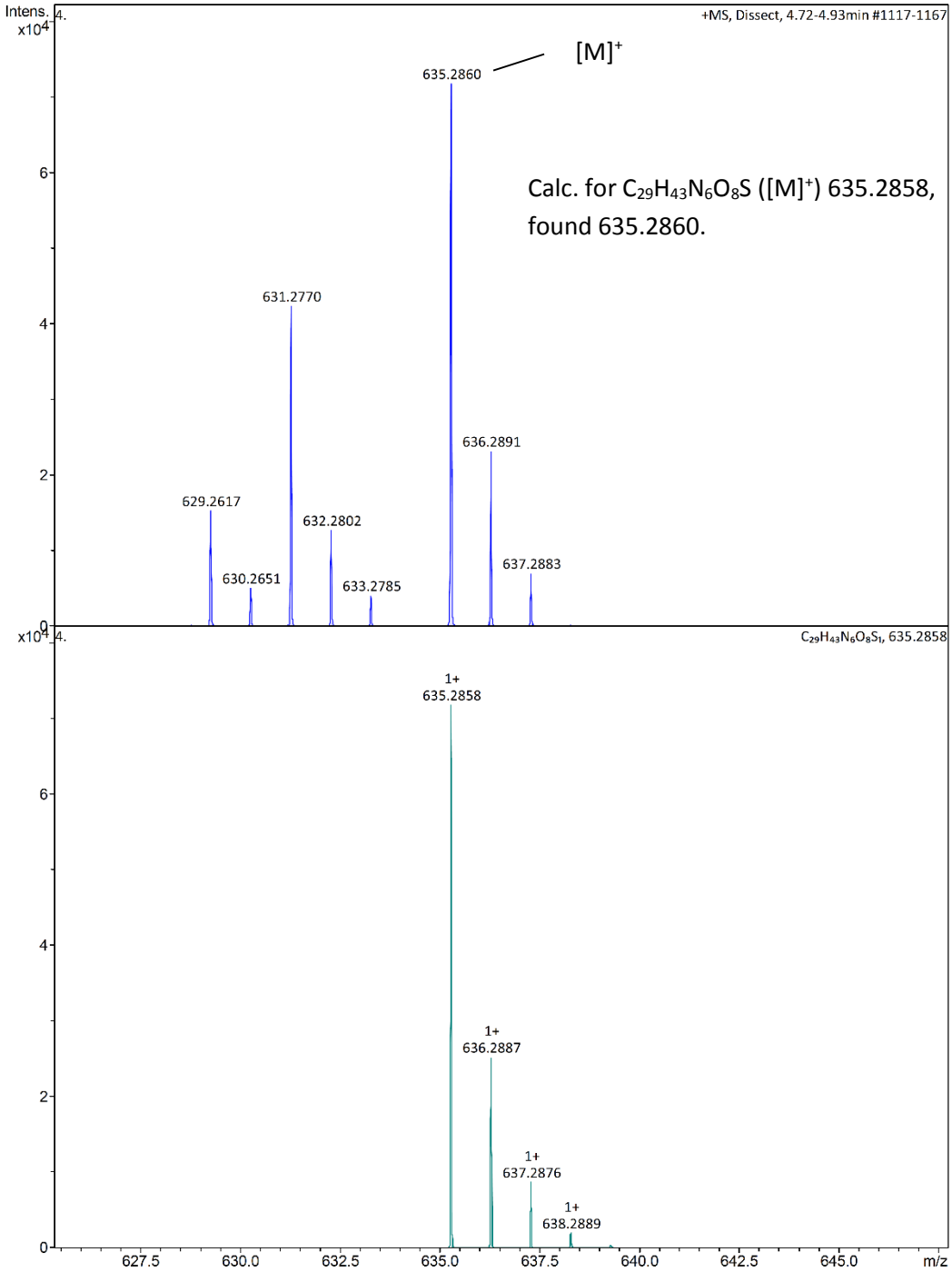


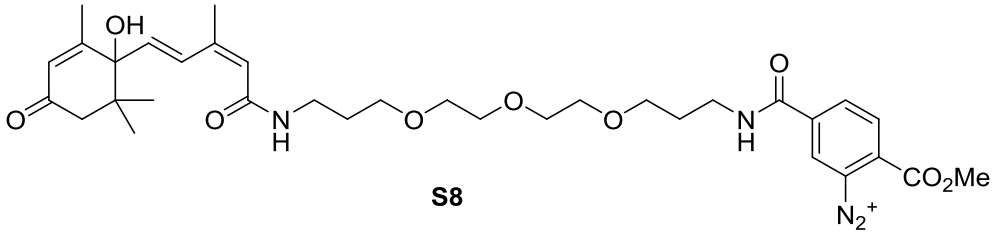
S6


Intens.
x10^5
2.
+MS, Dissect, 4.34-4.52min #1027-1071
[M-N2]+
583.3064
Calc. for C28H44N5O10 ([M]+) 610.3083,
found 610.3087.
[M]+
610.3087
606.2999
x10^5
2.
C28H44N5O10, 610.3083
1+
610.3083
m/z


S
H
H
HN
NH
O
O
N
H
O
O
O
N
H
O
CO2Me
N2+
TsO-

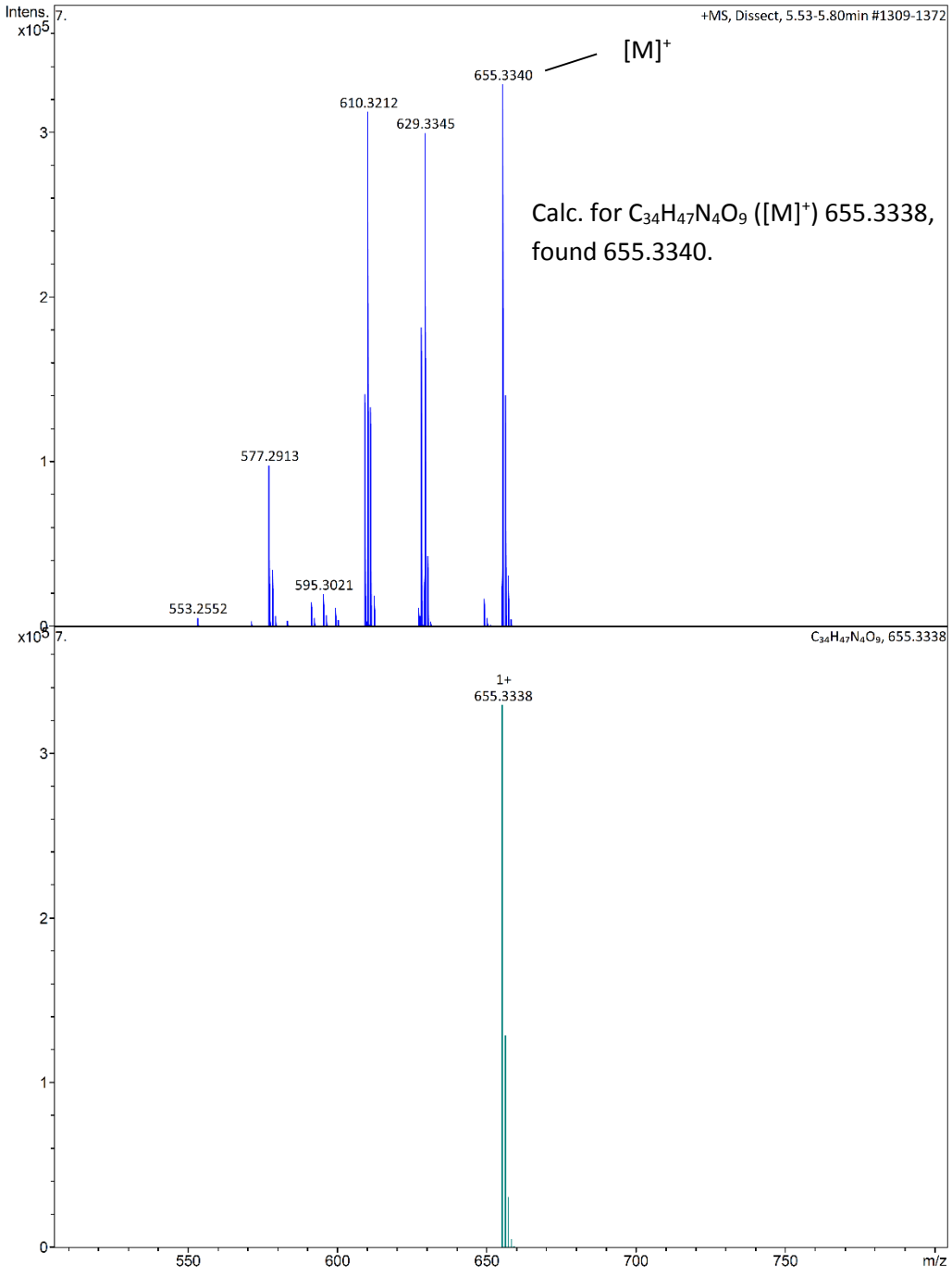
S7

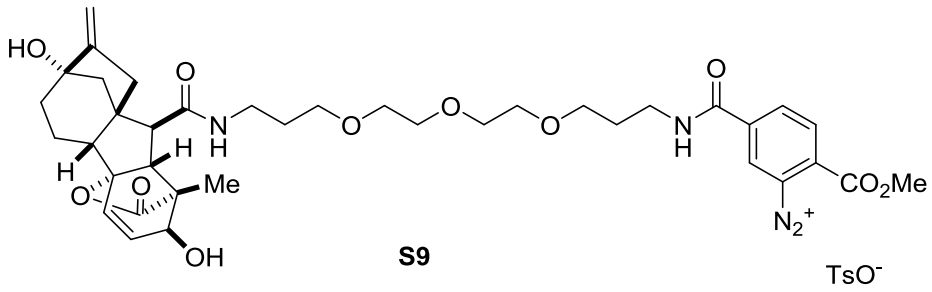

Intens.
x10^4
4.
+MS, Dissect, 4.72-4.93min #1117-1167
[M]+
635.2860
631.2770
636.2891
629.2617
632.2802
630.2651
633.2785
637.2883
Calc. for C29H43N6O8S ([M]+) 635.2858, found 635.2860.
C29H43N6O8S1, 635.2858
1+
635.2858
1+
636.2887
1+
637.2876
1+
638.2889
627.5
630.0
632.5
635.0
637.5
640.0
642.5
645.0
m/z



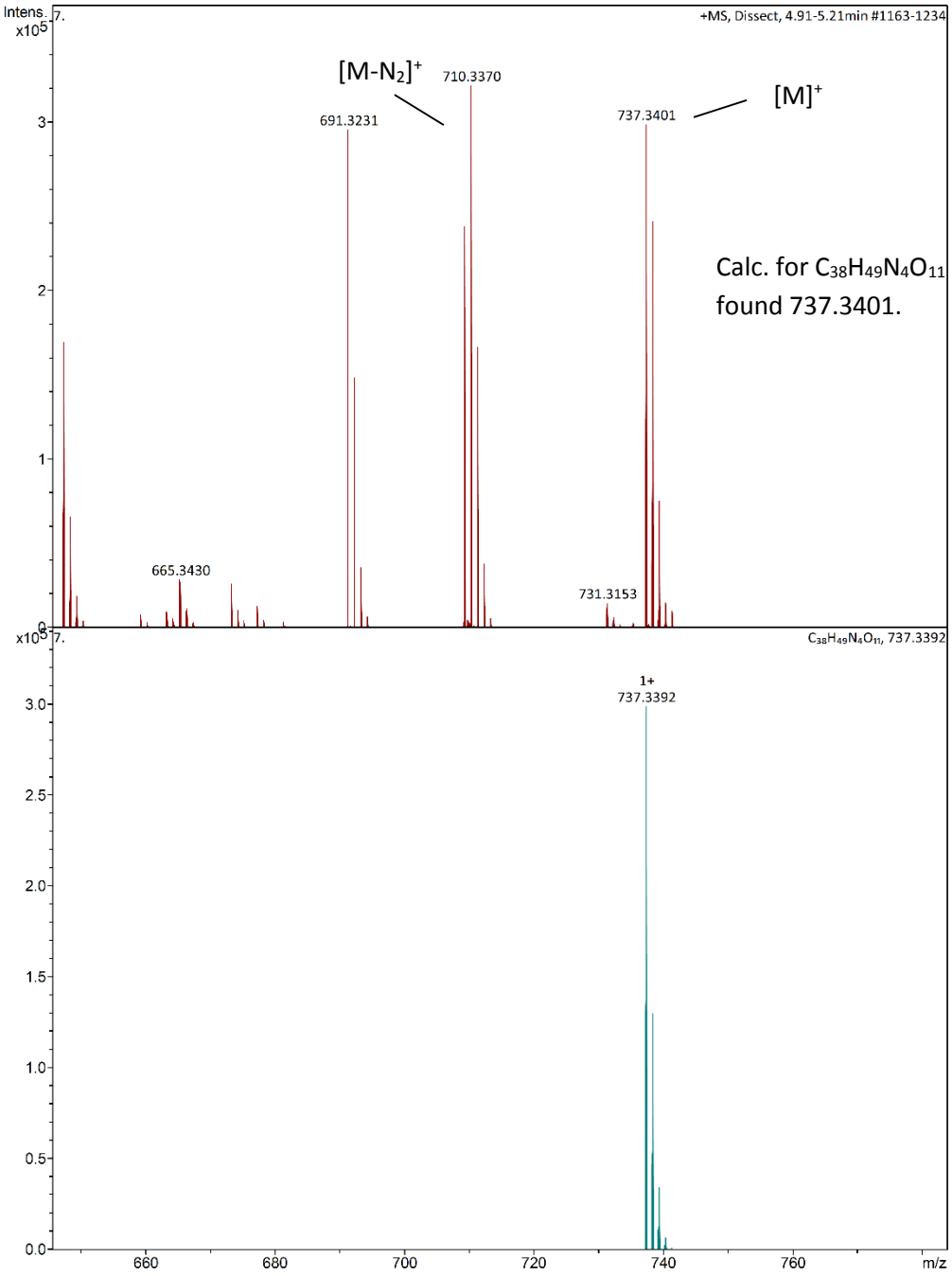
S8

TsO$^-$

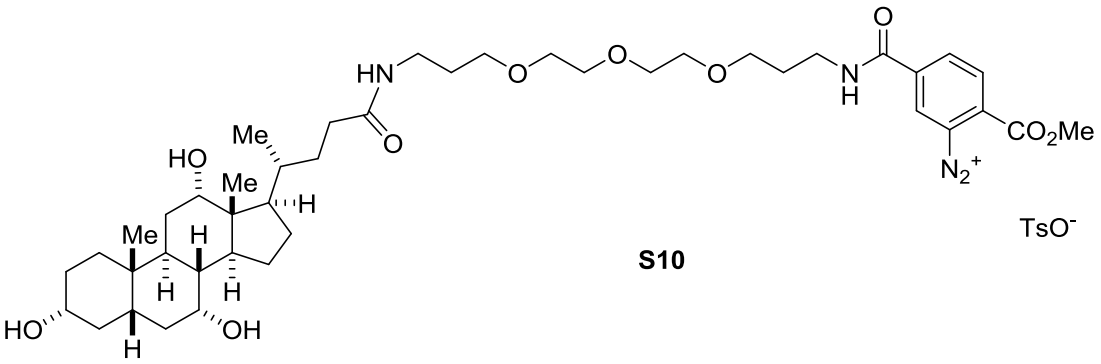

Intens.
x10^5
7.
+MS, Dissect, 5.53-5.80min #1309-1372
[M]+
655.3340
610.3212
629.3345
577.2913
595.3021
553.2552
Calc. for C34H47N4O9 ([M]+) 655.3338,
found 655.3340.
x10^5
7.
C34H47N4O9, 655.3338
1+
655.3338
m/z



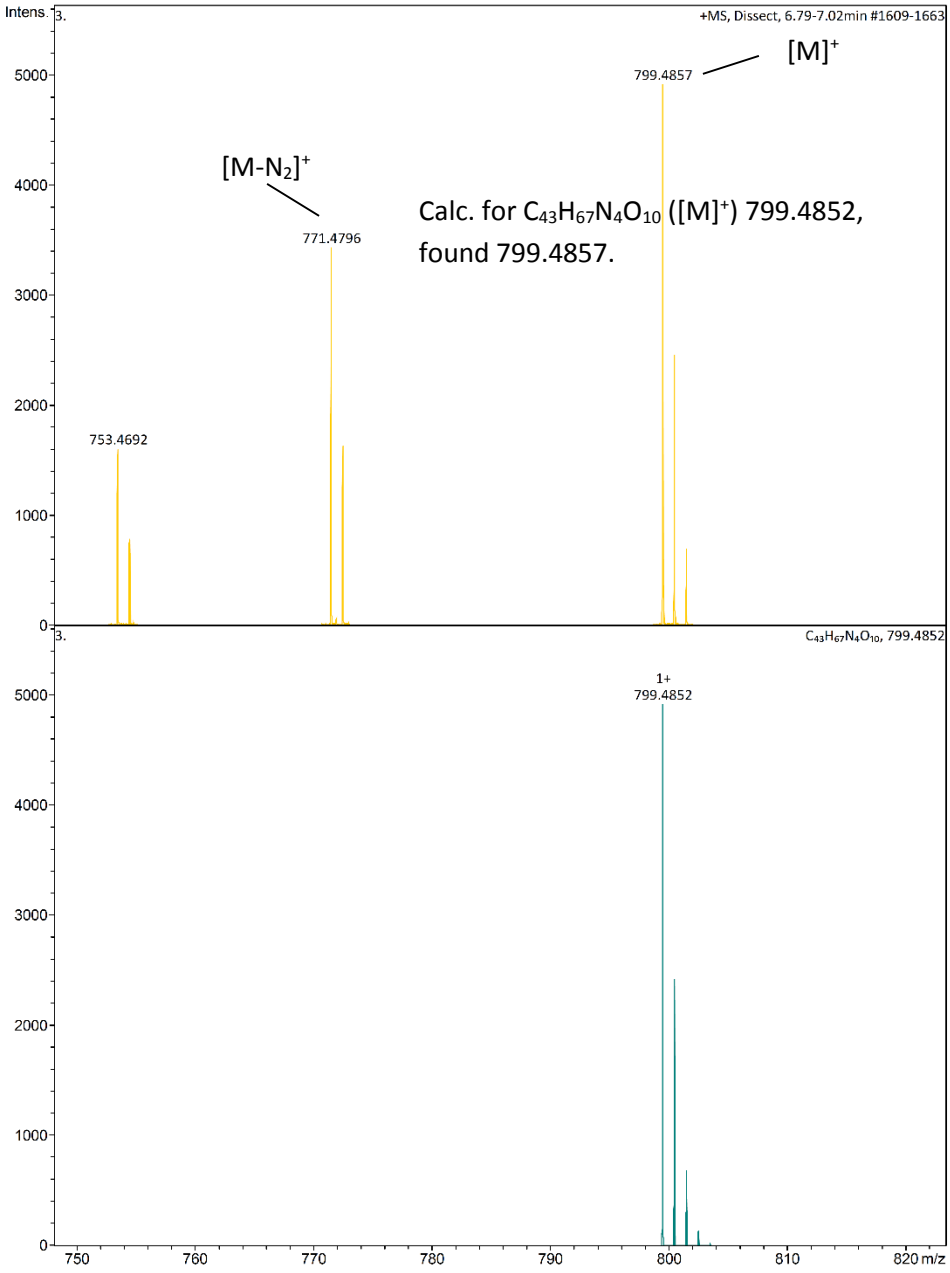
S9

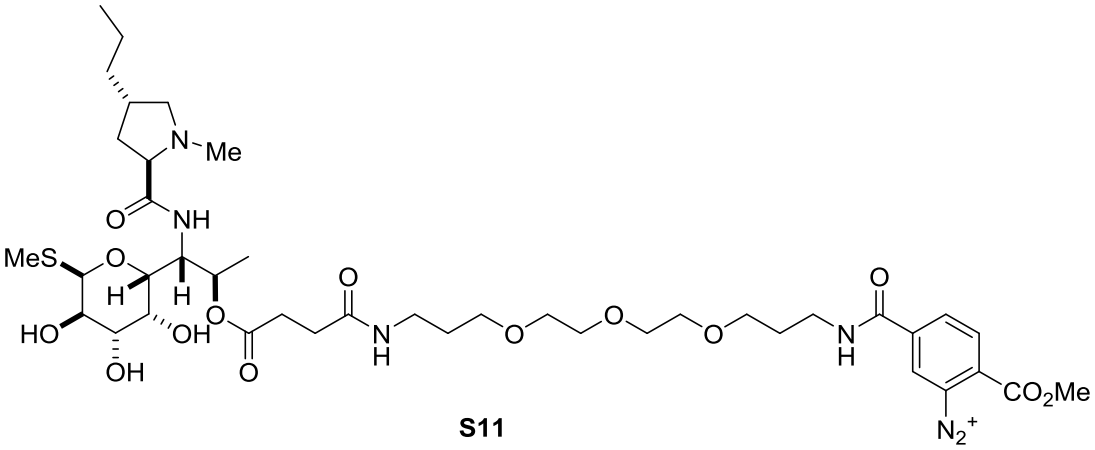


Calc. for $C_{38}H_{49}N_4O_{11}$ ($[M]^+$) 737.3392, found 737.3401.


HN
O
O
O
N
H
O
CO2Me
N2+
TsO-
Me
HO
Me
H
Me
H
H
H
HO
OH
H

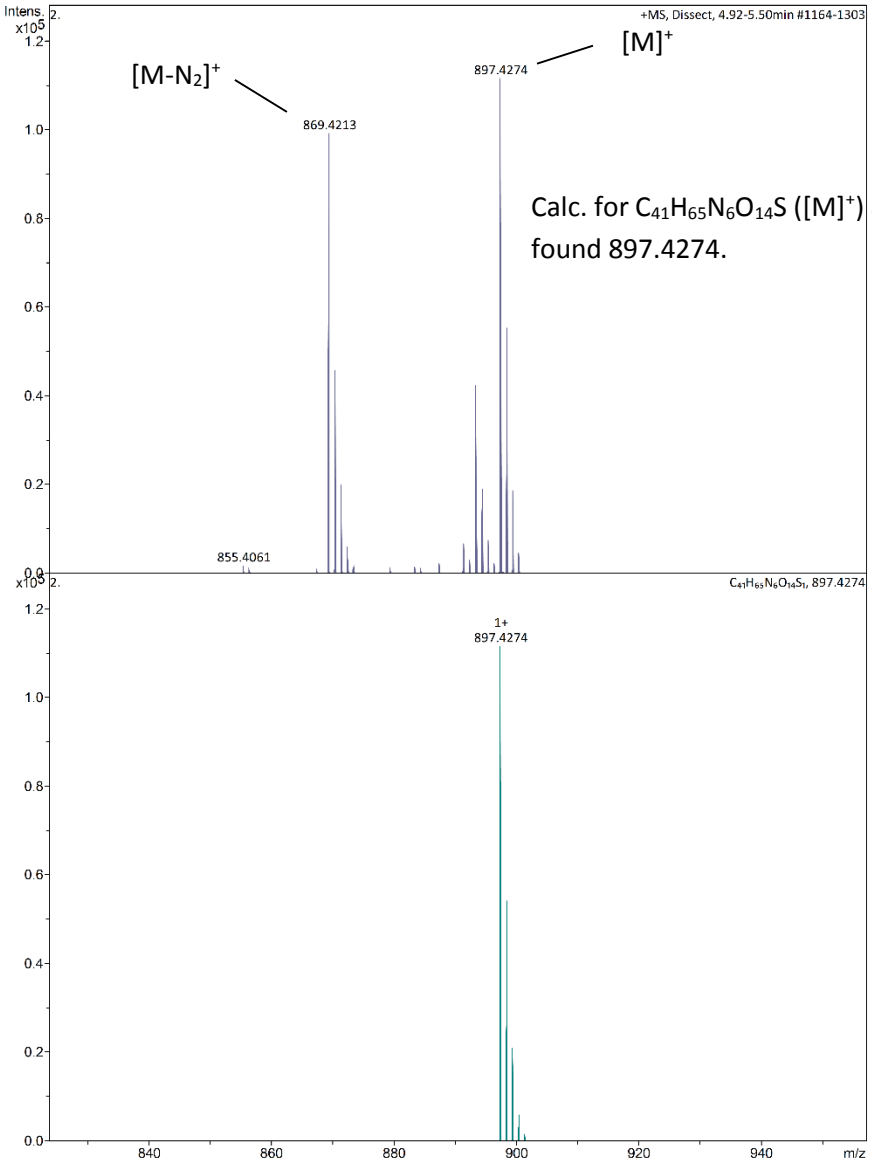
S10


Intens.
3.
+MS, Dissect, 6.79-7.02min #1609-1663
[M]+
799.4857
[M-N2]+
771.4796
753.4692
Calc. for C43H67N4O10 ([M]+) 799.4852,
found 799.4857.
C43H67N4O10, 799.4852
1+
799.4852
5000
4000
3000
2000
1000
0
750
760
770
780
790
800
810
820 m/z

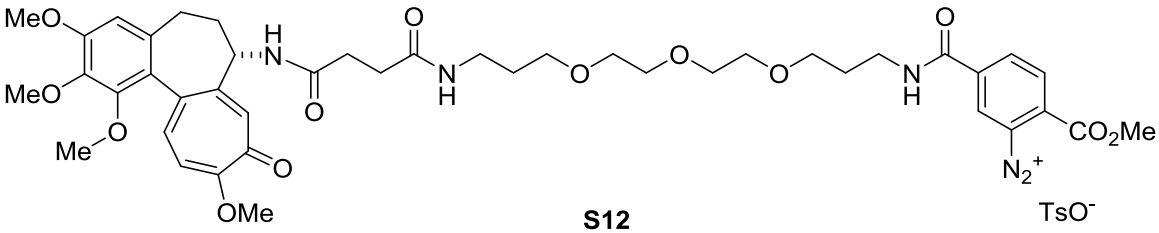


S11

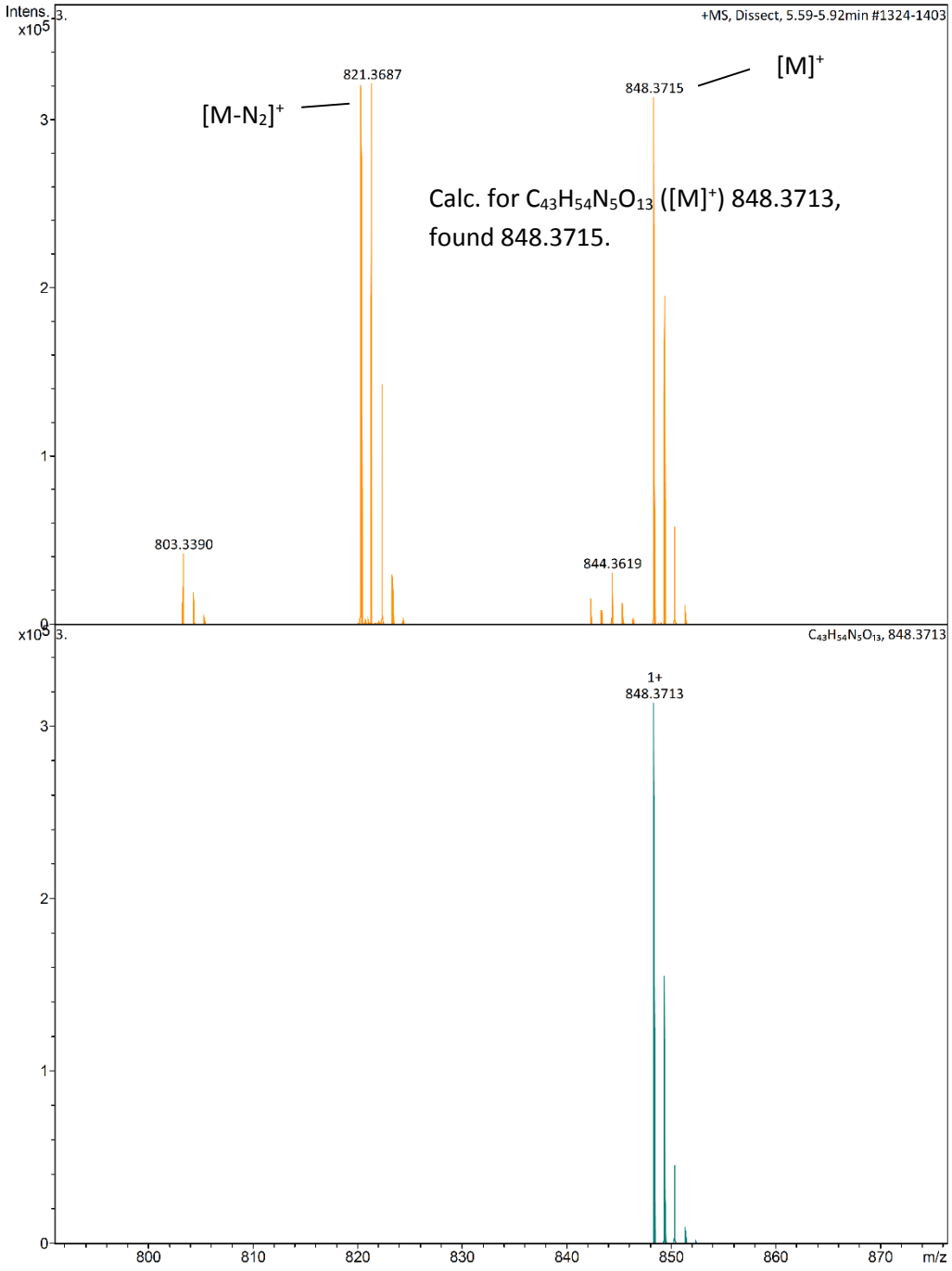
TsO^-



Calc. for $C_{41}H_{65}N_6O_{14}S$ ($[M]^+$) 897.4274, found 897.4274.


MeO
MeO
Me
O
OMe
H
N
O
O
N
H
O
O
O
O
N
H
O
CO2Me
N2+
TsO-

S12


Intens.
x10^5
+MS, Dissect, 5.59-5.92min #1324-1403
821.3687
848.3715
[M]+
[M-N2]+
Calc. for C43H54N5O13 ([M]+) 848.3713,
found 848.3715.
803.3390
844.3619
C43H54N5O13, 848.3713
1+
848.3713
800
810
820
830
840
850
860
870
m/z