

## Supporting Information

### Fully synthetic self-adjuvanting MUC1-fibroblast stimulating lipopeptide 1 conjugates as potential cancer vaccines

Yonghui Liu, Wenpeng Zhang, Qianqian He, Fan Yu, Tianbang Song, Tingting Liu,  
Zhenqing Zhang, Jun Zhou, Peng George Wang\* and Wei Zhao\*

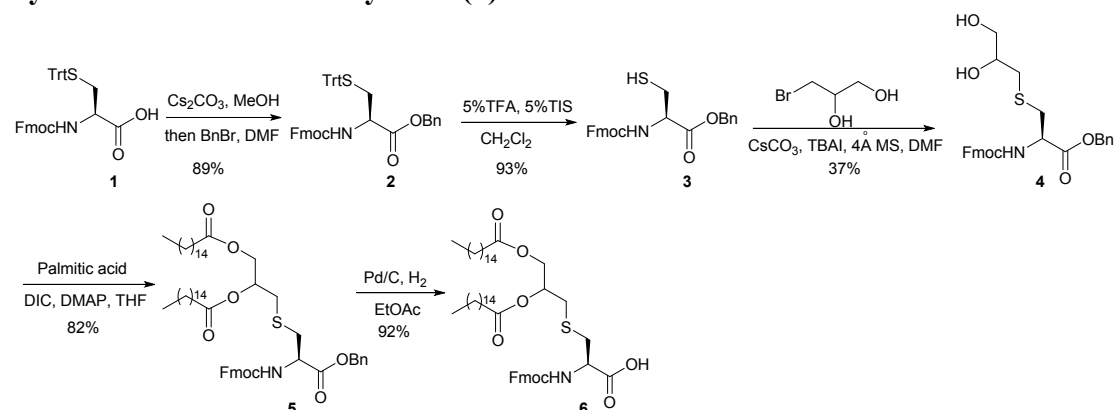
### Table of Contents

General procedures.....	S2
Synthesis of Fmoc-Pam <sub>2</sub> Cys-OH (6).....	S2
Analytical data for MUC1- lipopeptide.....	S6
Analytical data for MUC1 peptide and glycopeptide.....	S12
Immunological studies.....	S16
References.....	S18
<sup>1</sup> H NMR, <sup>13</sup> C NMR spectra.....	S18

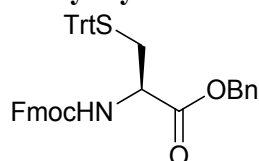
## General procedures

All reagents were purchased from commercial sources and were used without further purification. All solvents were available with commercially dehydrated or freshly dehydrated and distilled prior to use. Reactions were monitored using Thin Layer Chromatography (TLC) on silica gel GF254 plates with detection by short wave UV fluorescence ( $\lambda = 254$  nm) after being stained with p-anisaldehyde solution (ethanol/p-anisaldehyde/acetic acid/sulfuric acid 135:5:4:1.5) or 10% phosphomolybdic acid in EtOH, followed by heating on a hot plate. Column chromatography was conducted by silica gel (200-300 mesh) with petroleum ether and ethyl acetate as eluent.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were obtained using a Bruker AV 400 MHz spectrometer at 400 MHz and 100 MHz respectively. Signals are reported in terms of their chemical shift ( $\delta$  in ppm) relative to  $\text{CDCl}_3$  ( $^1\text{H}$ , 7.26 and  $^{13}\text{C}$ , 77.16). Coupling constants are reported in hertz. High-resolution electrospray-ionization mass spectra (HRESIMS) were obtained on a Varian QFT-ESI mass spectrometer. Matrix-assisted laser desorption/ionization time of flight mass spectra (MALDI-TOF MS) were performed using 2, 5-dihydroxybenzoic acid (DHB) or  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) as matrix on Varian 7.0T FTMS instrument. Reversed-phase HPLC separations were performed on a Waters system 2487 using solution A (0.1% trifluoroacetic acid in 100% acetonitrile) and solution B (0.1% trifluoroacetic acid in 100% water) for elution. UV absorption signals were detected with an UV detector at a wavelength of 220nm. Semi-preparative HPLC was used for separation and purification of the glycopeptides on a C-18 column (10 $\times$ 250 mm) at a flow rate of 2 mL/min.

## Synthesis of Fmoc-Pam<sub>2</sub>Cys-OH (6):



## N-Fluorenylmethoxycarbonyl-S-trityl-cysteine benzyl ester (2)



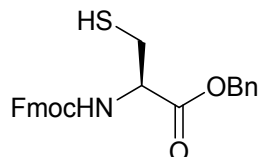
Fmoc-L-Cysteine (Trt)-OH **1** (10.00 g, 17.09 mmol) was dissolved in CH<sub>3</sub>OH (170 mL) and Cesium carbonate (3.42 g, 10.49 mmol) was added. After the solid powder was dissolved, the solution was concentrated. The complex was then dissolved in DMF (180 mL) and benzyl bromide (2.23 mL, 3.21 g, 18.79 mmol) was added. The solution was stirred for 24h at r.t. and then the solvent was evaporated. The crude mixture was poured into H<sub>2</sub>O (100 mL), extracted with EtOAc (3 × 500 mL) and washed with brine (100 mL). After drying with Na<sub>2</sub>SO<sub>4</sub> the solvent was evaporated, the product **2** (10.27 g, 89%) was put into next step directly without any purification. R<sub>f</sub> = 0.65 (PE/EtOAc = 3:1).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.78 – 7.73 (m, 2H), 7.59 (m, 2H), 7.37 – 7.33 (m, 9H), 7.32 – 7.28 (m, 6H), 7.24 – 7.16 (m, 9H), 5.30 (d, *J* = 8.3 Hz, 1H, NH), 5.14 (m, 2H, Ph-CH<sub>2</sub>), 4.35 (m, 3H, Fmoc-CH<sub>2</sub>, Cys-CH<sub>α</sub>), 4.20 (t, *J* = 7.0 Hz, 1H, Fmoc-CH), 2.69 (dd, *J* = 12.4, 6.4 Hz, 1H, Cys-CH<sub>2</sub><sub>a</sub>), 2.60 (dd, *J* = 12.4, 4.4 Hz, 1H, Cys-CH<sub>2</sub><sub>b</sub>).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 170.50, 155.71, 144.34, 144.00, 143.84, 141.41, 141.39, 135.23, 129.61, 128.67, 128.54, 128.37, 128.13, 127.83, 127.21, 127.08, 127.01, 125.28, 125.23, 120.09, 67.54, 67.29, 67.13, 53.15, 47.21, 34.23.

HRMS (ESI) calcd for [C<sub>44</sub>H<sub>37</sub>NO<sub>4</sub>S + NH<sub>4</sub>]<sup>+</sup> 693.2787, found 693.2790.

### ***N*-Fluorenylmethoxycarbonyl-cysteine benzyl ester (**3**)**



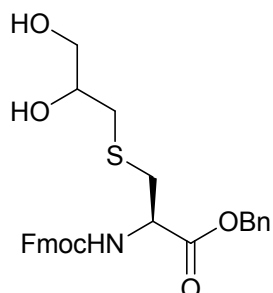
Compound **2** (10.2 g, 15.10 mmol) was dissolved in degassed, anhydrous CH<sub>2</sub>Cl<sub>2</sub> (180 mL). TIS (10 mL) and TFA (10 mL) were added to the solution slowly. The resulting solution was allowed to react under N<sub>2</sub> for 4.0 hours at r.t. and was monitored by TLC. The reaction was diluted in CH<sub>2</sub>Cl<sub>2</sub> (500 mL), washed with H<sub>2</sub>O (200 mL), saturated NaHCO<sub>3</sub> (200 mL), brine (200 mL), dried with Na<sub>2</sub>SO<sub>4</sub> and concentrated in vacuo. Flash column chromatography over silica gel, white solid **3** (6.08 g, 93%) was isolated. R<sub>f</sub> = 0.43 (PE/EtOAc = 3:1).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.78 (d, *J* = 7.5 Hz, 2H), 7.61 (d, *J* = 7.4 Hz, 2H), 7.44 – 7.30 (m, 9H), 5.74 (d, *J* = 7.4 Hz, 1H, NH), 5.23 (m, 2H, Ph-CH<sub>2</sub>), 4.72 (m, 1H, Cys-CH<sub>α</sub>), 4.43 (m, 2H, Fmoc-CH<sub>2</sub>), 4.23 (t, *J* = 6.9 Hz, 1H, Fmoc-CH), 3.10 – 2.93 (m, 2H, Cys-CH<sub>2</sub>).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 169.98, 155.76, 143.91, 143.72, 141.45, 141.42, 135.07, 128.84, 128.63, 127.89, 127.22, 125.22, 125.17, 120.16, 120.14, 67.84, 67.26, 55.33, 47.25, 27.32.

HRMS (ESI) calcd for [C<sub>25</sub>H<sub>23</sub>NO<sub>4</sub>S + H]<sup>+</sup> 434.1426, found 434.1423.

***N*-Fluorenylmethoxycarbonyl-*S*-[2,3-dihydroxy-(2*RS*)-propyl]-(*R*)-cysteine benzyl ester (4)**



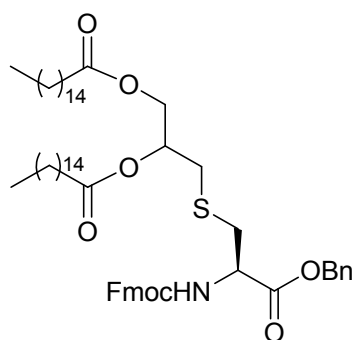
To a solution of *N*-Fluorenylmethoxycarbonyl-cysteine benzyl ester **3** (2.40 g, 5.54 mmol) in anhydrous DMF (27 mL) under N<sub>2</sub> atmosphere were added powdered, activated 4Å MS. Then cesium carbonate (1.81 g, 5.54 mmol) and tetrabutylammonium iodide (2.04 g, 5.54 mmol) were added in an ice-water bath, following by adding 3-Bromo-1,2-propanediol (1.29 g, 8.31 mmol) dropwise. The reaction solution was then warmed to rt and stirred for 3 hours. The mixture was filtered through celatom and diluted with CH<sub>2</sub>Cl<sub>2</sub> (120 mL). The solution were washed washed with H<sub>2</sub>O (30 mL), brine (30 mL), dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, the solvent was removed under vacuum. Flash column chromatography over silica gel, white solid **4** (1.03 g, 37%) was isolated. R<sub>f</sub>=0.2 (Toluene/EtOAc = 1:1).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) two diastereoisomers δ 7.76 (d, *J* = 6.5 Hz, 2H), 7.60 (d, *J* = 5.3 Hz, 2H), 7.40 – 7.31 (m, 9H), 5.93 (dd, *J* = 16.2, 8.1 Hz, 1H, NH), 5.26 – 5.14 (m, 2H, PhCH<sub>2</sub>), 4.68 (d, *J* = 4.1 Hz, 1H, Cys-CH<sub>α</sub>), 4.41 (m, 2H, Fmoc-CH<sub>2</sub>), 4.23 (m, 1H, Fmoc-CH), 3.92 (d, *J* = 4.0 Hz, 1H, *S*-glycery-CH), 3.75 (m, 1H, *S*-glycery-OCH<sub>2</sub>), 3.70 – 3.62 (m, 1H, *S*-glycery-OCH<sub>2</sub>), 3.08 – 2.93 (m, 2H, Cys-CH<sub>2</sub>), 2.75 – 2.53 (m, 2H, *S*-glycery-CH<sub>2</sub>).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 170.74, 156.23, 143.69, 141.34, 134.94, 128.76, 128.55, 127.85, 127.20, 125.17, 120.10, 71.52, 70.99, 70.76, 67.81, 67.37, 65.27, 65.13, 64.30, 54.17, 47.09, 36.42, 36.30, 35.46, 34.84.

HRMS (ESI) calcd for [C<sub>28</sub>H<sub>29</sub>NO<sub>6</sub>S + H]<sup>+</sup> 508.1794, found 508.1790.

***N*-Fluorenylmethoxycarbonyl-*S*-[2,3-bis(palmitoyloxy)-(2*RS*)-propyl]-(*R*)-cysteine benzyl ester (5)**



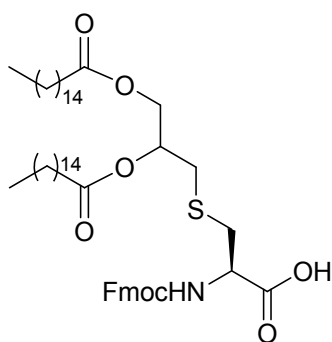
*N*-Fluorenylmethoxycarbonyl-*S*-[2,3-dihydroxy-(2*RS*)-propyl]-(*R*)-cysteine benzyl ester **4** (1.00 g, 1.97 mmol) was dissolved in THF (20 mL), then palmitic acid (1.52 g, 5.91 mmol) was added, following by adding *N,N*-dimethylaminopyridine (96 mg, 0.79 mmol) and *N,N'*-diisopropylcarbodiimide (1.09 mL, 7.09 mmol) under argon atmosphere. The solution was stirred overnight at rt. The solvent was removed under vacuum. Flash column chromatography over silica gel, white solid **4** (1.58 g, 82%) was isolated. *R<sub>f</sub>* = 0.66 (PE/EtOAc 3:1).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) two diastereoisomers δ 7.77 (d, *J* = 7.6 Hz, 2H), 7.61 (d, *J* = 7.2 Hz, 2H), 7.43 – 7.28 (m, 9H), 5.76 (m, 1H, NH), 5.25 – 5.16 (m, 2H, PhCH<sub>2</sub>), 5.12 (m, 1H, *S*-glycery-CH), 4.69 (m, 1H, Cys-CH<sub>α</sub>), 4.39 – 4.19 (m, 4H, *S*-glycery-OCH<sub>2</sub>a, Fmoc-CH<sub>2</sub>, Fmoc-CH), 4.07 (m, 1H, *S*-glycery-OCH<sub>2</sub>b), 3.16 – 2.95 (m, 2H, Cys-CH<sub>2</sub>), 2.69 (d, *J* = 6.0 Hz, 2H, *S*-glycery-CH<sub>2</sub>), 2.36 – 2.27 (m, 4H, Pal-CH<sub>2</sub>), 1.64 – 1.58 (m, 4H, Pal-CH<sub>2</sub>), 1.26 (m, 48H, Pal-CH<sub>2</sub>), 0.88 (t, *J* = 6.4 Hz, 6H, Pal-CH<sub>3</sub>).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 179.03, 173.47, 173.22, 173.17, 170.54, 155.88, 155.86, 143.91, 143.85, 141.41, 135.05, 128.80, 128.76, 128.59, 127.86, 127.21, 125.28, 120.12, 70.30, 70.27, 69.43, 67.81, 67.47, 65.79, 63.55, 54.05, 53.96, 47.20, 35.19, 35.02, 34.37, 34.22, 34.19, 34.02, 33.18, 33.13, 32.06, 29.84, 29.81, 29.73, 29.64, 29.58, 29.50, 29.43, 29.38, 29.27, 29.24, 29.20, 25.02, 25.00, 24.84, 22.88, 22.83, 14.27.

MALDI MS calcd for [C<sub>60</sub>H<sub>89</sub>NO<sub>6</sub>S + Na]<sup>+</sup> 1006.620, found 1006.039.

*N*-Fluorenylmethoxycarbonyl-*S*-[2,3-*bis*(palmitoyloxy)-(2*RS*)-propyl]-(*R*)-cysteine (**6**)



*N*-Fluorenylmethoxycarbonyl-*S*-[2,3-*bis*(palmitoyloxy)-(2*RS*)-propyl]-(*R*)-cysteine benzyl ester **5** (1.50 g, 1.52 mmol) was dissolved in EtOAc (30 mL), then Pd/C (250 mg) was added. The reaction atmosphere was replaced with argon three times, then substituted with hydrogen three times again. The reaction was stirred at hydrogen atmosphere for 30 minutes. Upon completion, the reaction was filtered and concentrated. Compound **6** (1.25 g, 92%) was obtained through column chromatography as white powder.  $R_f = 0.28$  (DCM/MeOH = 10:1).

$^1\text{H NMR}$  (400 MHz,  $\text{CDCl}_3$ ) two diastereoisomers  $\delta$  7.76 (d,  $J = 7.5$  Hz, 2H), 7.61 (d,  $J = 7.0$  Hz, 2H), 7.40 (t,  $J = 7.4$  Hz, 2H), 7.31 (t,  $J = 7.4$  Hz, 2H), 5.76 (m, 1H, NH), 5.21 – 5.10 (m, 1H, *S*-glycery-CH), 4.66 (m, 1H, Cys-CH $\alpha$ ), 4.44 – 4.21 (m, 4H, *S*-glycery-OCH $_{2a}$ , Fmoc-CH $_2$ , Fmoc-CH), 4.14 (m, 1H, *S*-glycery-OCH $_{2b}$ ), 3.22 – 3.02 (m, 2H, Cys-CH $_2$ ), 2.84 – 2.69 (m, 2H, *S*-glycery-CH $_2$ ), 2.38 – 2.26 (m, 4H, Pal-CH $_2$ ), 1.59 (m, 4H, Pal-CH $_2$ ), 1.25 (br s, 48H, Pal-CH $_2$ ), 0.88 (t,  $J = 6.7$  Hz, 6H, Pal-CH $_3$ ).

HRMS (ESI) calcd for  $[\text{C}_{53}\text{H}_{83}\text{NO}_8\text{S} + \text{NH}_4]^+$  911.6183, found 911.6192.

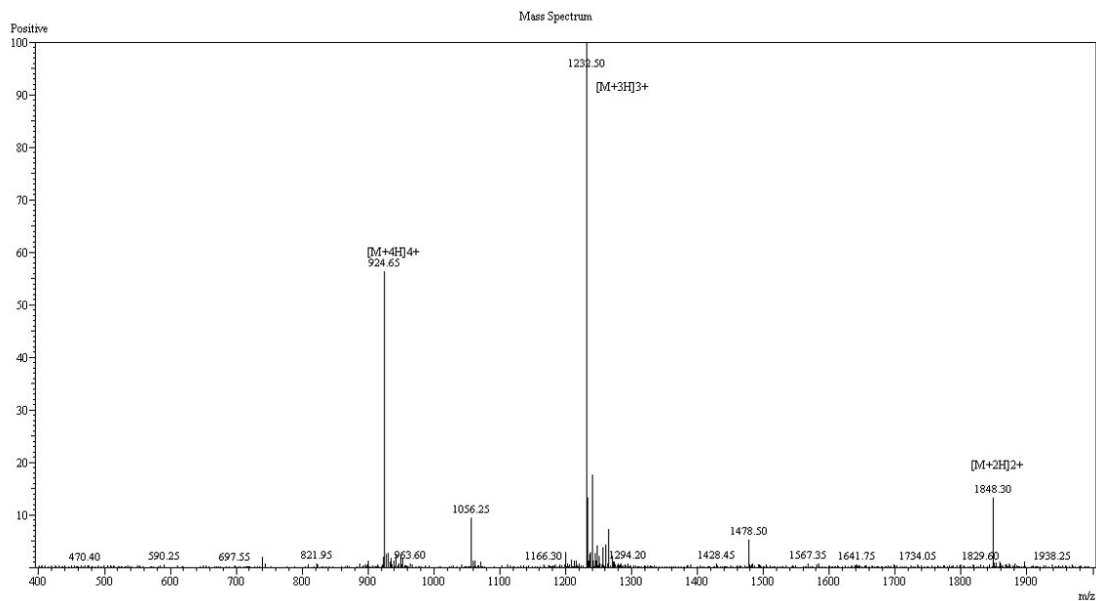
These data are consistent with the literature reported by Boons group.<sup>1</sup>

### General Procedure for peptides synthesis

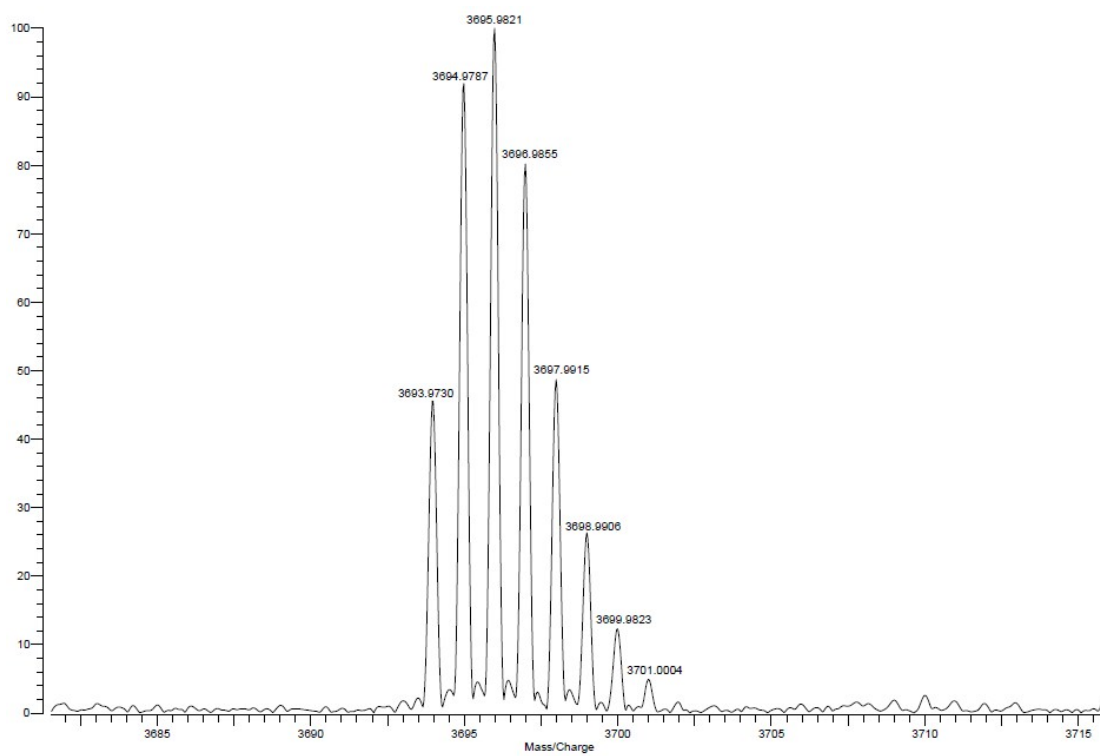
(Glycolipo)peptides were synthesized according to standard Solid Phase Peptide Synthesis (SPPS) procedure. The peptide synthesis was performed with natural Fmoc amino acids utilizing 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU, 4.0 equiv), 1-hydroxybenzotriazole (HOBT, 4.0 equiv), *N,N*-diisopropylethylamine (DIPEA, 8.0 equiv). Fmoc O-glycosylated amino acid and Fmoc-Pam $_2$ Cys-OH were introduced using more reactive 1-[*dis*(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate (HATU, 2.0 equiv), 1-hydroxy-7-azabenzotriazole (HOAT, 2.0 equiv), DIPEA (4.0 equiv). The acetyl moieties of the Tn antigen were removed by adding MeONa/MeOH (PH=10~11). All acid-sensitive side chain protecting groups were removed and the (glycolipo)peptides were detached from the resin by adding 90% TFA, 5% TIPS, 5% H $_2$ O. The (glycolipo)peptides were purified by semi-preparative HPLC on a C18 column.

### Analytical data for MUC1-(glycol)lipopeptide



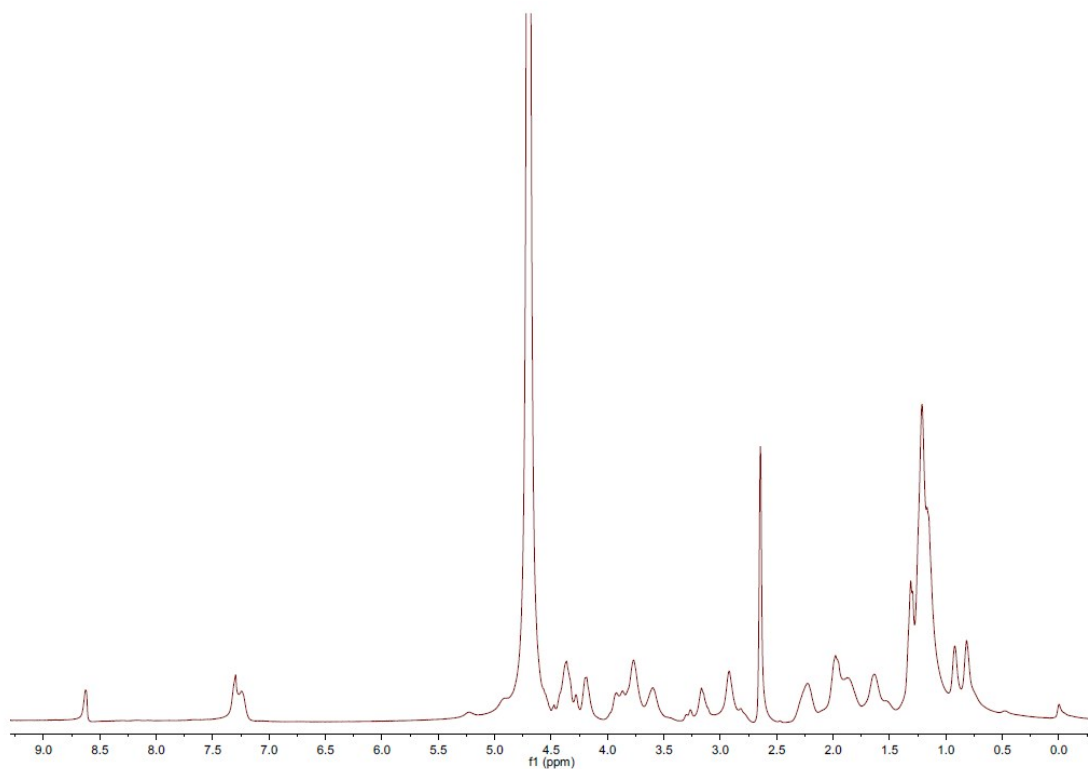


ESI-MS:  $m/z$  for  $C_{169}H_{273}N_{41}O_{47}S_2$   $[M+2H]^{2+}$  calcd 1847.99, found 1848.30;  $[M+3H]^{3+}$  calcd 1232.33, found 1232.50;  $[M+4H]^{4+}$  calcd 924.50, found 924.65.



MALDI-TOF-MS:  $C_{169}H_{273}N_{41}O_{47}S_2$   $[M+H]^+$  calcd 3693.9752, found 3693.9730.

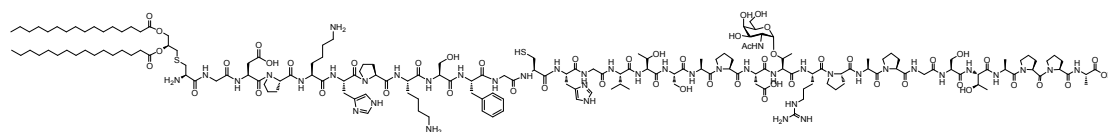


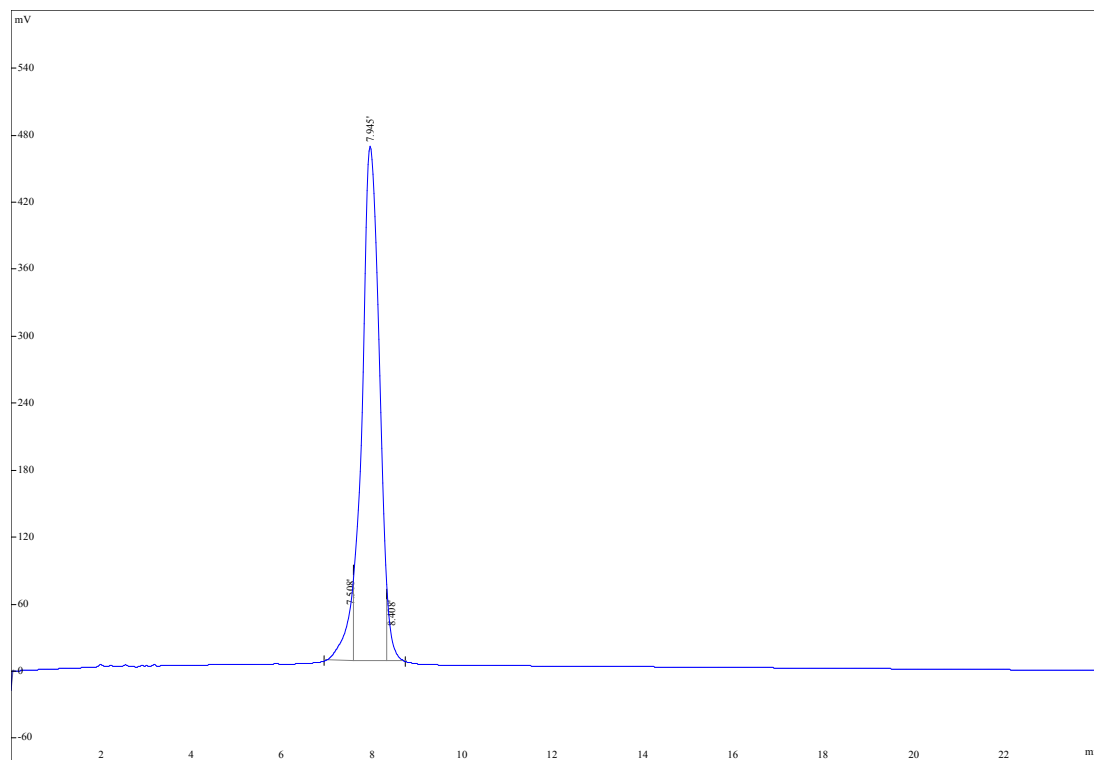


$^1\text{H}$  NMR (400 MHz, 90%  $\text{D}_2\text{O}$ +10%  $\text{DMSO-}d_6$ , selected signals)  $\delta$  8.71 (s, 2H, 2 $\times$ His-Im- $\text{H}_2$ ), 7.38 – 7.33 (m, 7H, Phe- $\text{C}_6\text{H}_5$ , 2 $\times$ His-Im- $\text{H}_4$ ), 1.39 (m, 12H, 4 $\times$ Ala- $\text{CH}_3$ ), 1.30 – 1.25 (m, 57H, 24 $\times$ Pal- $\text{CH}_2$ , 3 $\times$ Thr- $\text{CH}_3$ ), 1.01 (m, 6H, Val- $\text{CH}_3$ ), 0.91 (m, 6H, 2 $\times$ Pal- $\text{CH}_3$ ).

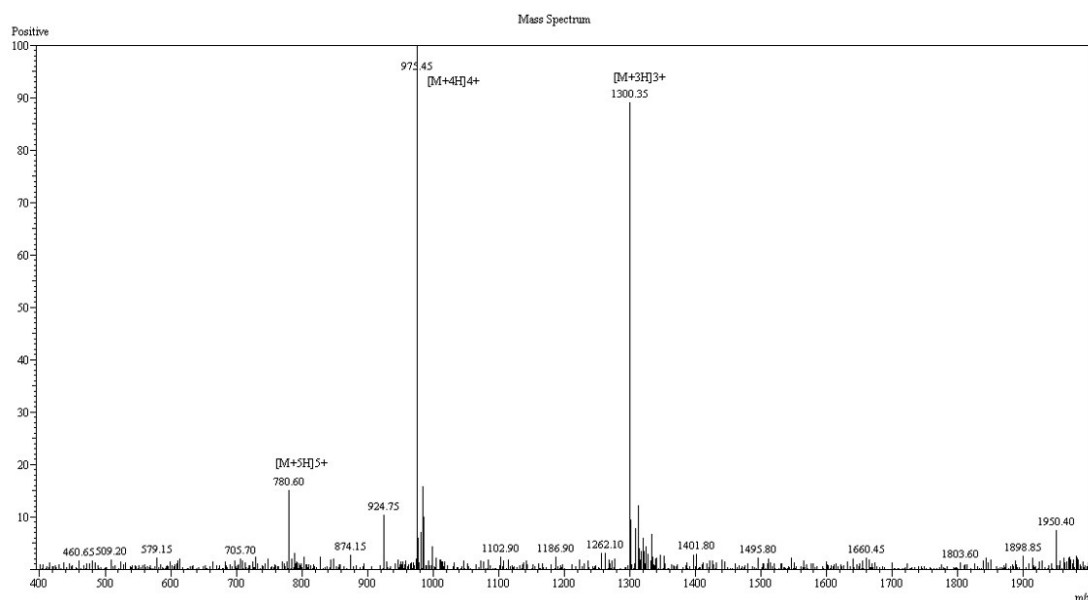
### Vaccine 8

H-S-[2,3-Bis(palmitoyloxy)-(2RS)-propyl]-(R)-cysteine-Gly-Asp-Pro-Lys-His-pro-Lys-Ser-Phe-Gly-Cys-His-Gly-Val-Thr-Ser-Ala-Pro-Asp-Thr( $\alpha$ -D-GalNAc)-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala-Pro-Pro-Ala-OH.

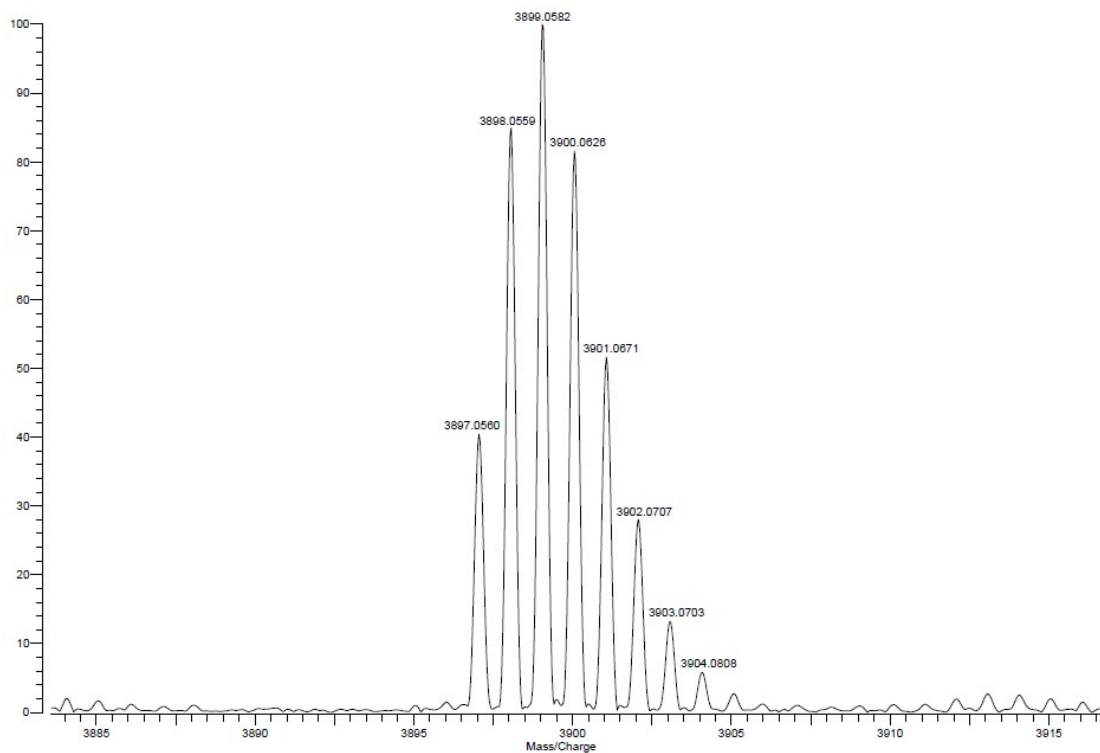




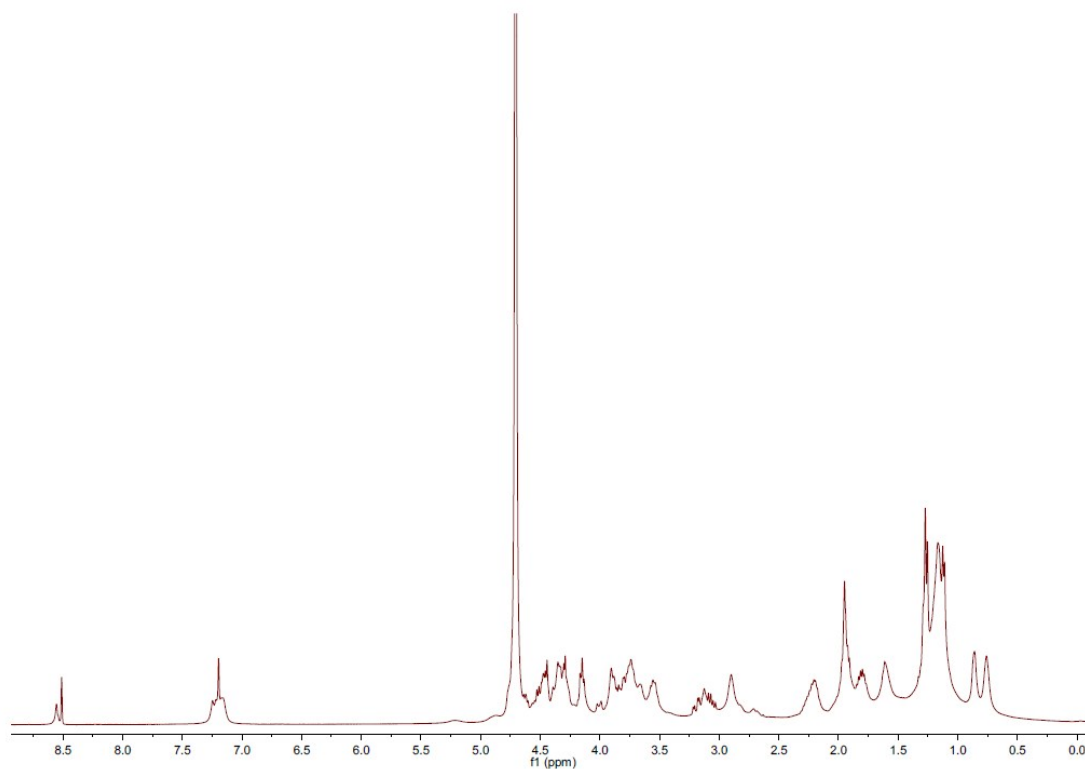
Analytical HPLC:  $R_t$  (retention time) = 7.9 min (60-100% of acetonitrile and 0.1% trifluoroacetic acid over 30 min on a C-18 column,  $\lambda=220$  nm). Yield 28% (8.9 mg).



ESI-MS:  $m/z$  for  $C_{177}H_{286}N_{42}O_{52}S_2$   $[M+2H]^{2+}$  calcd 1949.53, found 1950.40;  $[M+3H]^{3+}$  calcd 1300.02, found 1300.35;  $[M+4H]^{4+}$  calcd 975.27, found 975.45;  $[M+5H]^{5+}$  calcd 780.41, found 780.60.



MALDI-TOF-MS:  $C_{177}H_{286}N_{42}O_{52}S_2$   $[M+H]^+$  calcd 3897.0546, found 3897.0560.  
 $NH_2(CH_2CH_2O)_2CH_2C=O$

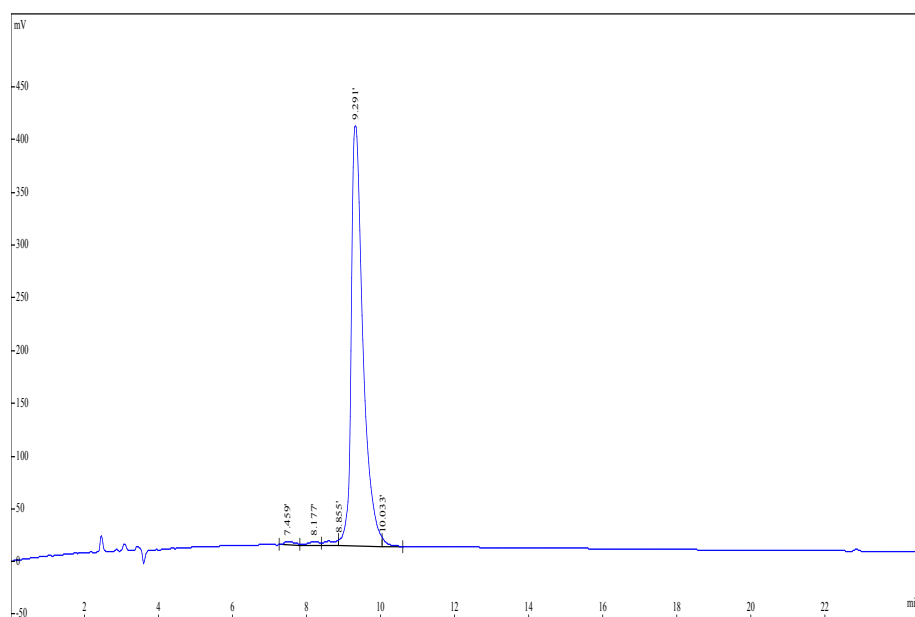
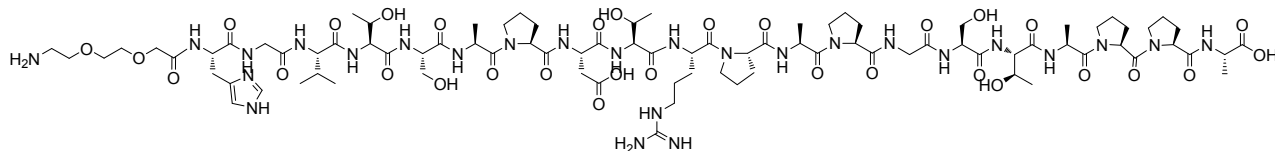


$^1H$  NMR (400 MHz,  $D_2O$ , selected signals)  $\delta$  8.63 (s, 1H, His-Im- $H_2$ ), 8.59 (s, 1H, His-Im- $H_2$ ), 7.33 – 7.24 (m, 7H, Phe- $C_6H_5$ , 2 $\times$ His-Im- $H_4$ ), 1.36 – 1.34 (m, 21H,

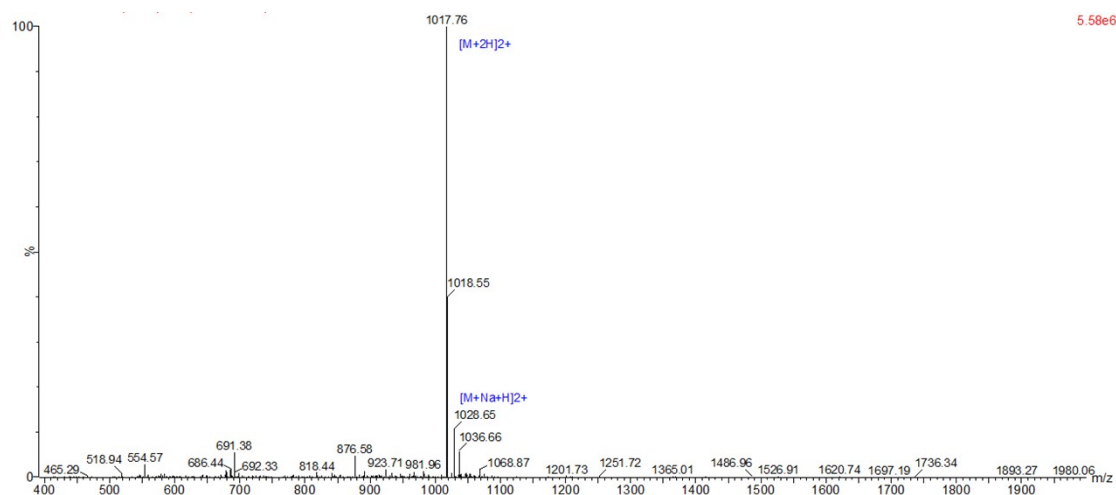
4×Ala-CH<sub>3</sub>, 3×Thr-CH<sub>3</sub>), 1.31 – 1.05 (m, 48H, 24×Pal-CH<sub>2</sub>), 0.94 (m, 6H, Val-CH<sub>3</sub>), 0.84 (m, 6H, 2×Pal-CH<sub>3</sub>).

### Antigen for ELISA plate coating

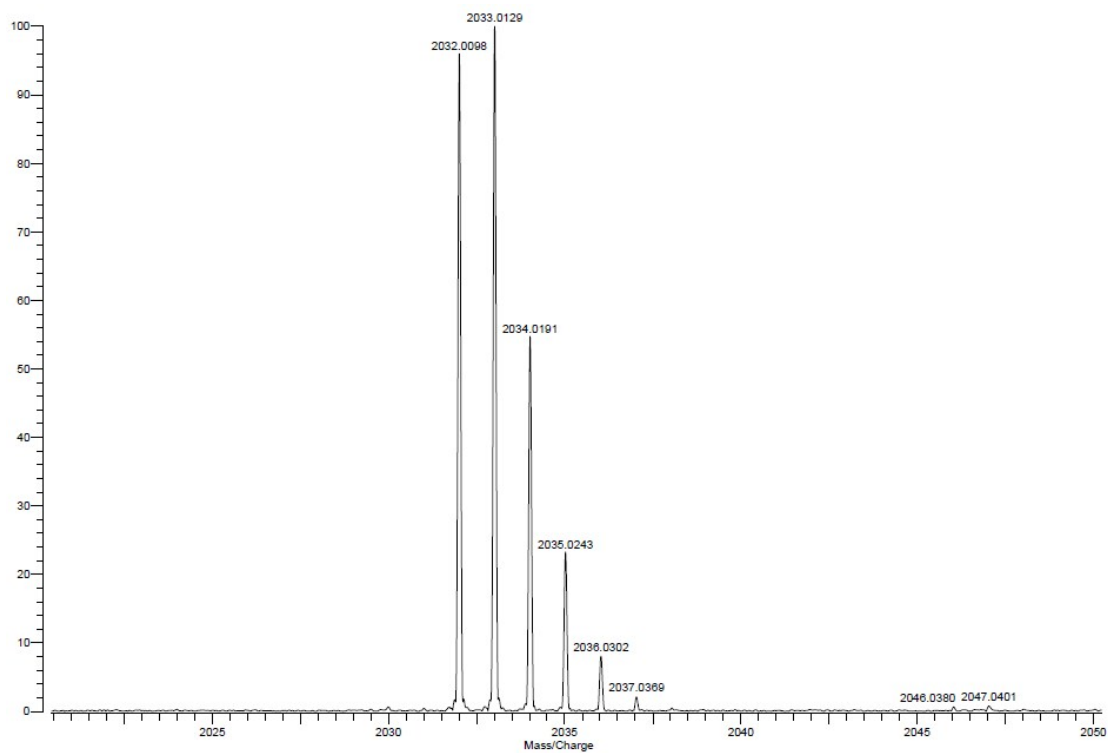
1) NH<sub>2</sub>(CH<sub>2</sub>CH<sub>2</sub>O)<sub>2</sub>CH<sub>2</sub>CO-His-Gly-Val-Thr-Ser-Ala-Pro-Asp-Thr-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala-Pro-Pro-Ala-OH (**12**)



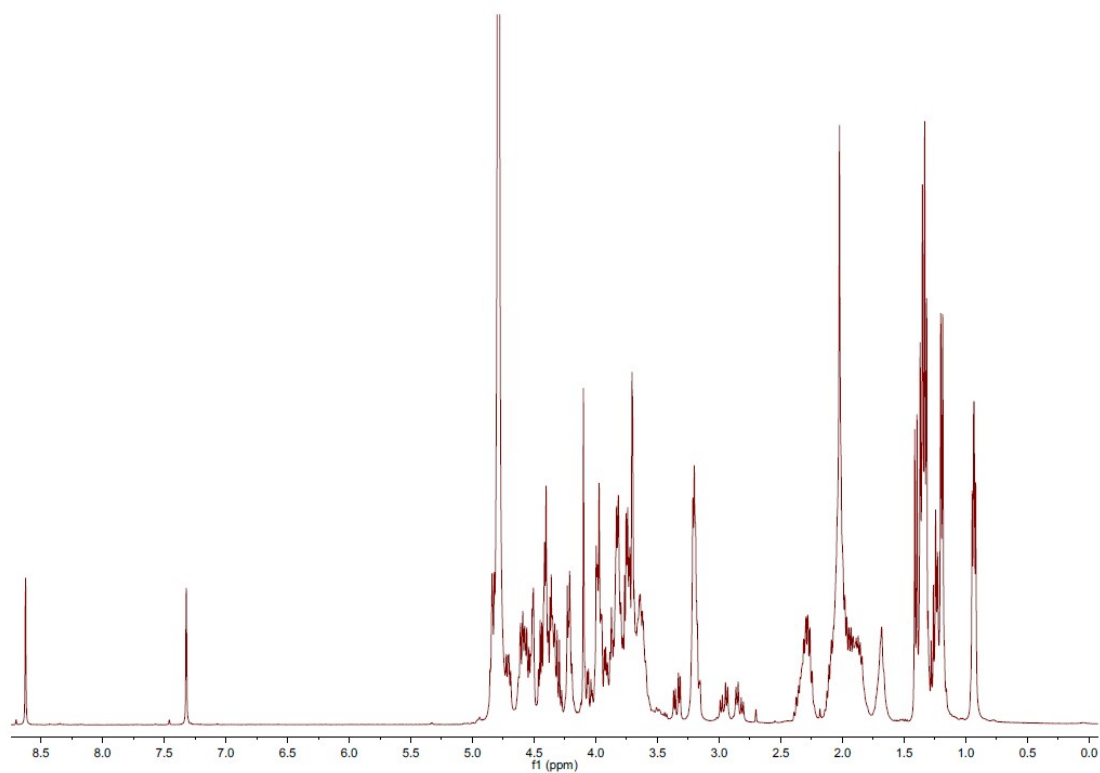
Analytical HPLC: R<sub>t</sub> (retention time) = 9.2 min (12-37% of acetonitrile and 0.1% trifluoroacetic acid over 25 min on a C-18 column, λ=220 nm). Yield 67% (15.4mg).



ESI-MS: *m/z* for C<sub>86</sub>H<sub>138</sub>N<sub>26</sub>O<sub>31</sub> [M+2H]<sup>2+</sup> calcd 1016.50, found 1017.76; [M+Na+H]<sup>2+</sup> calcd 1027.49, found 1028.65.

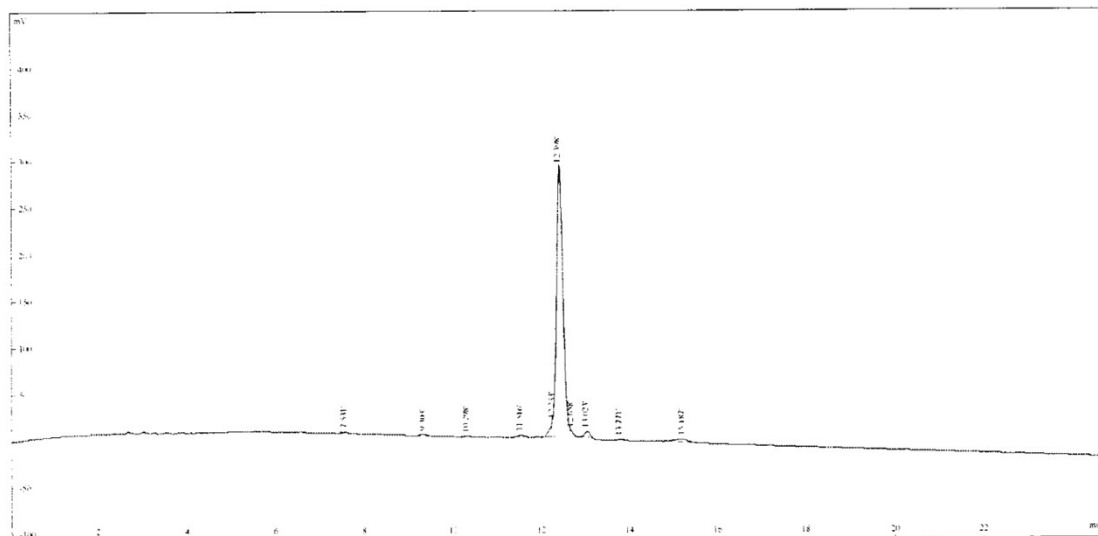
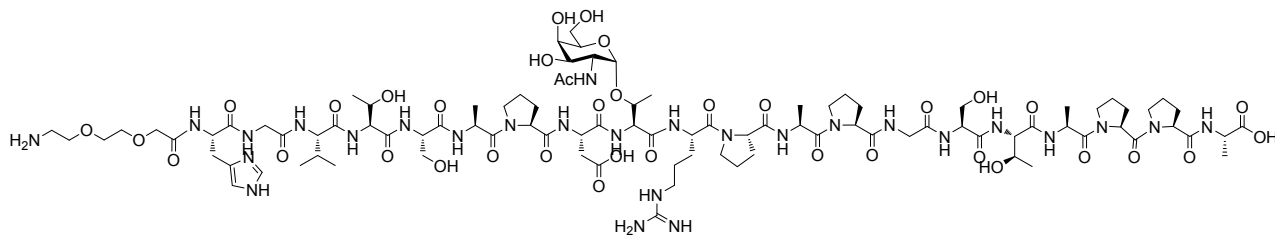


MALDI-TOF-MS:  $C_{86}H_{138}N_{26}O_{31} [M+H]^+$  calcd 2032.0100, found 2032.0098.

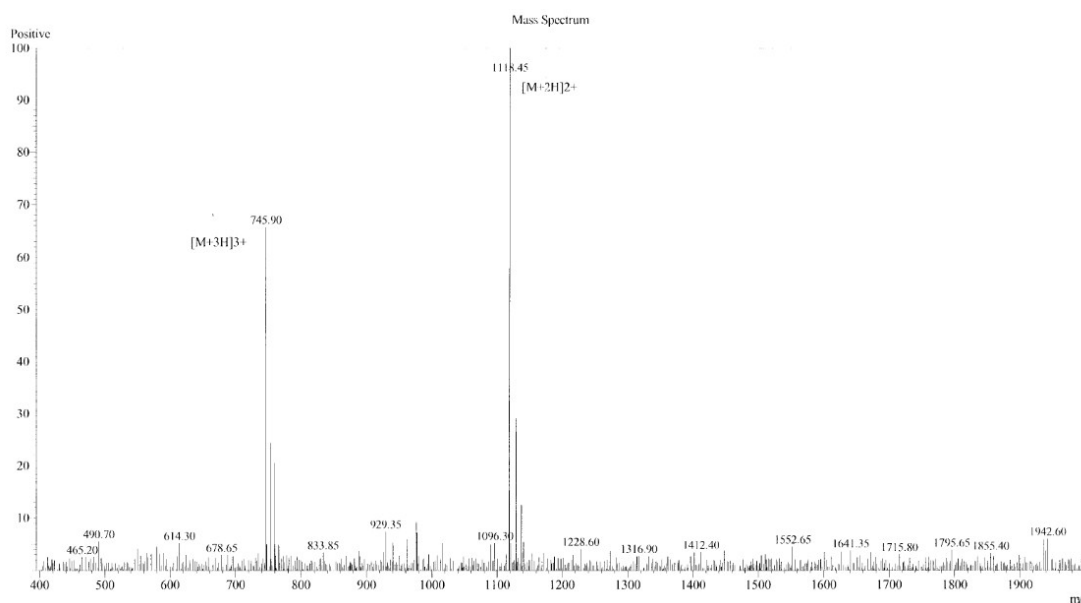


$^1H$  NMR (400 MHz,  $D_2O$ , selected signals)  $\delta$  8.62 (d,  $J = 1.1$  Hz, 1H, His-Im- $H_2$ ), 7.32 (s, 1H, His-Im- $H_4$ ), 1.37 – 1.31 (m, 12H, 4 $\times$ Ala- $CH_3$ ), 1.20– 1.88 (m, 9H, 3 $\times$ Thr- $CH_3$ ), 0.95– 0.92 (m, 6H, Val- $CH_3$ ).

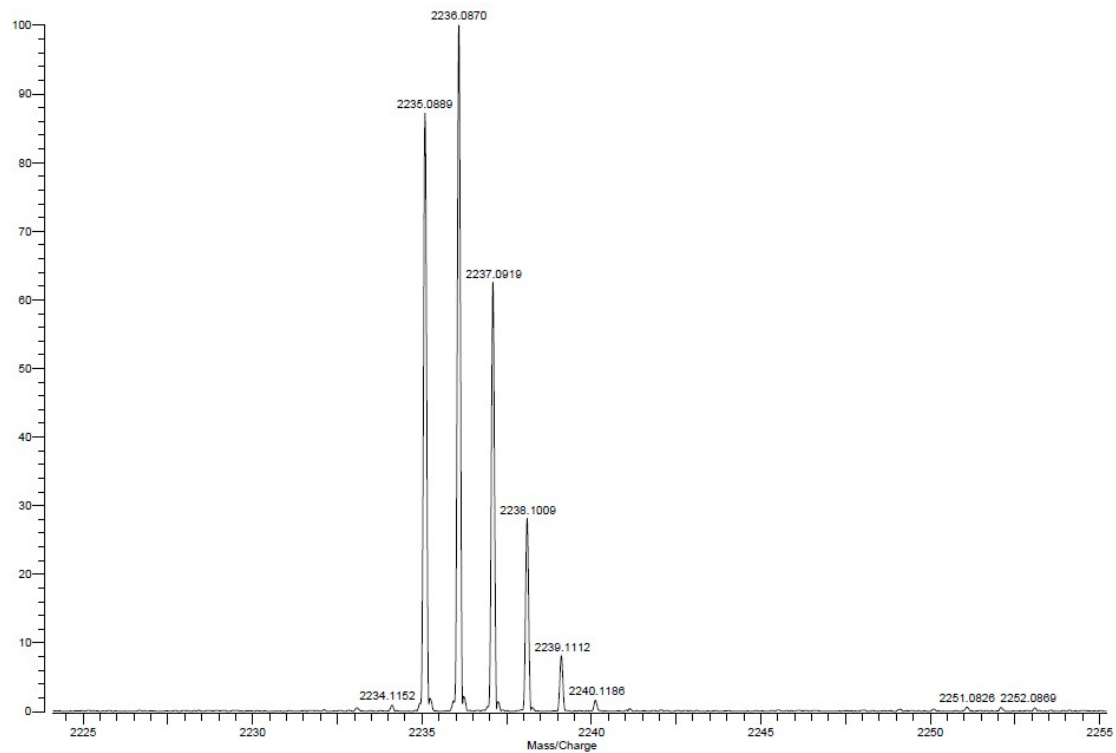
2)  $\text{NH}_2(\text{CH}_2\text{CH}_2\text{O})_2\text{CH}_2\text{CO-His-Gly-Val-Thr-Ser-Ala-Pro-Asp-Thr}(\alpha\text{-D-GalNAc})\text{-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala-Pro-Pro-Ala-OH}$  (**13**)



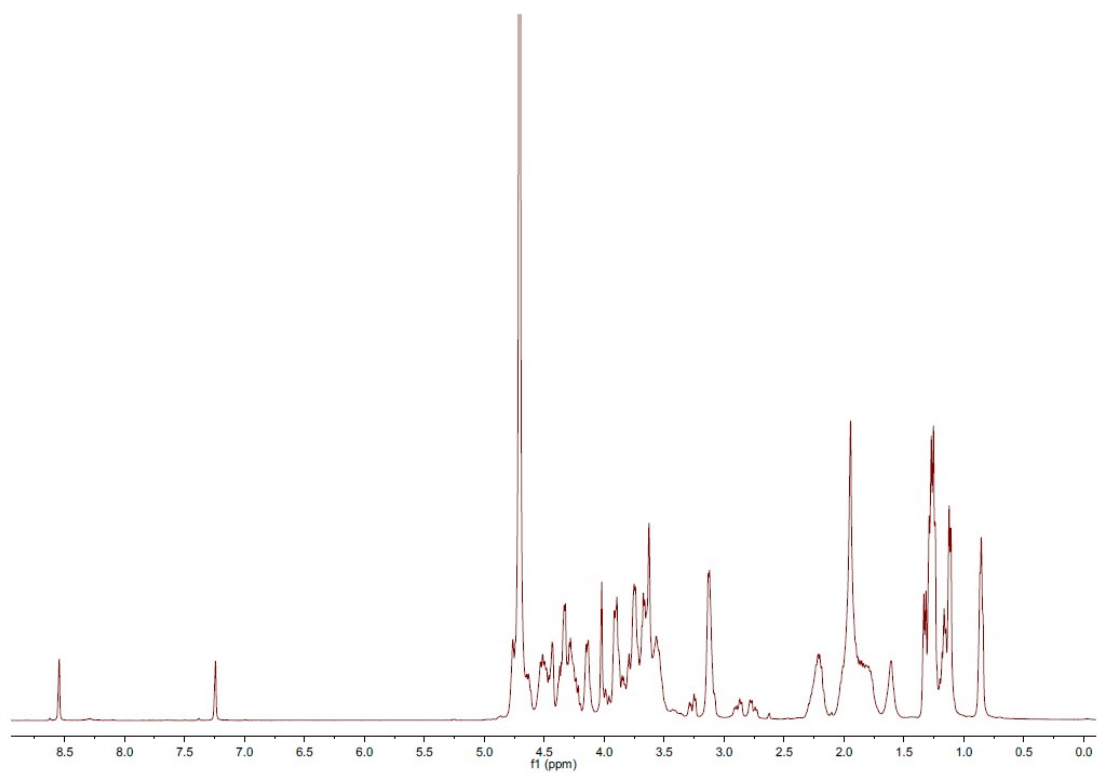
Analytical HPLC:  $R_t$  (retention time) = 12.3 min (10-35% of acetonitrile and 0.1% trifluoroacetic acid over 25 min on a C-18 column,  $\lambda=220$  nm). Yield 64% (13.8mg).



ESI-MS:  $m/z$  for  $\text{C}_{94}\text{H}_{151}\text{N}_{27}\text{O}_{36}$   $[\text{M}+2\text{H}]^{2+}$  calcd 1118.55, found 1118.45;  $[\text{M}+3\text{H}]^{3+}$  calcd 746.03, found 745.90;



MALDI-TOF-MS:  $C_{94}H_{151}N_{27}O_{36} [M+H]^+$  calcd 2235.0893, found 2235.0889.



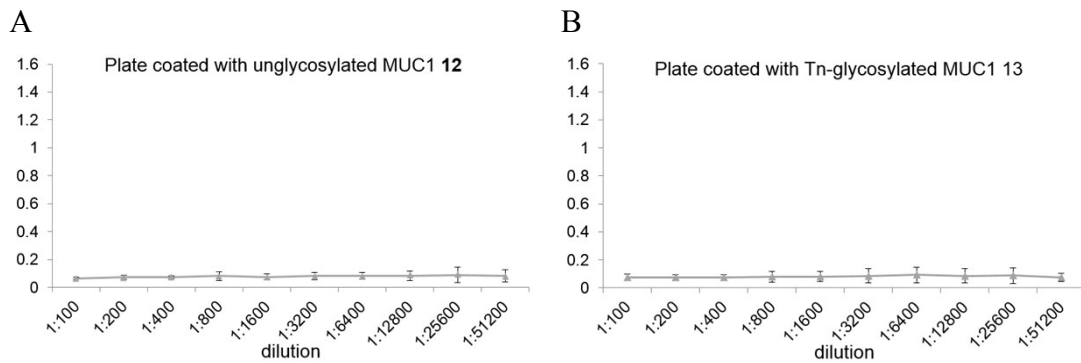
$^1H$  NMR (400 MHz,  $D_2O$ , selected signals)  $\delta$  8.63 (s, 1H, His-Im- $H_2$ ), 7.33 (s, 1H, His-Im- $H_4$ ), 1.37 – 1.32 (m, 12H, 4 $\times$ Ala- $CH_3$ ), 1.22 (m, 9H, 3 $\times$ Thr- $CH_3$ ), 0.95 (m, 6H, Val- $CH_3$ ).

## Immunological studies

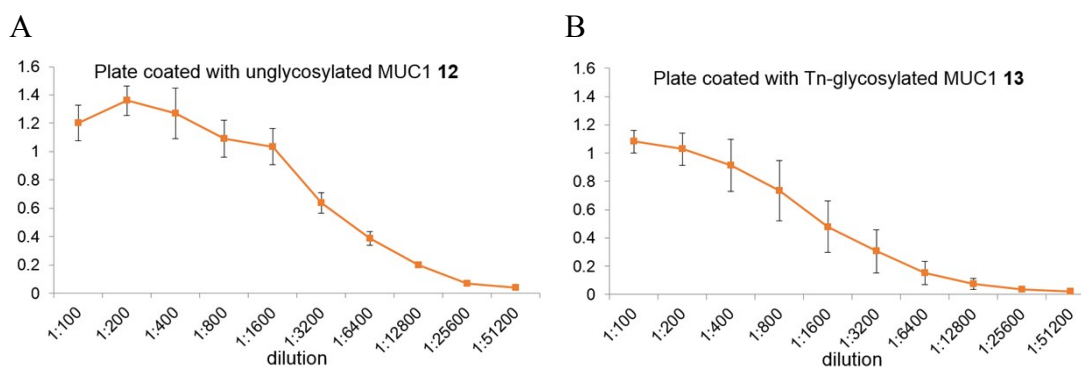
All the animal experiments were conducted at Beijing Institute of Pharmacology and Toxicology, followed the protocol of the Institutional Animal Care and Use Committee, which was in compliance with the guidelines of the Association for Assessment and Accreditation of Laboratory Animal Care International. Female Balb/c mice (6-8 weeks) were obtained and housed in institute's SPF grade animal facility, maintaining on a 12-hour light/dark cycle with free access to water and laboratory chow.

Mice were vaccinated via subcutaneous injections of 3.5 nmol vaccine candidates in 100  $\mu$ L PBS weekly for four times total. Blood samples were collected via tail vein seven days after the last vaccination and the anti-sera were prepared for further analysis and validation.

## MUC1-specific IgG cross-reactivity studies

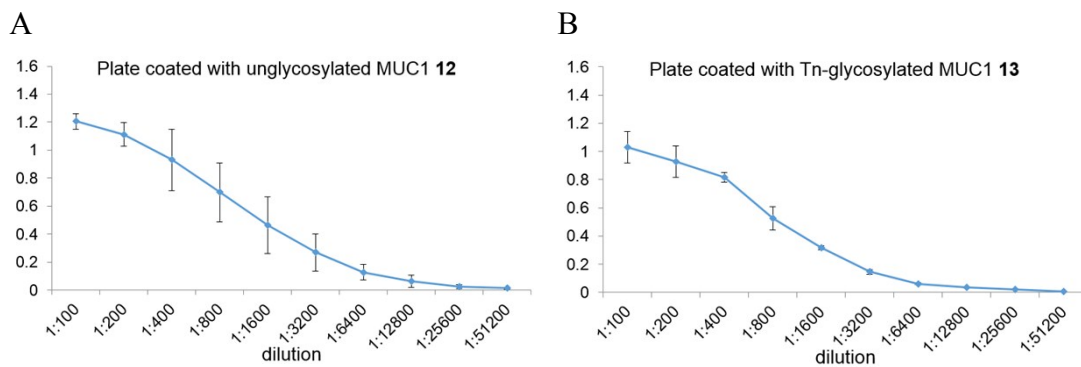


**Figure 1** ELISA of total IgG antibody specific for (A) unglycosylated MUC1 or (B) Tn-glycosylated MUC1. Sera were collected from preimmune mice.



**Figure 2** ELISA of total IgG antibody specific for (A) unglycosylated MUC1 or (B) Tn-glycosylated MUC1. Sera were collected from unglycosylated vaccine candidate 7 injected mice.





**Figure 3** ELISA of total IgG antibody specific for (A) unglycosylated MUC1 or (B) Tn-glycosylated MUC1. Sera were collected from Tn-glycosylated vaccine candidate **8** injected mice.

### Antibody Isotyping ELISA

For the qualitative determination of antigen specific antibody isotype in the antisera, the Mouse Monoclonal Antibody Isotyping kit (Sigma-Aldrich, USA) was applied here. The measurement was carried out according to the kit manual. The specific antigen parts for vaccine **7** and **8**, (glyco)peptides **9** and **10**, respectively, was coated on the 96-well plates and blocked by 10% fetal bovine serum before incubating the gradient diluted corresponding antisera. Each of the isotype antibodies from the kit was then applied, followed by HRP-labeled secondary antibody. TMB ELISA Substrates solution (Thermo, USA) was applied and  $H_2SO_4$  solution was added to stop the reaction before the plated being read by a plate reader.

### Mice sera cytokine profile after immunization

The cytokine profile in the sera after immunization by **7** and **8** was screened using Mouse Cytokine Antibody Array (Abcam, USA). The experiment was carried out as described in the manual. The cytokine antibody array membranes were incubated with 10-times diluted antisera after being blocked by blocking buffer. Then, the biotin-conjugated anti-cytokines antibody mixture was applied to incubate with each membrane, following by the incubation with HRP-conjugated streptavidin. The membranes were sent to exposure on X-ray film with chemiluminescence detection buffer.

After exposure, the X-ray film was scanned and the images were processed by ImageJ software (NIH, USA) to obtain the densitometry data of each image. The densitometry data from different membranes was subtracted for the background and normalized to the positive control spot on each membrane before the further compare between membranes.

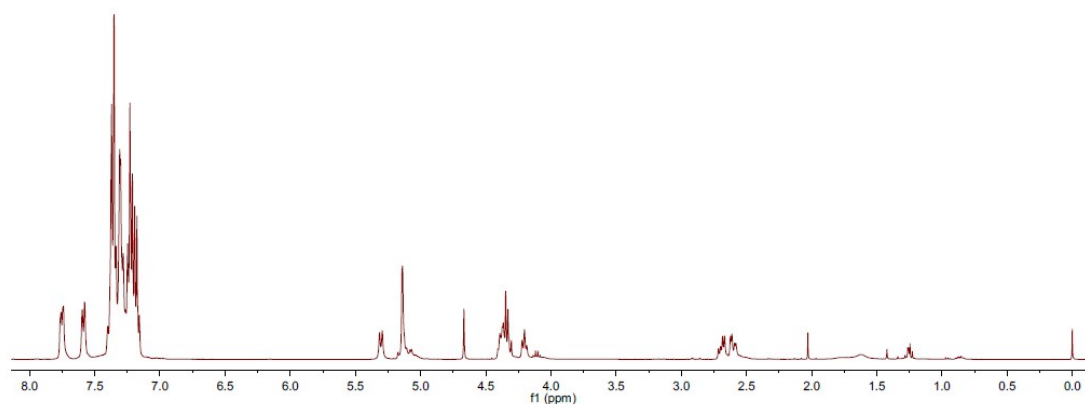
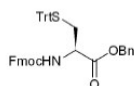
### Binding between antigen specific antibody and MCF-7 cells

Human breast cancer cells MCF-7 and human malignant melanoma cells SK-MEL-28 were cultured in DMEM culture medium containing fetal bovine serum (FBS, 10%) at 37 °C. The cells were digested with 0.25% (w/v) trypsin solution and washed three times with PBS solution containing 1% FBS. The cells suspensions  $5 \times 10^5$  per Eppendorf tube were incubated with antiserum (1:50 in PBS containing 1% FBS) at 4 °C for 1h. After washing with PBS containing FBS(1%) three times, the cells were incubated with FITC-conjugated rabbit anti-mouse IgG antibody (diluted 1:1000 in PBS containing 1% FBS, Sigma) at 4 °C for 1h. After washing three times with PBS containing 1% FBS, the cells were suspended in washing buffer (1 mL) and filtered through 200-mesh sieve, FACS analysis was conducted on BD FACSAria III flow cytometry.

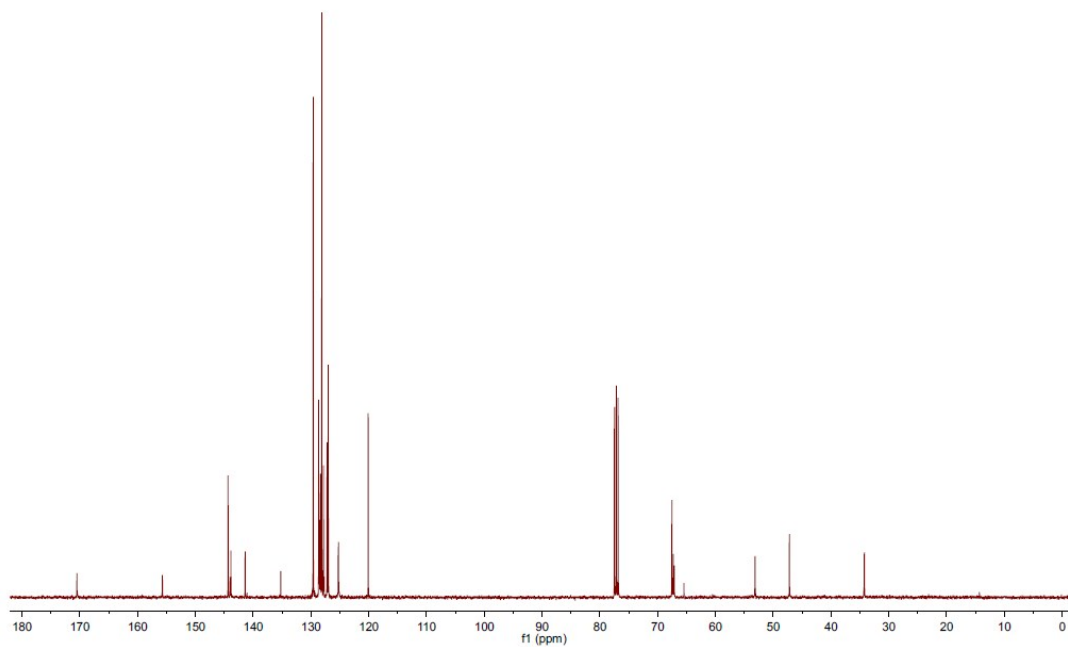
1. F. Reichel, A. M. Roelofsen, H. P. M. Geurts, T. I. Hämäläinen, M. C. Feiters and G.-J. Boons, *J. Am. Chem. Soc.* 1999, **121**, 7989-7997.

### $^1\text{H}$ NMR, $^{13}\text{C}$ NMR spectra

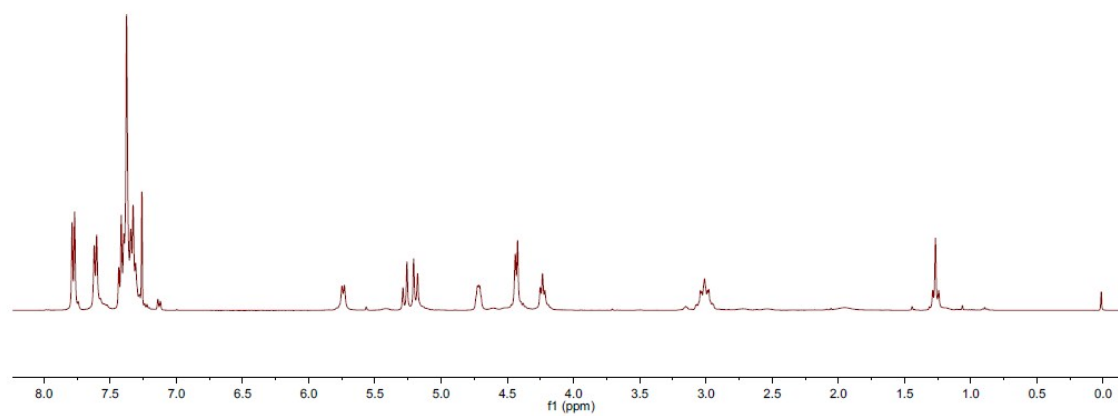
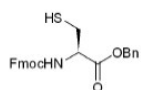
$^1\text{H}$  NMR Spectrum of compound **2** in  $\text{CDCl}_3$  (400 MHz)



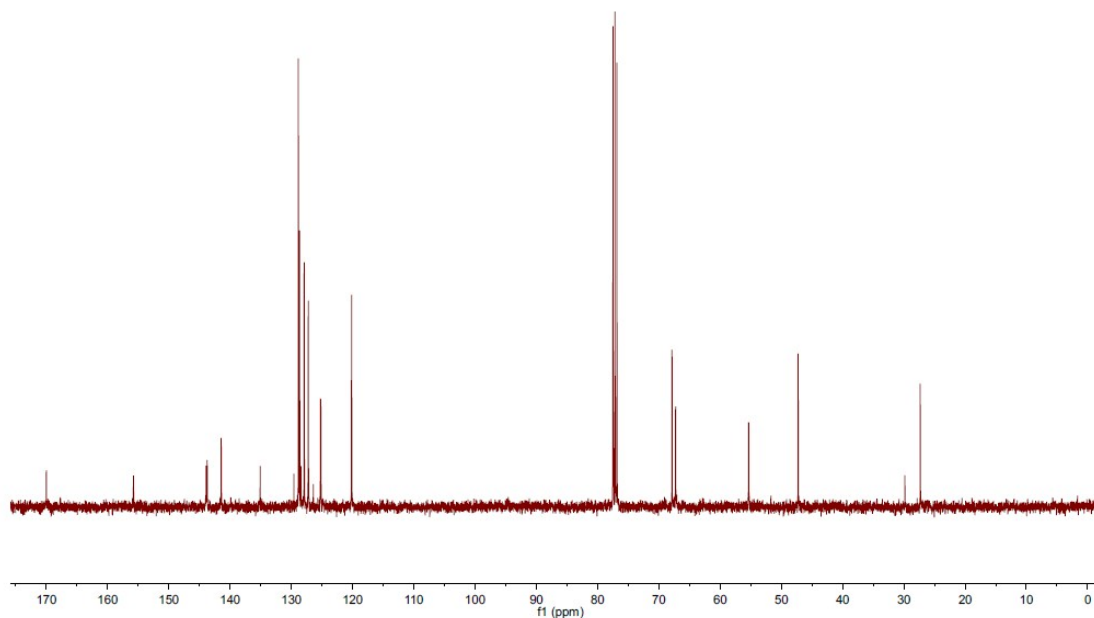
$^{13}\text{C}$  NMR Spectrum of compound **2** in  $\text{CDCl}_3$  (100 MHz)



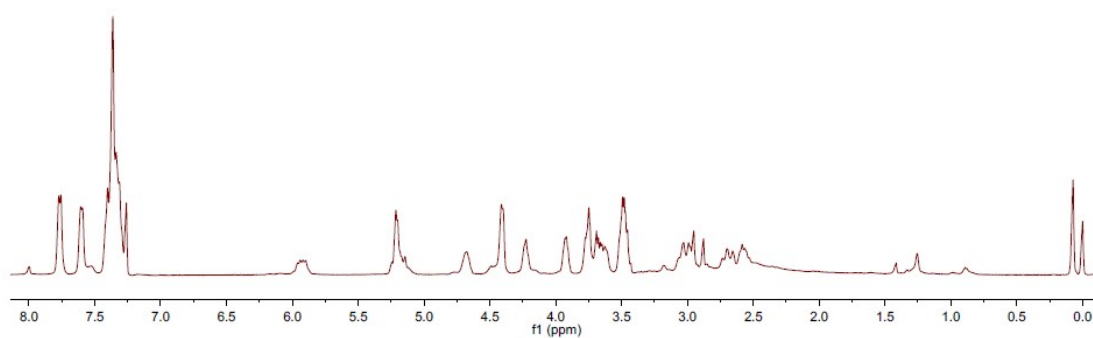
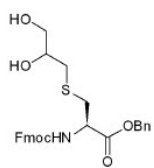
$^1\text{H}$  NMR Spectrum of compound **3** in  $\text{CDCl}_3$  (400 MHz)



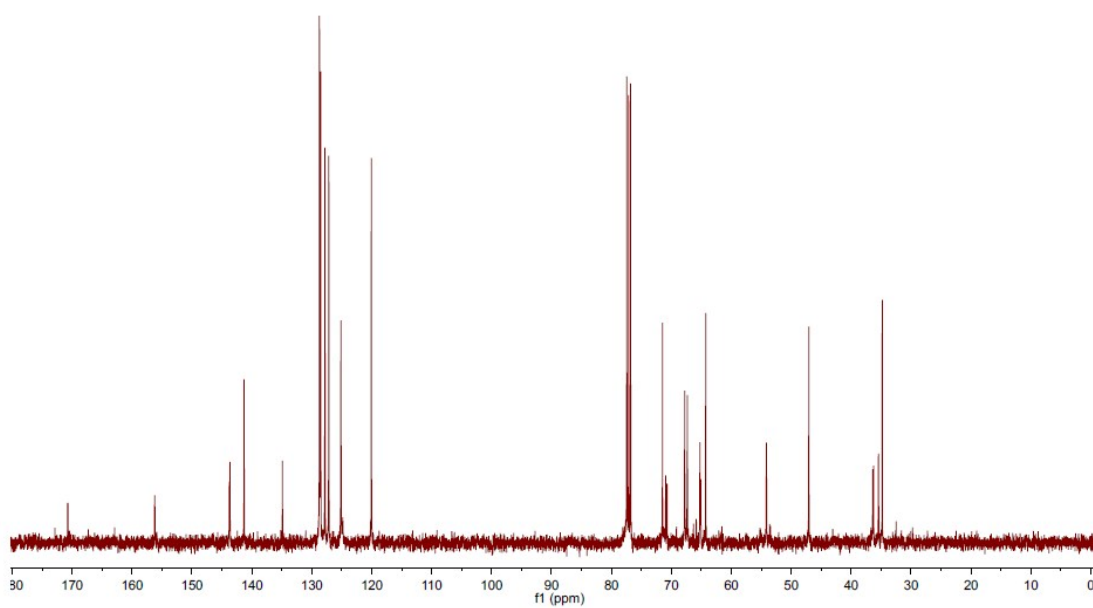
$^{13}\text{C}$  NMR Spectrum of compound **3** in  $\text{CDCl}_3$  (100 MHz)



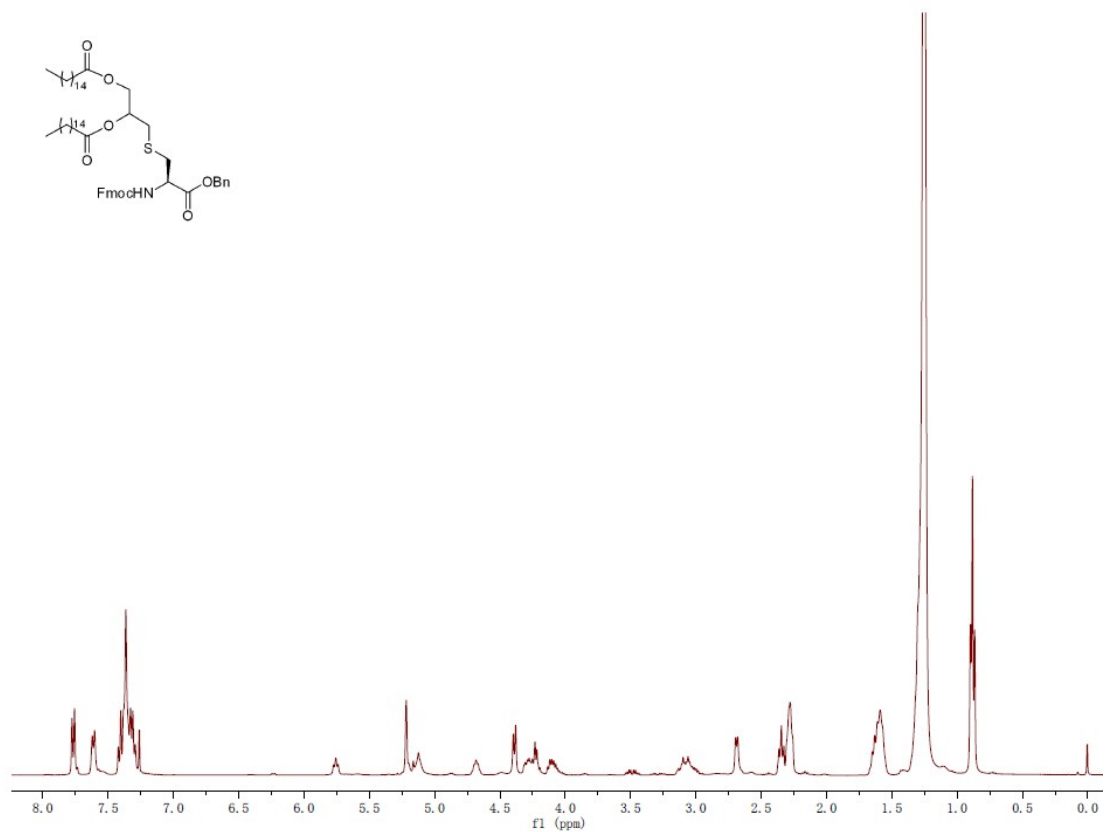
$^1\text{H}$  NMR Spectrum of compound **4** in  $\text{CDCl}_3$  (400 MHz)



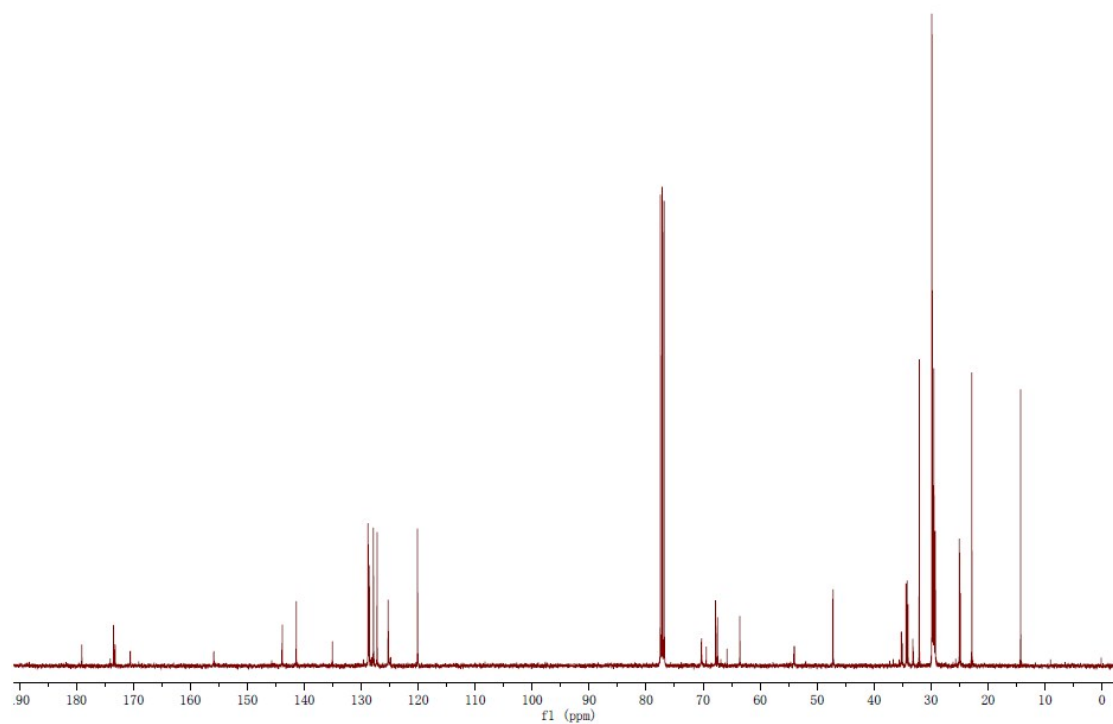
<sup>13</sup>C NMR Spectrum of compound 4 in CDCl<sub>3</sub> (100 MHz)



<sup>1</sup>H NMR Spectrum of compound 5 in CDCl<sub>3</sub> (400 MHz)



<sup>13</sup>C NMR Spectrum of compound 5 in CDCl<sub>3</sub> (100 MHz)



<sup>1</sup>H NMR Spectrum of compound 6 in CDCl<sub>3</sub> (400 MHz)

