Supporting Information for

Site-Specific Labeling of an Anti-MUC1 Antibody: Probing the Effects of Conjugation and Linker Chemistry on the Internalization Process

Hongjiao Xu[+,1], *Lu Gan*[+,1], *Ying Han*[1], *Yifan Da*[1], *Jiale Xiong*[1], *Sihua Hong*[1], *Qian Zhao*[1], *Nazi Song*[1], *Xiaoqing Cai*[*,1], *and Xianxing Jiang*

[1]School of Pharmaceutical Sciences, Sun Yat-sen University, 132 East Outer Ring Road, Guangzhou, 510006, China

[+]These two authors contributed equally to this study.

1. Preparation and characterization of SM3

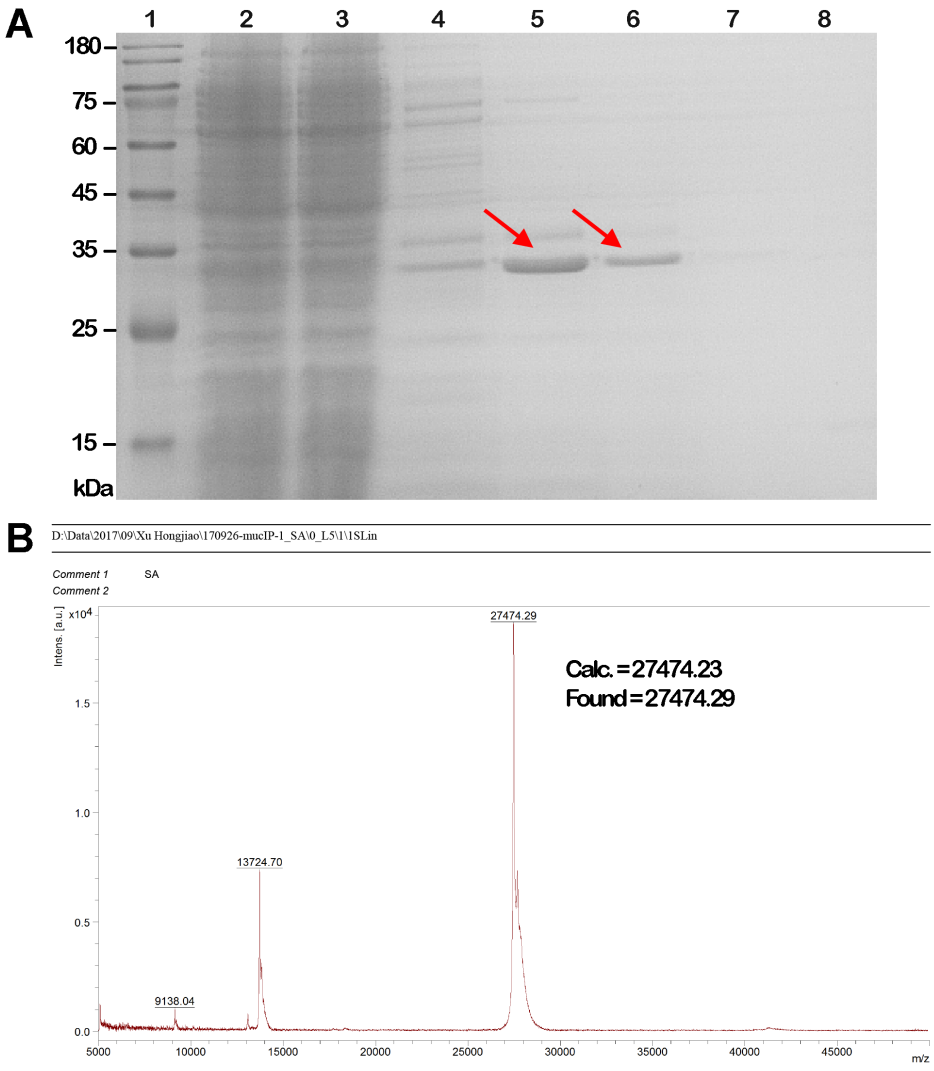


Figure S1. Bacterial expression and analysis of SM3. (A) Analysis on SDS-PAGE gel of the expressed SM3. Lane 1: marker; lane 2: lysate of *E.coli*; lane 3: flow through; lanes 4−8: elution with buffer contained 20 mM, 50 mM, 100 mM, 200 mM, 800 mM imidazole. Lane 5 and 6 were collected. (B) MALDI-TOF MS analysis of SM3.

2. Preparation of SortaseA and SM3-N_3

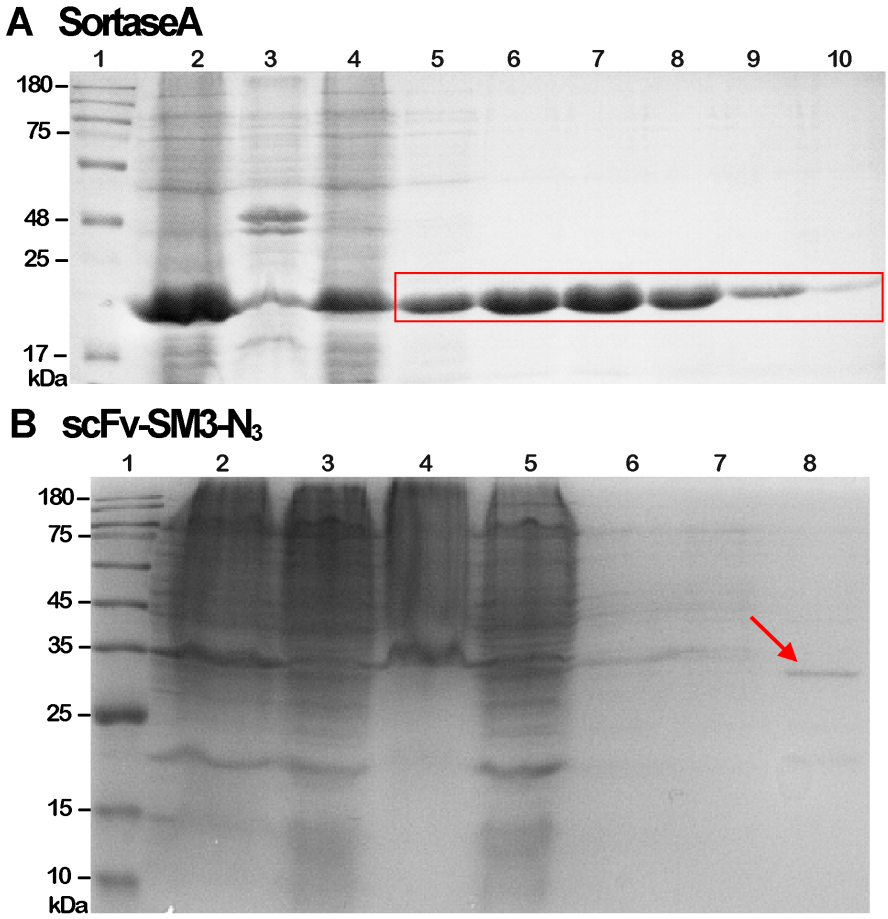


Figure S2. Bacterial expression and analysis of SortaseA and SM3-N_3. (A) Analysis on SDS-PAGE gel of the expressed SortaseA. Lane1: marker; lane 2: lysate of *E.coli*; lane 3: flow through; lanes 4−10: elution with buffer contained 20 mM, 50 mM, 80 mM, 100 mM, 200 mM, 400 mM and 800 mM imidazole. Lanes 5−10 were collected. (B) Analysis on SDS-PAGE gel of the expressed SM3-N_3. Lane1: marker; lane 2: lysate of *E.coli*; lane 3: supernatant; lane 4: sediment; lane 5: flow through; lanes 6−8: elution with buffer contained 20 mM, 50 mM and 200 mM imidazole. Lane 8 was collected.

3. Gel filtration chromatography of SM3 and SM3-N_3

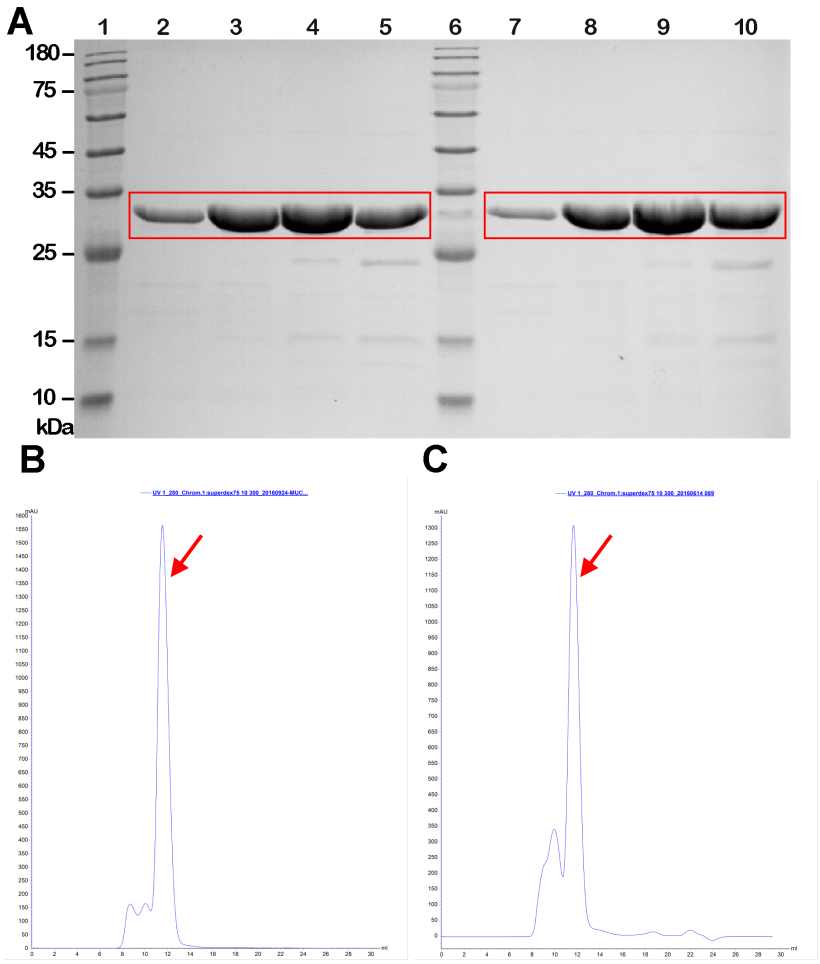
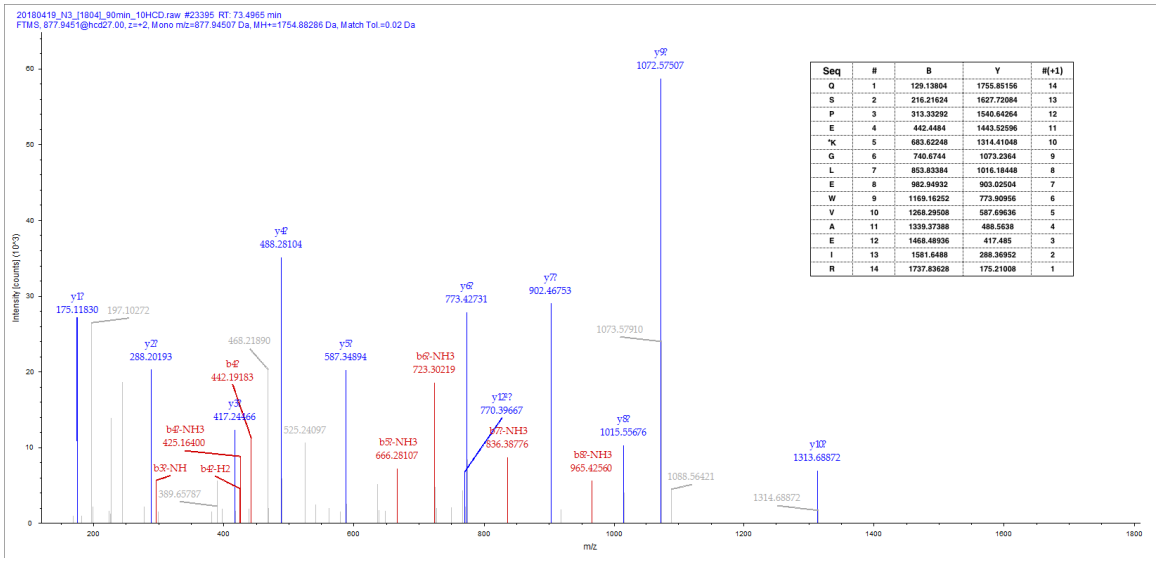


Figure S3. (A) SDS-PAGE analysis of the collected SM3 and SM3-N_3 purified by gel filtration chromatography. Lanes 1 and 6: marker; lanes 2−5: aliquots of SM3 peak fraction (10-13 ml of retention); lanes 7−10: aliquots of SM3-N_3 peak fraction (10-13 ml of retention). (B) Analytical gel filtration analysis of collected SM3. The retention volume is 11.5 ml. (C) Analytical gel filtration analysis of collected SM3-N_3. The retention volume is 11.7 ml.

4. Peptide mapping and MS/MS analysis of SM3-N_3



Seq	#	B	Y	#(+1)
Q	1	129.13804	1755.85156	14
S	2	216.21624	1627.72084	13
P	3	313.33292	1540.64264	12
E	4	442.4484	1443.52596	11
*K	5	683.62248	1314.41048	10
G	6	740.6744	1073.2364	9
L	7	853.83384	1016.18448	8
E	8	982.94932	903.02504	7
W	9	1169.16252	773.90956	6
V	10	1268.29508	587.69636	5
A	11	1339.37388	488.5638	4
E	12	1468.48936	417.485	3
I	13	1581.6488	288.36952	2
R	14	1737.83628	175.21008	1

Figure S4. Fragmentation (MS/MS) spectra of the QSPE*KGLEWVAEIR peptide of SM3-N_3 confirmed that the K43 modification (+113 Da) is present on Y10, which is indicated with *.

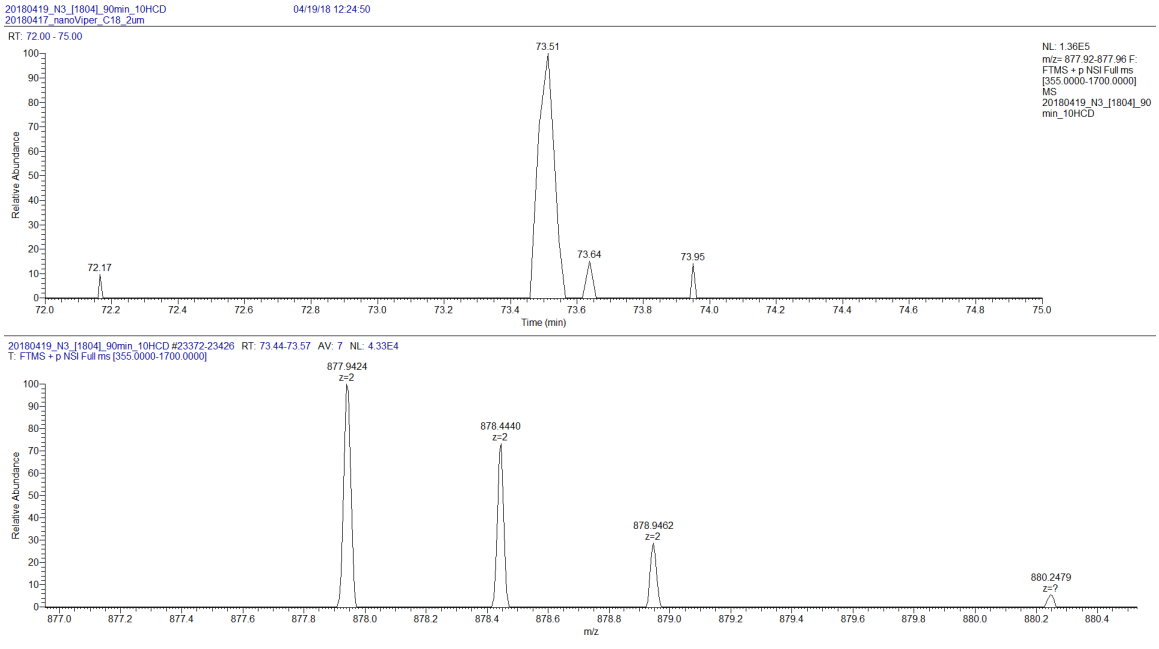


Figure S5. The UV profiles showed the peak time (73.51 min) and the protein mass spectrum of SM3-N_3.

5. In vitro characterization of SM3 and SM3-N_3

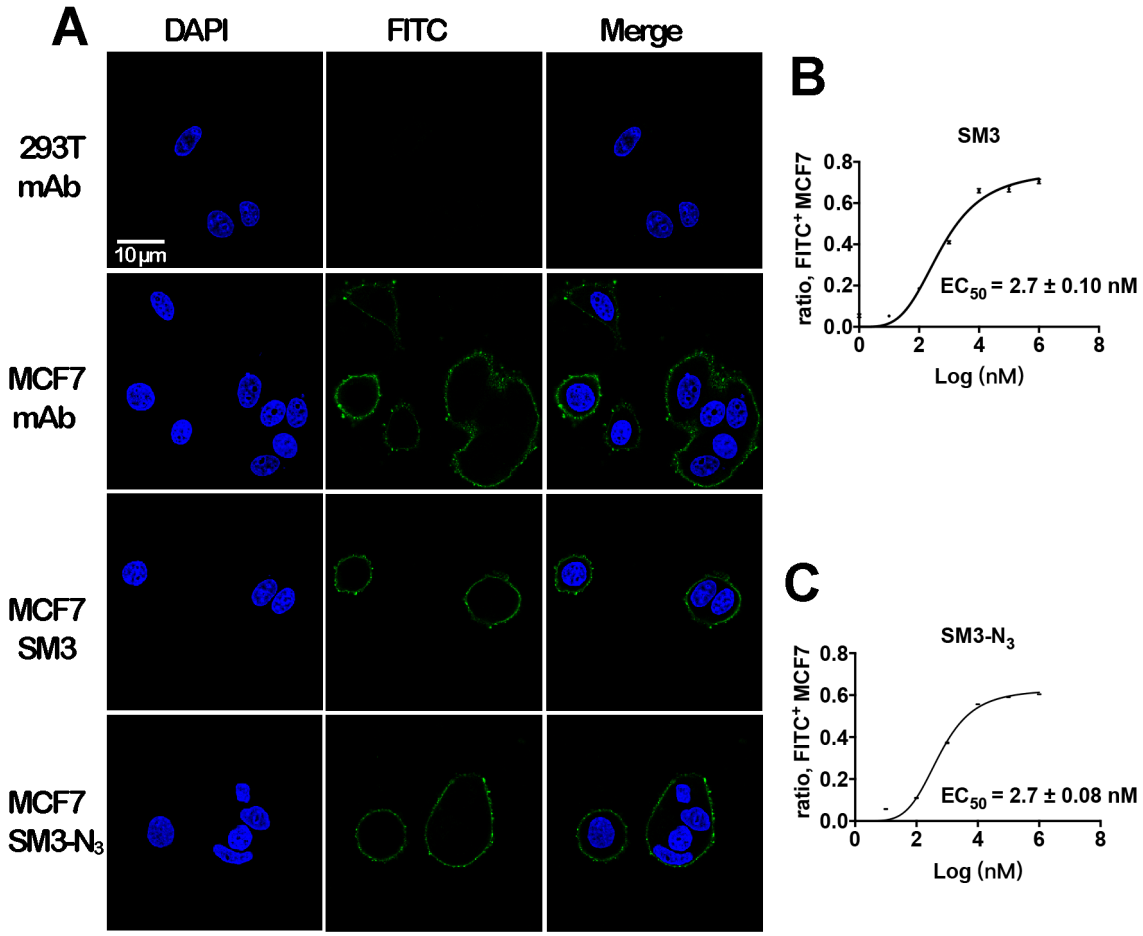


Figure S6. *In vitro* characterization of SM3 and SM3-N_3. (A) Confocal images of 293T and MCF7 cells incubated with mAb, SM3 or SM3-N_3, FITC-labeled goat anti-mouse IgG or FITC-labeled anti-His antibody was used as secondary antibody. Commercial full -length mAb was used as the positive control. 293T cell line was used as the negative control. (scale bar, 10 μm) (B, C) Determination of the half maximal effective binding concentration (K_d, EC_{50}) of SM3 and SM3-N_3 with human breast cancer cell line MCF7. The 293T cell line was used as the negative control. All assays were performed in triplicate (n = 3, error bar = means ± SEM).

6. Labeling of SM3 with FITC via random lysine conjugation

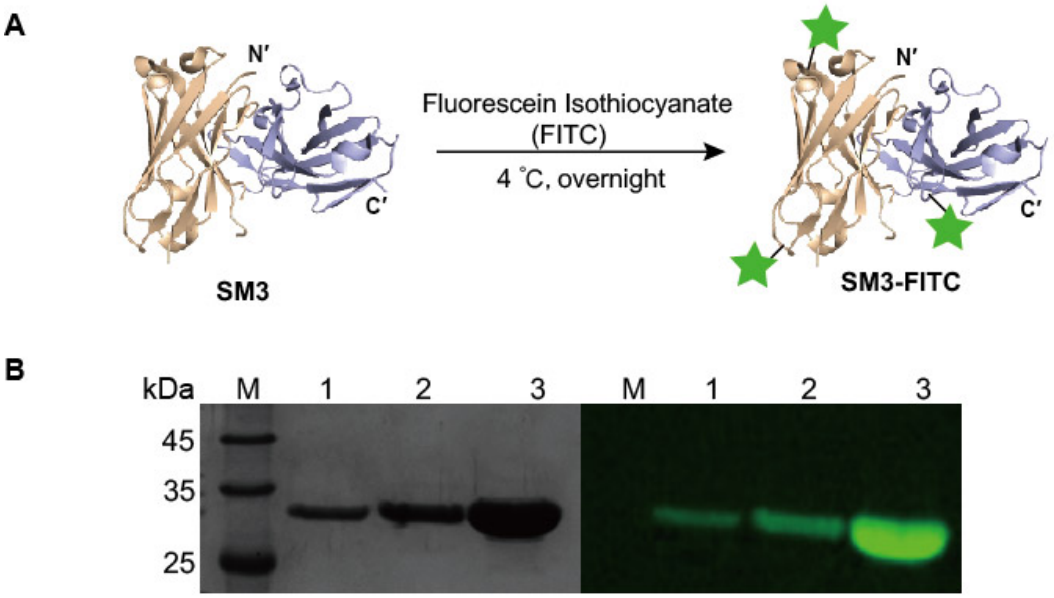


Figure S7. Labeling of SM3 with FITC via random lysine. (A) Preparation of SM3-FITC. Condition: Na_2CO_3, pH = 9.0, 4 ^{o}C, overnight. (B) SDS-PAGE gel indicating the formation of SM3-FITC. Left panel: Coomassie-stained gel. Right panel: in-gel fluorescence excited at 480 nm. Lanes 1−3: SM3-FITC with different concentrations.

7. In vivo and in vitro assay of MUC1 antigen distribution

Tumor bearing mice were sacrificed after the *in vivo* imaging assay. The sera were obtained when the tumors were removed. Sera were used for SDS-Page gel assay and western blot assay. SM3 was used as the primary antibody, and HRP-labeled anti-His antibody was used as the secondary antibody. The removed tumor was cut into 5 μm-slice after paraformaldehyde fixing and paraffin embedding. Hematoxylin staining was performed to identify the tumor cells. Immunohistochemistry staining was performed to confirm the existence of MUC1 antigen.

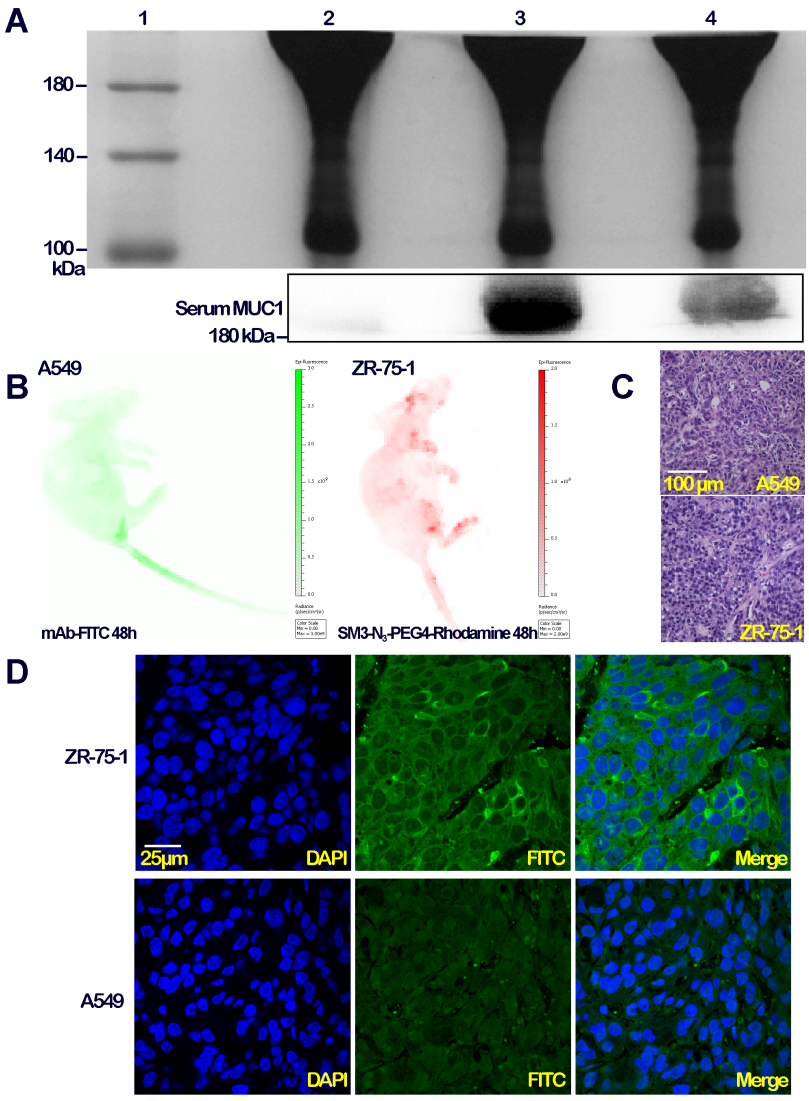
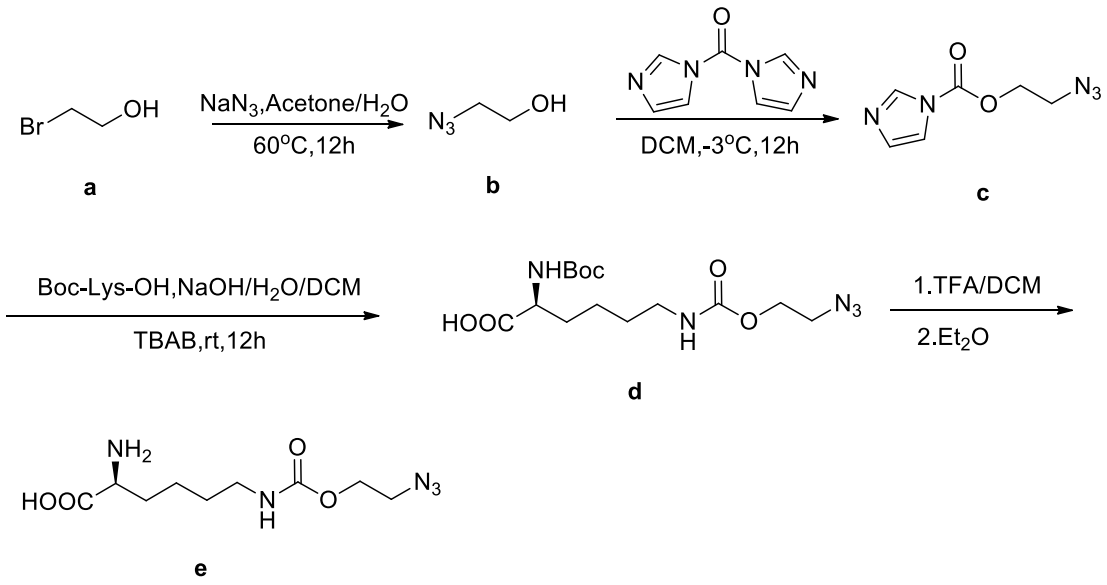


Figure S8. *In vivo* and *in vitro* assay of MUC1 antigen distribution. (A) Analysis on SDS-PAGE gel and western blot assay of the tumor mice serum. Top: lane1: marker; lane 2: normal nude mice serum; lane 3: ZR-75-1 tumor bearing mice serum; lane 4: A549 tumor bearing mice serum. Down: lane 2: normal nude mice serum; lane 3: ZR-75-1 tumor bearing mice serum; lane 4: A549 tumor bearing mice serum. (B) Left: representative IVIS imaging of nude mice bearing A549 tumor administrated with mAb-FITC at 48 h. λ_{ex} = 480 nm, λ_{em} = 520 nm. Right: representative IVIS imaging of nude mice bearing ZR-75-1 tumor administrated with SM3-N_3-PEG4-Rhodamine at 48 h. λ_{ex} = 560 nm, λ_{em} = 620 nm. (C) The staining of tumor tissue sections (scale bar, 100 μm). (D) Immunohistochemistry staining of tumor tissue sections (scale bar, 25 μm).

8. Synthesis of G_3K-Rhodamine, azidolysine 1, compound 2, compound 3 (Alkyne-Rhodamine) and compound 4 (Alkyne-PEG4-Rhodamine)

G_3K-Rhodamine Synthesis. The synthesis of G_3K-Rhodamine was performed based on the Fmoc strategy by solid-phase peptide synthesis as previously reported.[1] Briefly, the coupling was carried out by adding a solution of Fmoc-protected amino acid (3 equiv), HBTU (3 equiv), HOBt (3 equiv) and DIPEA (6 equiv) in DMF (5 ml) to the Rink Amide-MBHA resin (100-200 mesh, 0.341mmol/g). Fmoc-Lysine (Mtt) was coupled to the resin firstly, followed by the coupling of three Fmoc-Gly-OH. The Fmoc-protecting group of each amino acid was removed by exposing the resin to 80:20 DMF/piperidine for 15 min. After the completion of amino acid ligation, the 4-methyl trityl (Mtt)–protecting group was selectively removed with a solution of 97% DCM, 2% TIS and 1% TFA by shaking the resin for 30 min. A solution of Rhodamine (3 equiv), HBTU (3 equiv), HOBt (3 equiv) and DIPEA (6 equiv) in DMF (5 ml) added to the resin, and the resulted mixture was shaken for 2 h. Subsequently, the Fmoc-protecting group was removed and the cleavage of peptide from the resin was achieved by incubating the resin with 3 ml of cleavage cocktail (95% trifluoroacetic acid, 2.5% H_2O, 2.5% triisopropyl silane) for 2 h at room temperature. The peptide was precipitated in 30 ml of diethyl ether and purified using semi-preparative HPLC to yield a red trifluoroacetate. (column: Hypersil GOLD^{TM} 5 μm C18, 21.2 × 250 mm; solvent system: A = 0.1% TFA/H_2O, B = MeCN, flow rate 10 ml/min, 5%−40% B 0−40 min).



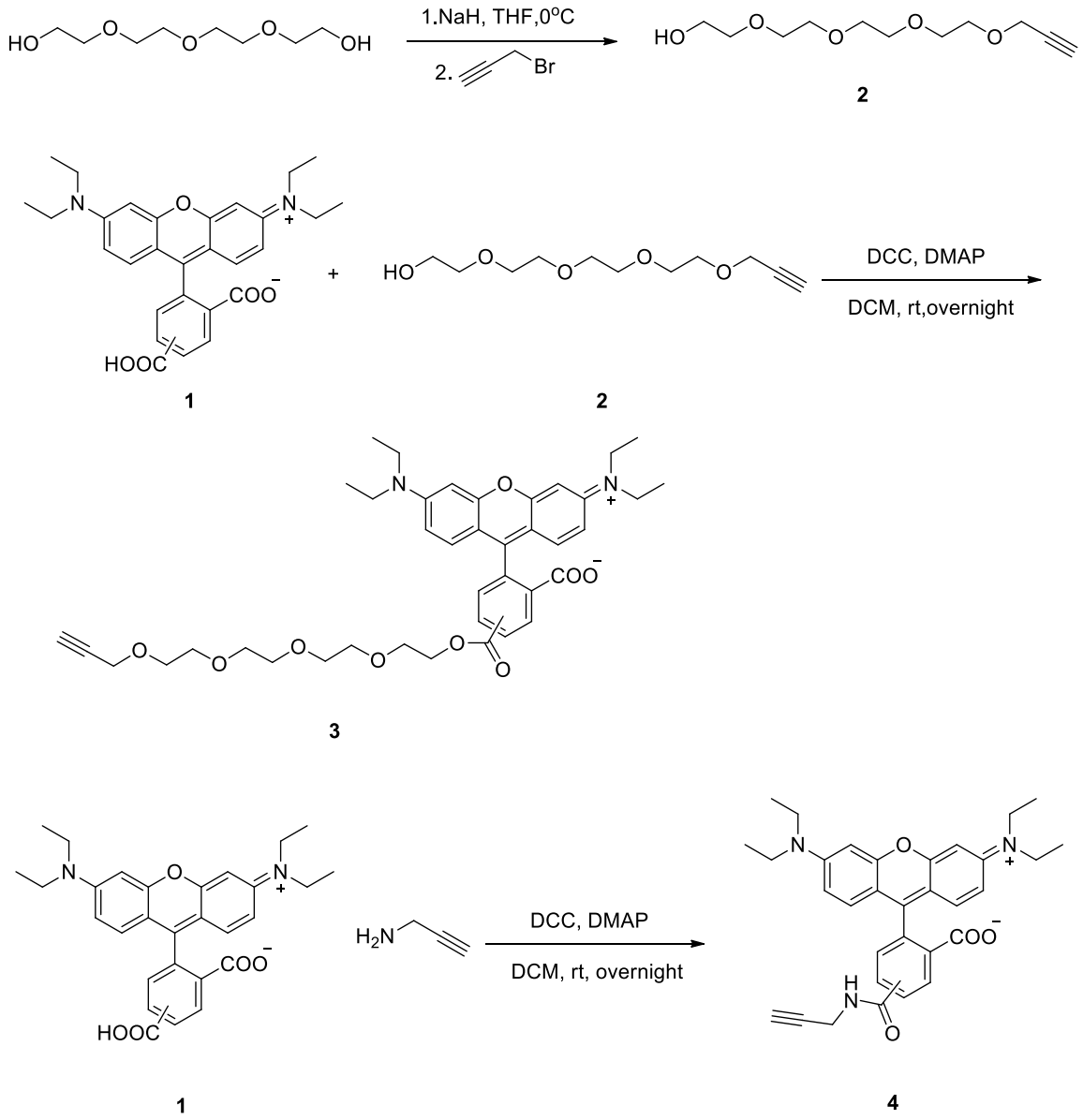
Synthesis of Compound b: 2-Bromoethanol (8 g, 64 mmol) and NaN_3 (6.24 g, 96 mmol) were added into acetone (60 ml) and water (30 ml). The reaction mixture was refluxed at 60 ^oC for 12 h. After completion of the reaction, acetone was removed under diminished pressure. The mixture was extracted with diethyl ether. The organic layer was separated, washed with brine twice and dried over Na_2SO_4, and concentrated under diminished pressure to obtain compound **b** (4.45 g, 80%). Compound **b** was used directly for the next step without further purified.

Synthesis of Compound c: Compound **b** (4.45 g, 51.2 mmol) was dissolved in DCM (60 ml), the resultant solution was slowly added at -3 ^oC into a suspension of N, N'-carbonyldiimidazole (12.44 g, 76.8 mmol) in DCM (30 ml). The reaction mixture was stirring at -3 ^oC for 12 h. Subsequently, the reaction mixture was quenched with water (100 ml), and the organic layer was washed with brine twice, dried over Na_2SO_4, and

concentrated under diminished pressure. The residue was further purified by silica gel chromatography to obtain compound **c** (8.34g, 90%).

Synthesis of Compound d: Compound **c** (8.34 g, 46 mmol) was dissolved in DCM (50 ml), the resultant solution was added into Boc-Lys-OH (9.4 g, 38 mmol) in 1M NaOH aqueous solution (30 ml), followed by the addition of TBAB (126 mg, 0.38 mmol). The reaction mixture solution was stirred at room temperature for 12 h, and then cooled to 0 ^{o}C, and treated with 1M HCl until pH 2−3. The aqueous layer was extracted with DCM, and the organic layer was washed with brine twice, dried with Na_2SO_4, and concentrated. The residue was purified by silica gel chromatography to obtain compound **d** as colorless oil (10.1 g, 74%).

Synthesis of Compound e: To a solution of compound **d** (10.1 g, 28 mmol) in DCM (20 ml) was slowly added trifluoroacetic acid (5 ml). The reaction solution was stirred at room temperature for 1h, and then concentrated under diminished pressure. The resultant residue was re-dissolved in MeOH, and precipitated in ethyl ether. The precipitate was collected and concentrated to obtain compound **e** as a white solid (5.2 g, 72%).



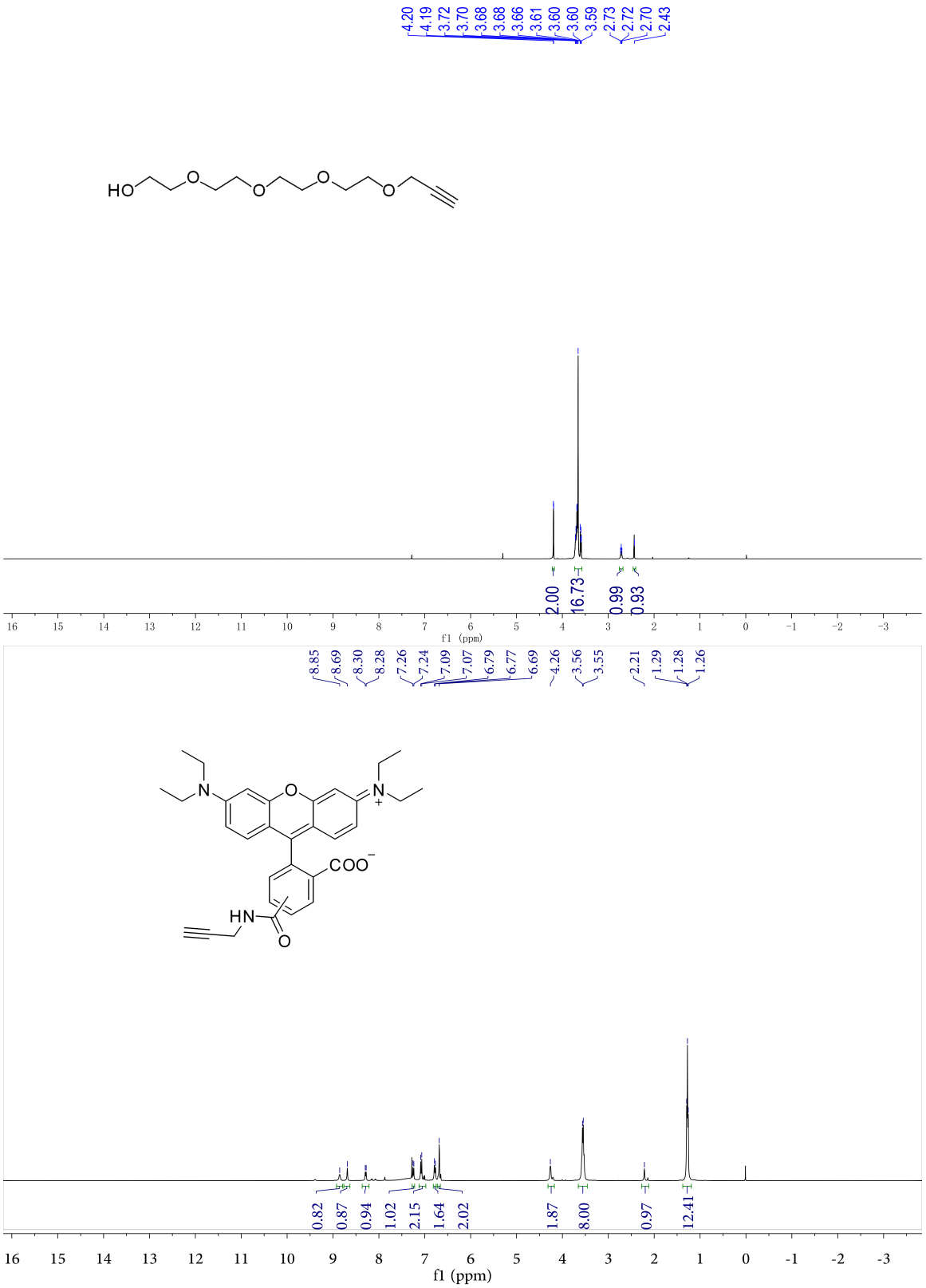
Synthesis of Compound 2: To a solution of tetraethylene glycol (3.38 g, 20 mmol) in THF (50 mL) was slowly added NaH (60% oil dispersion, 0.52 g, 13 mmol) at 0 ^{o}C. The reaction was stirred at 0 ^{o}C for 30 min. Propargyl bromide (1.1 mL, 10 mmol) was added to the reaction solution, and the reaction was stirred at room temperature for 12 h. Reaction was quenched by the addition of ice cold water, and the resulted mixture was extracted with dichloromethane twice. The combined organic layer was washed with

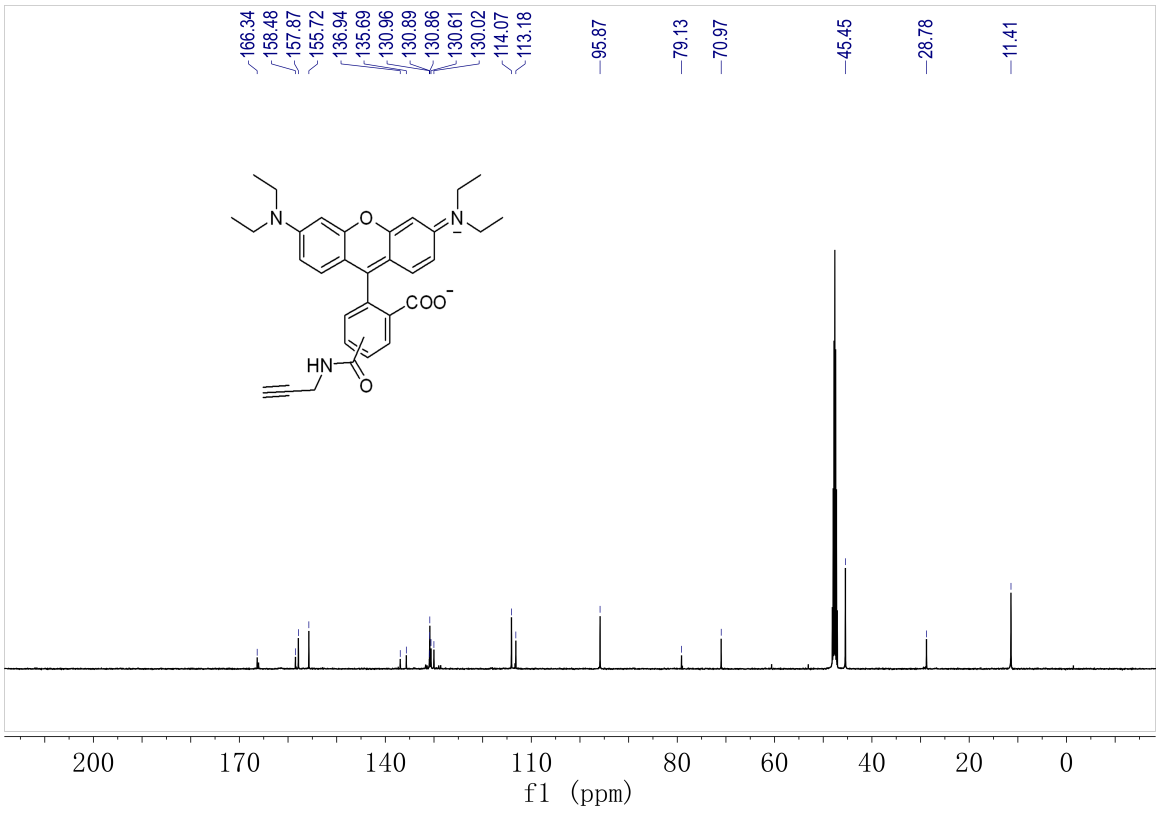
brine, and dried over anhydrous Na_2SO_4. After filtration, the solvent was removed under diminished pressure. The resulted residue was purified by silica gel chromatography using petroleum ether and ethyl acetate as eluent to afford compound **1** as a colorless oil (1.74 g, 75%). ^{1}H NMR data were confirmed by literature data.[2] ^{1}H NMR (400MHz, $CDCl_3$) δ 4.19 (d, *J*=2 Hz, 2H), 3.74-3.8 (m, 16H), 2.72 (t, *J*=6 Hz, 1H), 2.43 (t, *J*=2 Hz, 1H).

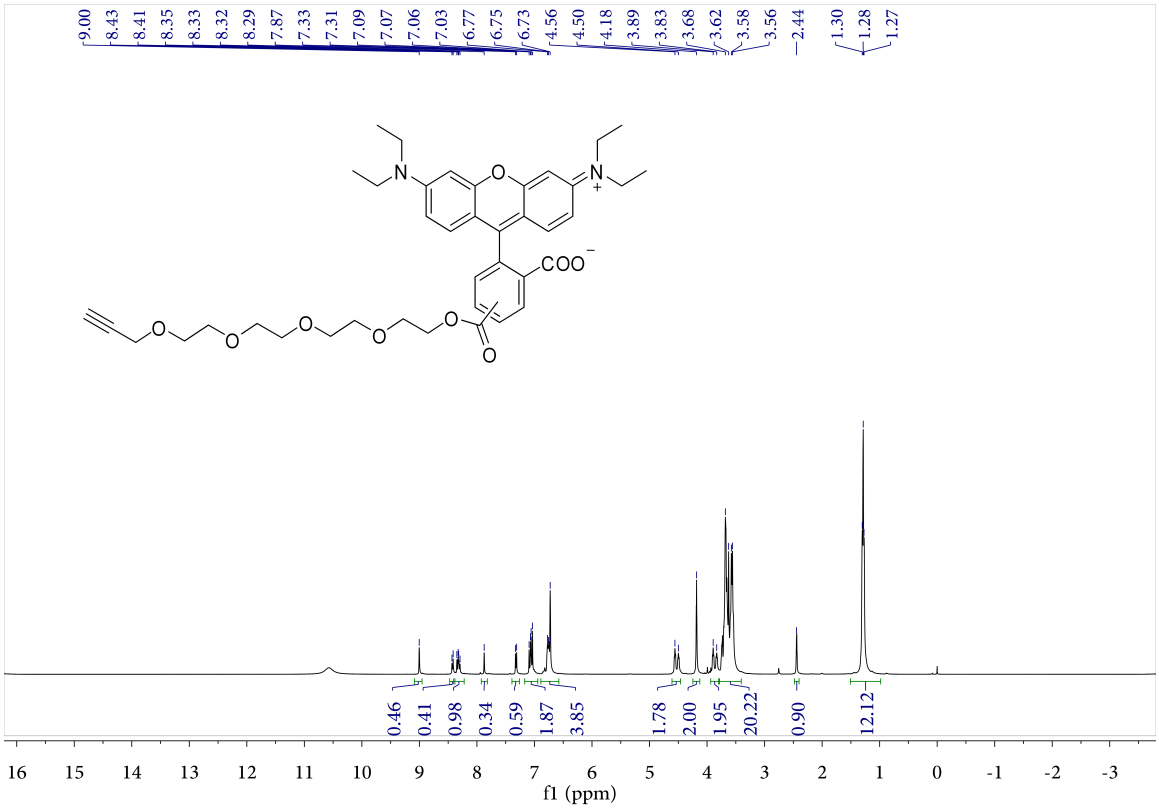
Synthesis of Compound 3: A solution of compound **1** (244 mg, 0.5 mmol), DMAP (6.1 mg, 0.05 mmol) and DCC (124 mg, 0.6 mmol) in anhydrous DCM (5 mL) was stirred at room temperature for 30 min, followed by the addition of compound **2** (143 mg, 0.55 mmol). The resulting mixture was stirred at room temperature overnight. The resulted mixture was then filtered, and the filtrate was concentrated under diminished pressure. The residue was purified by a semi-preparative HPLC column (C18 column, Hypersil GOLDTM, 5 μm, 21.2 × 250 mm) at a flow rate of 10.0 mL/min: solvent A, 0.1% TFA in H_2O; solvent B, MeCN. Gradient: 5% → 95% in 30 min. Pure product were collected and lyophilized to afford compound **3** as a dark purple solid (161 mg, 48%). ^{1}H NMR (400MHz, $CDCl_3$) δ 9.00 (s, 0.5H), 8.42 (d, *J*=8 Hz, 0.5H), 8.37-8.26 (m, 1H), 7.87 (s, 0.5H), 7.32 (d, *J*=8 Hz, 0.5H), 7.13-6.97 (m, 2H), 6.80-6.63 (m, 4H), 4.63-4.40 (td, *J*=24, 9.2 Hz, 2H), 3.97-3.79 (td, *J*=24, 9.2 Hz, 2H), 3.77-3.44 (m, 20H), 2.44 (s, 1H), 1.28 (t, *J*=6.8 Hz, 12H). ^{13}C NMR (100 MHz, CD_3OD) δ 164.98, 157.90, 157.85, 155.73, 138.30, 133.67, 132.93, 132.07, 131.91, 131.44, 130.90, 130.81, 114.14, 113.36, 113.08, 95.92, 79.24, 74.58, 70.28, 70.20, 70.15, 70.09, 69.95, 68.70, 64.64, 57.62, 45.47, 11.43; HRMS (ESI): calcd for $C_{40}H_{48}N_2O_9$ $[M+H]^+$ 701.3433, found 701.3435.

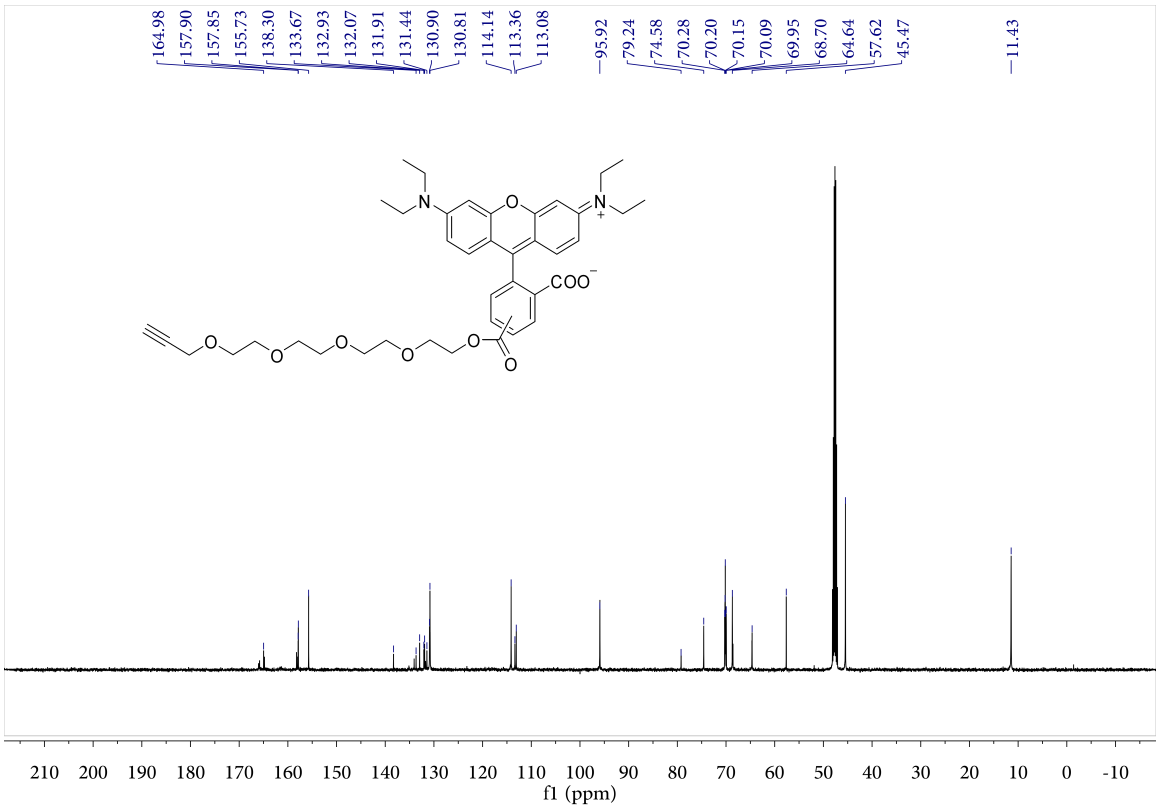
Synthesis of Compound 4: A solution of compound **1** (195 mg, 0.4 mmol), DMAP (4.9 mg, 0.04 mmol) and DCC (99 mg, 0.48 mmol) in anhydrous DCM (5 mL) was stirred at room temperature for 30 min, followed by the addition of 2-propynylamine (26.5 mg, 0.48 mmol). The resulting mixture was stirred at room temperature overnight. The resulted mixture was then filtered, and the filtrate was concentrated under diminished pressure. The residue was purified by a semi-preparative HPLC column (C18 column, Hypersil GOLDTM, 5 μm, 21.2 × 250 mm) at a flow rate of 10.0 mL/min: solvent A, 0.1% TFA in H_2O; solvent B, MeCN. Gradient: 5% → 95% in 30 min. Pure product were collected and lyophilized to afford compound **4** as a dark purple solid (84 mg, 40%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.85 (s, 1H), 8.69 (s, 1H), 8.29 (d, *J*=8 Hz, 1H), 7.25 (d, *J*=8 Hz, 1H), 7.12-6.98 (m, 2H), 6.84-6.79 (m, 2H), 6.72-6.63 (m, 2H), 4.26 (s, 2H), 3.55 (q, *J*=6.8 Hz, 8H), 2.21 (s, 1H), 1.27 (t, *J*=6.8 Hz, 12H); ^{13}C NMR (100 MHz, CD_3OD) δ 166.34, 158.48, 157.87, 155.72, 135.69, 130.89, 130.86, 130.61, 130.02, 114.07, 113.18, 95.87, 79.13, 70.79, 45.45, 28.78, 11.41; HRMS (ESI): calcd for $C_{32}H_{33}N_3O_4$ $[M+H]^+$ 524.2544, found 524.2546.

9. ^{1}H and ^{13}C NMR Spectra of compound 2, 3 and 4.




166.34
158.48
157.87
155.72
136.94
135.69
130.96
130.89
130.86
130.61
130.02
114.07
113.18
95.87
79.13
70.97
45.45
28.78
11.41
COO-
HN
O
200
170
140
110
80
60
40
20
0
f1 (ppm)


9.00
8.43
8.41
8.35
8.33
8.32
8.29
7.87
7.33
7.31
7.09
7.07
7.06
7.03
6.77
6.75
6.73
4.56
4.50
4.18
3.89
3.83
3.68
3.62
3.58
3.56
2.44
1.30
1.28
1.27
COO-
0.46
0.41
0.98
0.34
0.59
1.87
3.85
1.78
2.00
1.95
20.22
0.90
12.12
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
f1 (ppm)



References

1. Guimaraes, C. P., Witte, M. D., Theile, C. S., Bozkurt, G., Kundrat, L., Blom, A. E. M., Ploegh, H. L., *Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. Nat. Protoc.*, 2013. 8(9), 1787–1799.
2. Huang, J., Wu, F., Yu, Y., Huang, H., Zhang, S., and You, J., *Lipoic acid based core cross-linked micelles for multivalent platforms: design, synthesis and application in bio-imaging and drug delivery. Org. Biomol. Chem., 2017,* 15(22), 4798–802.