

Supporting Information

Glycoconjugate synthesis using chemoselective ligation

Shuihong Cheng, Paeton L. Wantuch, Megan E. Kizer, Dustin R. Middleton,

Ruitong Wang, Mikaela DiBello, Mingli Li, Xing Wang, Xuebing Li, Fikri Y. Avci,

Fuming Zhang, Xing Zhang*, and Robert J. Linhardt*

Authors to whom correspondence should be addressed at linhar@rpi.edu and xingzhang-cas@hotmail.com

Table of Contents

General materials and methods.....	S2
General procedure for glycoconjugate synthesis.....	S3
MHCII binding studies of glycopeptides dp2- and dp4-OVA.....	S6
HPLC chromatograms, NMR and HRMS data.....	S8

I. General materials and methods

All chemicals unless noted otherwise were purchased from commercial sources and used as received. Yields are given after purification. NMR spectra were recorded on a Bruker 600 spectrometer with topspin 2.1.6 software at 298 K. Mass data were acquired by high-resolution ESI-MS (Thermo LTQ XL Orbitrap, Bremen, Germany).

II. General procedure for glycopeptides synthesis

a) Carbohydrate-ADH (adipic acid dihydrazide) synthesis

Carbohydrate (2.0 mg, 1.0 equiv) and ADH (50.0 equiv) were dissolved in 200 μ L acetate buffer (pH 5.5, 100 mM, containing 1 mM PhNH₂) at 50 °C and incubated for 24 h. The conjugation product was then recovered by Sephadex G-10 gel filtration (for small size oligosaccharides-ADH, Table S1, entries 1-7) or size exclusion spin column (MWCO 3000 Da) (for GAGs-ADH, Table S1, entries 8-15) and washed by distilled water and lyophilized.

Table S1. Carbohydrate substrate scope in dihydrazide reaction^[a]

Entry	Carbohydrate	Yield (%)
1	Lactose	99
2	Maltotriose	95
3	Maltotetraose	91
4	Maltooligosaccharide	87
5	6'-Sialyllactose	97
6	Glc-GlcA-Glc-GlcA	85
7	GlcA-Glc-GlcA-Glc	89
8	Heparin dodecasaccharide	80
9	Heparan sulfate	80 ^b
10	Heparin	80 ^b
11	LMW-heparin	80 ^b
12	<i>N</i> -desulfated heparin	90 ^b
13	Chondroitin sulfate A	80 ^b
14	Chondroitin sulfate E	80 ^b
15	Dermatan sulfate	50 ^b

^[a]The ADH reaction with carbohydrate was conducted at PhNH₂/acetate buffer (pH 5.5) conditions described in entry 6, Table 1. ^[b]Yields of ADH reaction at the reducing end of a heterogeneous, polydisperse polysaccharides were estimated by ¹H-NMR.

b) Chemoselective ligation with N-terminal cysteine-peptides

Peptide^a (2.0 mg, 1.0 equiv) and carbohydrate-ADH (2.0 equiv) were dissolved in 0.2 mL of ligation buffer (3.0 M $\text{Gn}\cdot\text{HCl}$, 0.2 M Na_2HPO_4 , pH 3.8), 20 μL of NaNO_2 (200 mM) was added dropwise, and the reaction mixture was stirred for 20 min at -10°C . After that, 0.2 mL of freshly prepared MPAA (200 mM, dissolved in ligation buffer pH 7.0) was added, and the acidity of the mixed solution was adjusted to pH 7.0 with NaOH (2.0 M) slowly. The reaction mixture was stirred at room temperature for 2 hours. Before analysis, the reaction solution was reduced by 30 mM neutral tris(2-carboxyethyl)phosphine(TCEP). The reactions were monitored by HPLC and conjugation product were recovered by size exclusion spin column (MWCO 3000 Da) and Sephadex G-10 gel filtration washed by distilled water. The linkage of the corresponding conjugate is amido bond.

^a N-terminal cysteine-peptides with no modification.

c) Chemoselective ligation with other peptides

Peptide^b (2.0 mg, 1.0 equiv) and carbohydrate-ADH (2.0 equiv) were dissolved in 0.2 mL of ligation buffer (3.0 M $\text{Gn}\cdot\text{HCl}$, 0.2 M Na_2HPO_4 , pH 3.8), 20 μL of NaNO_2 (200 mM) was added dropwise, and the reaction mixture was stirred for 20 min at -10°C . After that, 0.2 mL of freshly prepared MPAA (200 mM, dissolved in ligation buffer pH 7.0) was added, and the acidity of the mixed solution was adjusted to pH 7.0 with NaOH (2.0 M) slowly. The reaction mixture was stirred at room temperature for 2 hours. The

reactions were monitored by HPLC and conjugation product were recovered by size exclusion spin column (MWCO 3000 Da) and Sephadex G-10 gel filtration washed by distilled water. The linkage of the corresponding conjugate is thioester bond.

^bC-terminal, internal cysteine-peptides or acetyl protected N-terminal cysteine-peptides.

Table S2. Conjugation of carbohydrate-adipic acid dihydrazide (ADH) with thiol containing compounds.^a

Entry	Carbohydrate-ADH	Aglycones	Conditions	Yield
1	Lactose	3-mercaptopropionic acid	NaNO ₂ , MPAA	86
2	Lactose	T20-Cys ^b	NaNO ₂ , MPAA	80
3	Maltotriose	T20-Cys	NaNO ₂ , MPAA	88
4	Lactose	C34(internal Cys) ^c	NaNO ₂ , MPAA	45
5	6'-Sialyllactose	T20-Cys	NaNO ₂ , MPAA	89

^aThe linkages of all the conjugation products are thioester bonds. ^bT20-Cys is a modified version of ENF with a cysteine at its C-terminus and the sequence is Ac-YTSLIHSLEESQNQQEKNEQELLELDKWASLWNWFC; ^cC34 (internal Cys), Ac-WMEWDREINCYTSLIHSLEESQNQQEKNEQELL, having a cysteine in the middle of its chain C34, which is a fusion polypeptide inhibitor of HIV.

III. MHCII binding studies of glycopeptides dp2- and dp4-OVA.

a) Disaccharide GlcA(1→4)Glc and tetrasaccharide

GlcA(1→4)Glc(1→3)GlcA(1→4)Glc preparation

Pure type 3 capsular polysaccharide (CPS, 40 mg) (ATCC 172-X) was added to 20 mL distilled H₂O in a flask and heated on a hot plate. When most of the CPS is dissolved, 20 mL of 0.6 M TFA was added to the flask and continued to heat until fully dissolved and then incubated for exactly 2 hours at 100 °C. Immediately cool on ice and aliquot into tubes, speed vac to dryness. Re-suspend pellets in H₂O and speed vac to dryness again. Re-suspended contents of all tubes in 2 mL H₂O, centrifuged and the supernatant was subjected to Superdex 30 column to obtain the target compounds.

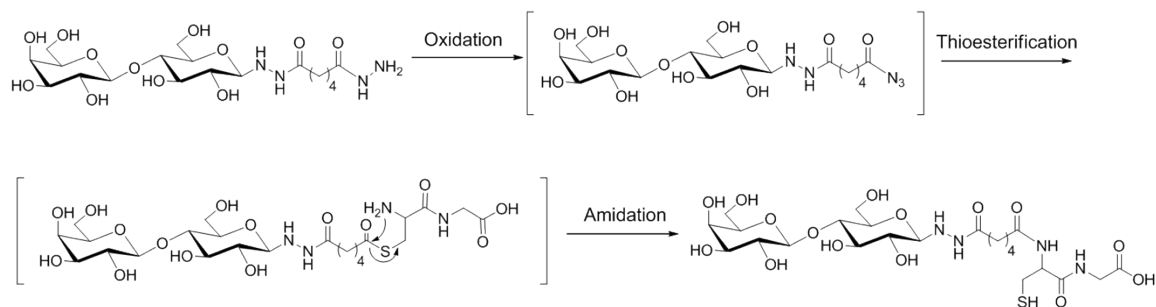
b) Synthesis of glycopeptides dp2- and dp4-OVA

The procedures of dp2-ADH and dp4-ADH, and the corresponding conjugates with Cys-OVA (sequence: CISQAVHAAHAEINEAGR) are described in Section II.

c) MHCII binding test

Purified MHCII monomers (mouse allele I-A^d) were graciously provided by the NIH tetramer facility. Monomers came loaded with a placeholder peptide with a 3C protease cleavage site. Peptide is cleaved by treating monomer with 3C protease (Pierce™ HRV 3C Protease Solution Kit) for 8 hours at room temperature. Cleaved monomer with empty binding groove is then loaded with desired peptide or glycan-peptide through an

exchange reaction. Exchange reactions use 200 μg cleaved monomer with 270 μM peptide (or glycan-peptide) sample at pH 5.0 in a citrate buffer. Samples for this experiment were OVA peptide 323-339, scrambled OVA peptide, OVA peptide 323-339 with an added Cys residue at the N terminus, disaccharide-peptide conjugate, and tetrasaccharide-peptide conjugate. Reactions were incubated for 5 days at room temperature. At the end of the incubation, reactions were neutralized with 1 M sodium phosphate buffer pH 7.5 and spun down at max speed for 10 minutes to remove aggregates. Absorbance was measured at 280, all samples except the scrambled OVA peptide reaction gave significant concentration values. Binding was confirmed by isoelectric focusing (NovexTM pH 3-7 Protein IEF gels). Gel was visualized using silver stain (PierceTM Silver Stain Kit).



Scheme S1. Synthesis of lactose-dipeptide (Cys-Gly).

IV. HPLC chromatograms, NMR and HRMS data.

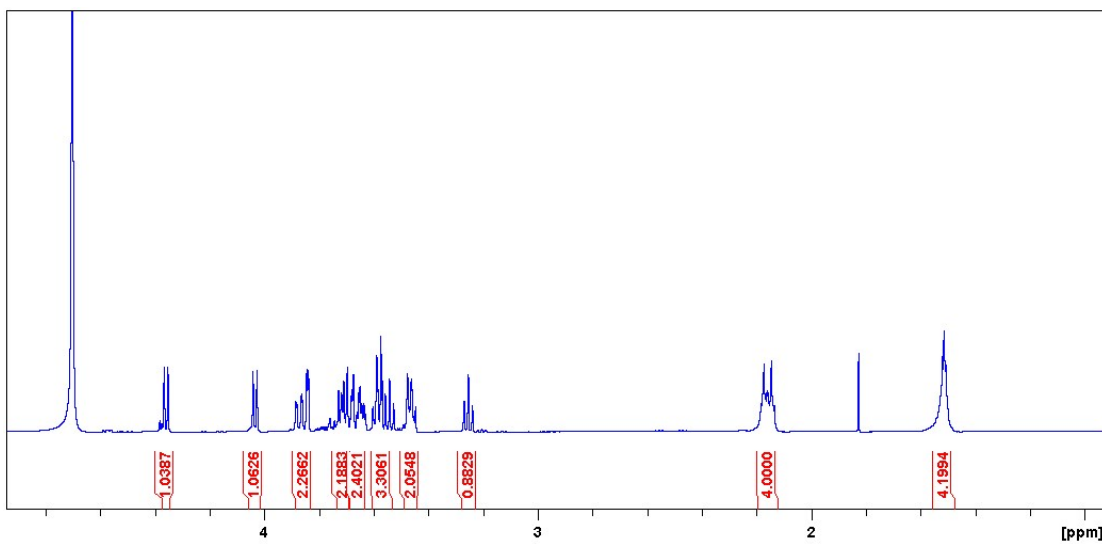


Figure S1. ^1H NMR spectra of lactose-ADH. White solid, ^1H NMR (600 MHz, D_2O) δ : 4.36 (d, $J = 7.79$, 1H, Gal-H1), 4.04 (d, $J = 9.03$, 1H, Glc-H1), 3.89-3.83 (m, 2H), 3.73-3.69 (m, 2H), 3.68-3.62 (m, 2H), 3.60-3.52 (m, 3H), 3.50-3.44 (m, 2H), 3.27-3.23 (m, 1H), 2.20-2.12 (m, 4H), 1.57-1.47 (m, 4H).

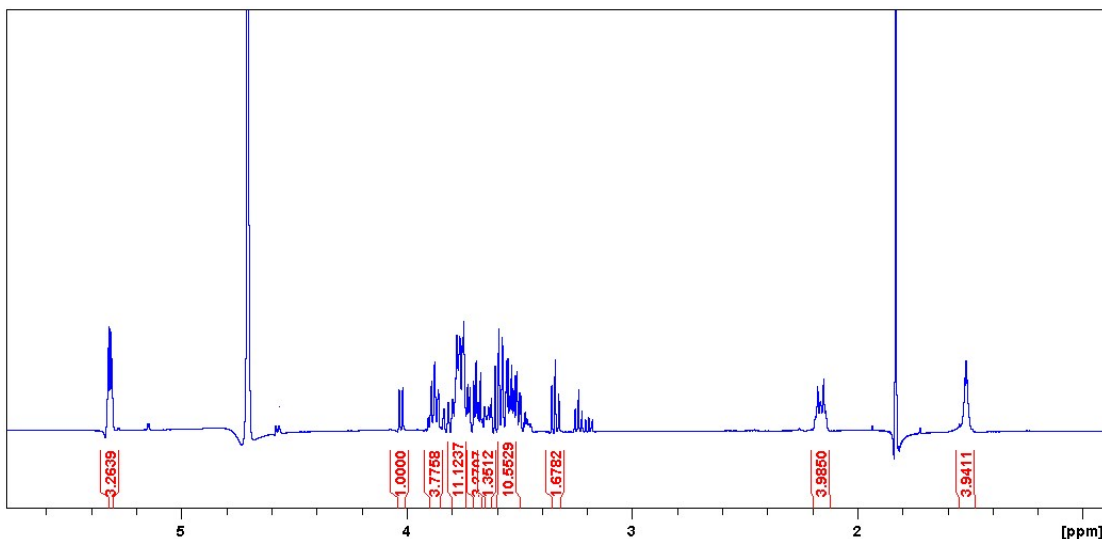


Figure S2. ^1H NMR spectra of maltotetraose-ADH. White solid, ^1H NMR (600 MHz, D_2O) δ : 5.32-5.30 (m, 3H), 4.02 (d, $J = 9.23$, 1H), 3.89-3.85 (m, 3H), 3.80-3.73 (m, 11H), 3.70-3.66 (m, 3H), 3.65-3.62 (m, 1H), 3.60-3.49 (m, 10H), 3.40 (t, $J = 9.48$, 1H), 2.19-2.13 (m, 4H), 1.53-1.49 (m, 4H).

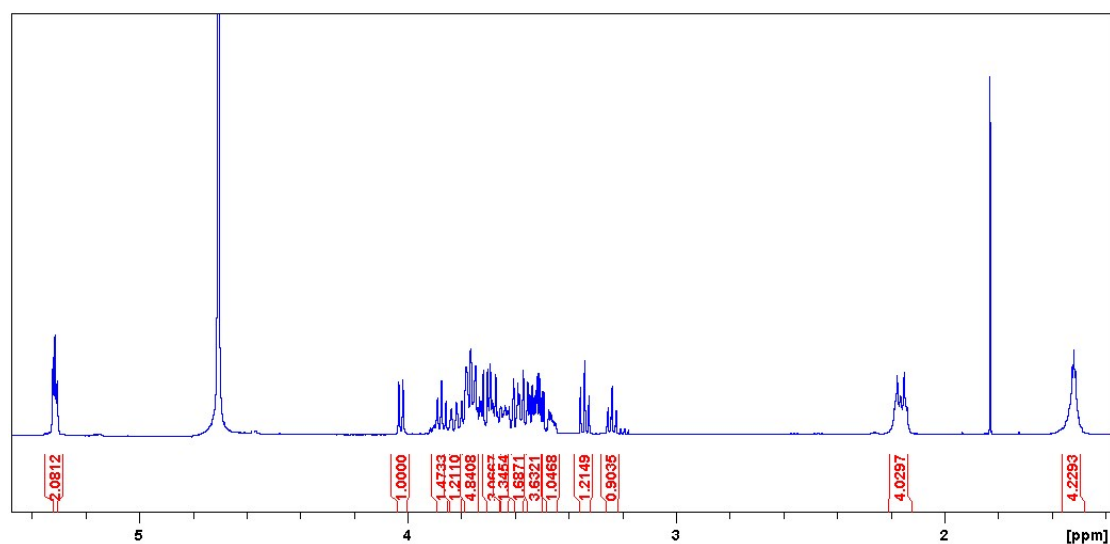


Figure S3. ^1H NMR spectra of maltotriose-ADH. White solid, ^1H NMR (600 MHz, D_2O) δ : 5.39-5.29 (m, 2H), 4.02 (d, $J = 9.02$, 1H), 3.89-3.85 (m, 1H), 3.84-3.80 (m, 1H), 3.79-3.73 (m, 5H), 3.70-3.66 (m, 3H), 3.65-3.62 (m, 1H), 3.60-3.56 (m, 1H), 3.55-3.49 (m, 3H), 3.47-3.44 (m, 1H), 3.34 (t, $J = 9.50$, 1H), 3.26-3.22 (m, 1H), 2.20-2.13 (m, 4H), 1.53-1.49 (m, 4H).

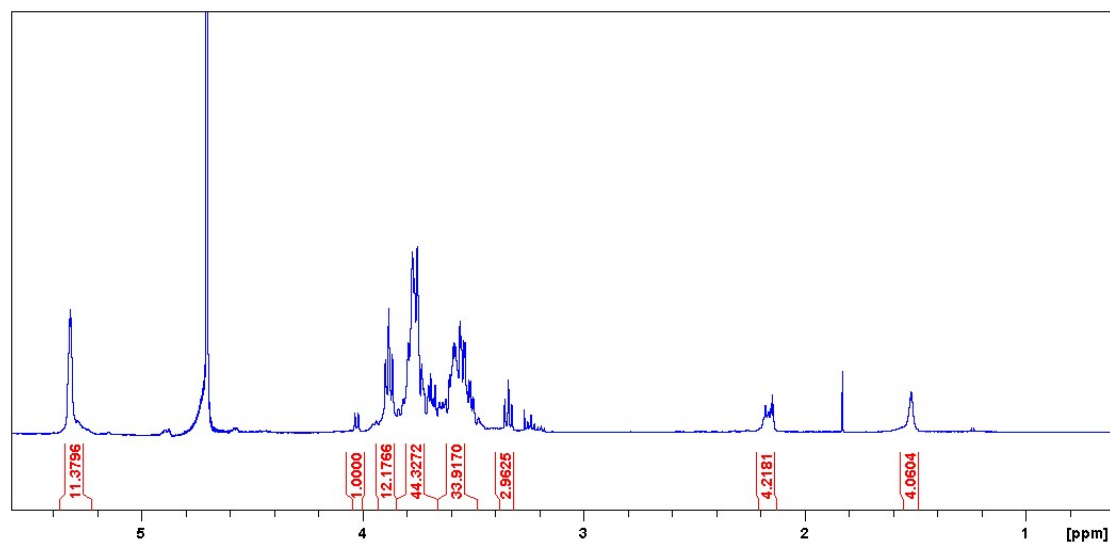


Figure S4. ^1H NMR spectra of maltooligosaccharide-ADH. White solid, ^1H NMR (600 MHz, D_2O) δ : 5.36-5.21 (m, 11H), 4.02 (d, $J = 9.02$, 1H), 3.92-3.85 (m, 12H), 3.83-3.71 (m, 44H), 3.66-3.47 (m, 34H), 3.36-3.31 (m, 3H), 2.20-2.13 (m, 4H), 1.54-1.50 (m, 4H).

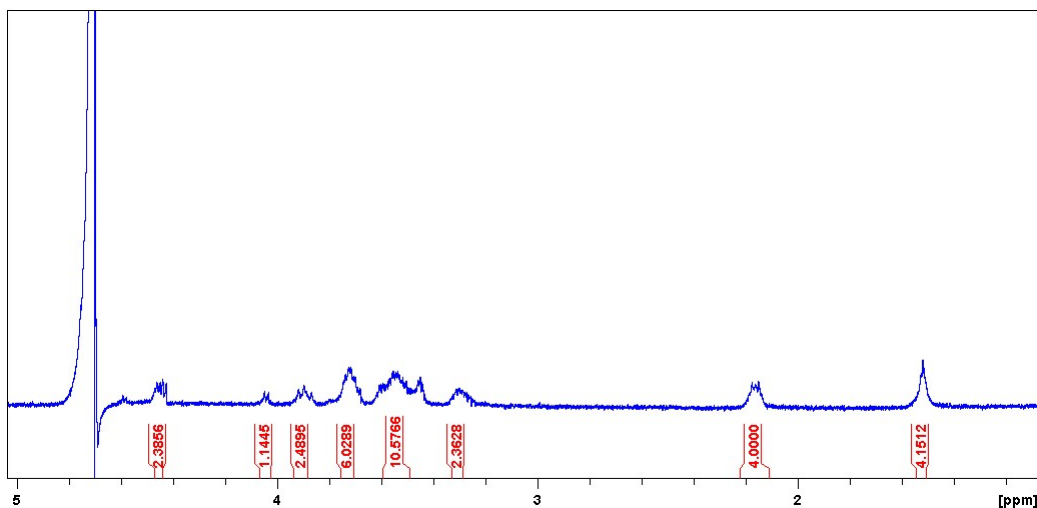


Figure S5. ^1H NMR spectra of Pn3P-tetrasaccharide GlcA-Glc-GlcA-Glc-ADH. White solid, ^1H NMR (600 MHz, D_2O) δ : 4.70 (1H), 4.50-4.42 (m, 2H), 4.04 (d, $J = 9.02$, 1H), 3.94-3.86 (m, 2H), 3.77-3.67 (m, 6H), 3.62-3.42 (m, 10H), 3.34-3.24 (m, 2H), 2.20-2.13 (m, 4H), 1.56-1.48 (m, 4H).

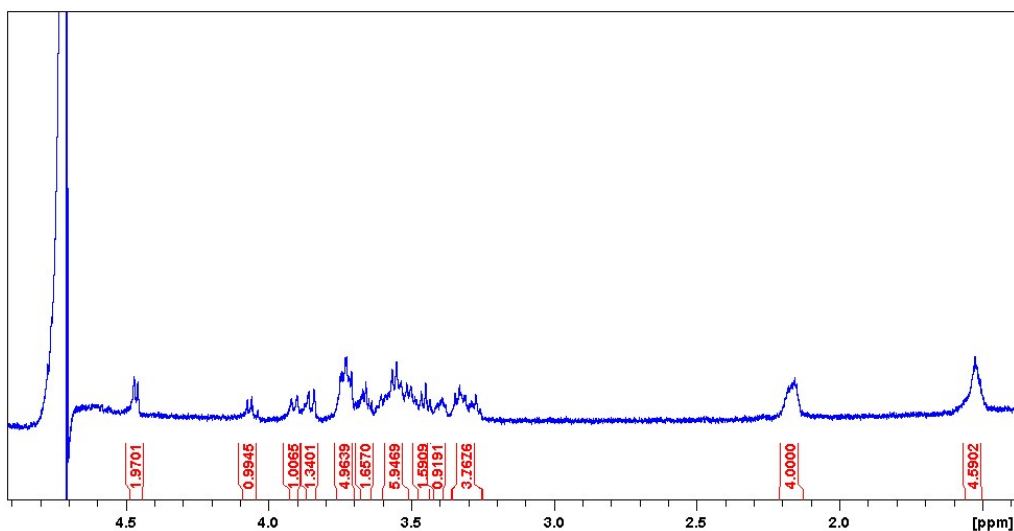


Figure S6. ^1H NMR spectra of Pn3P-tetrasaccharide Glc-GlcA-Glc-GlcA-ADH. White solid, ^1H NMR (600 MHz, D_2O) δ : 4.70 (2H), 4.47-4.44 (m, 2H), 4.07-4.03 (m, 1H), 3.94-3.90 (m, 1H), 3.87-3.82 (m, 1H), 3.77-3.69 (m, 5H), 3.69-3.63 (m, 1H), 3.60-3.51 (m, 6H), 3.48-3.44 (m, 1H), 3.41-3.37 (m, 1H), 3.35-3.25 (m, 3H), 2.20-2.13 (m, 4H), 1.56-1.48 (m, 4H).

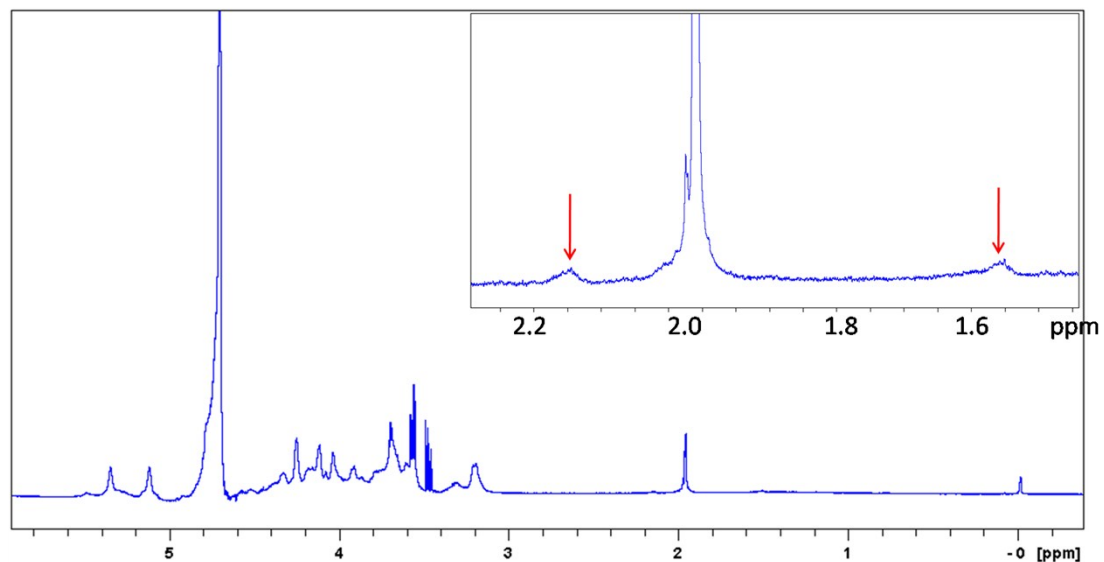


Figure S7. ^1H NMR spectra of heparin-ADH. Two sets of peaks at 1.51 and 2.16 ppm in the ^1H NMR spectra (insert) correspond to dihydrazide adipic linker.

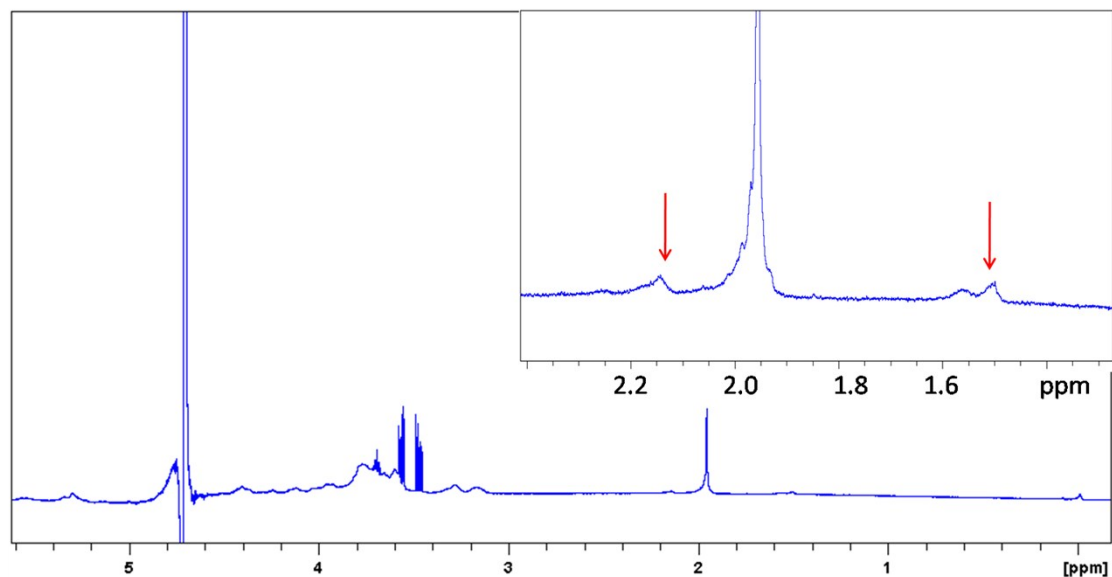


Figure S8. ^1H NMR spectra of heparan sulfate-ADH. Two sets of peaks at 1.51 and 2.16 ppm in the ^1H NMR spectra (insert) correspond to dihydrazide adipic linker.

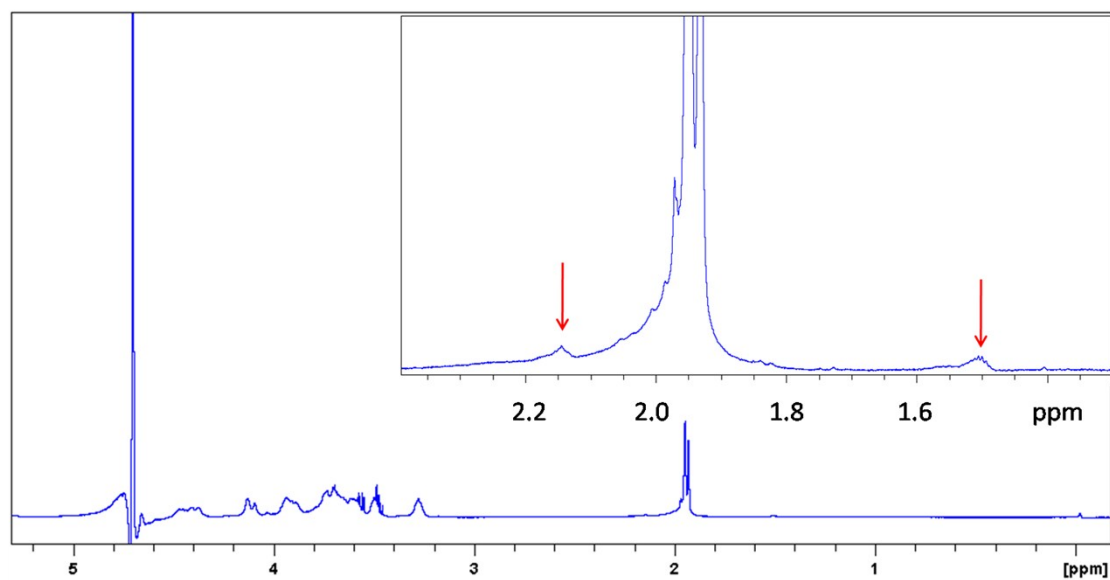


Figure S9. ^1H NMR spectra of chondroitin sulfate A-ADH. Two sets of peaks at 1.51 and 2.16 ppm in the ^1H NMR spectra (insert) correspond to dihydrazide adipic linker.

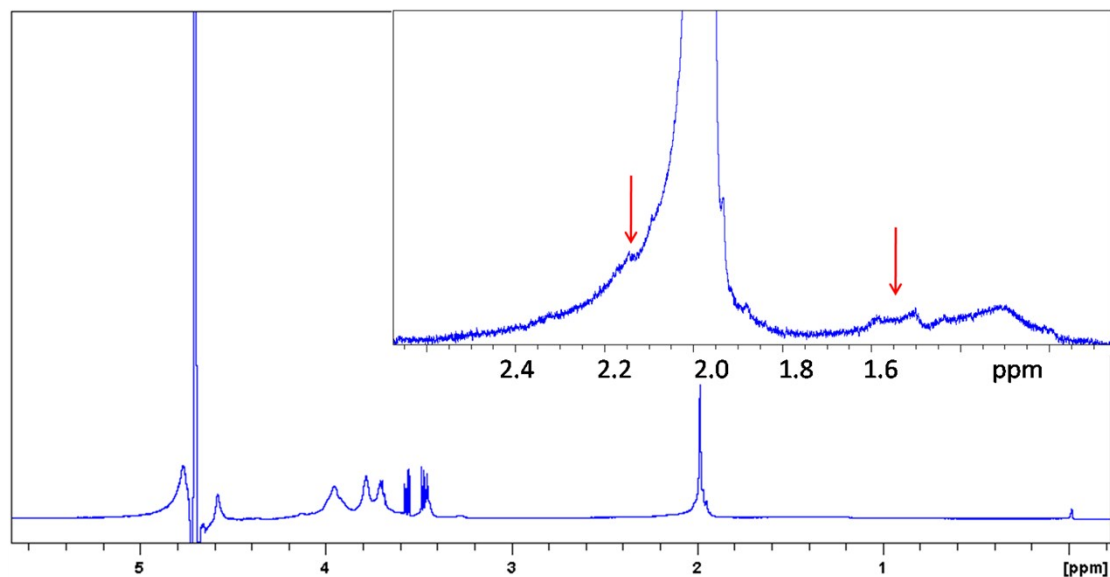


Figure S10. ^1H NMR spectra of dermatan sulfate-ADH. Two sets of peaks at 1.51 and 2.16 ppm in the ^1H NMR spectra (insert) correspond to dihydrazide adipic linker.

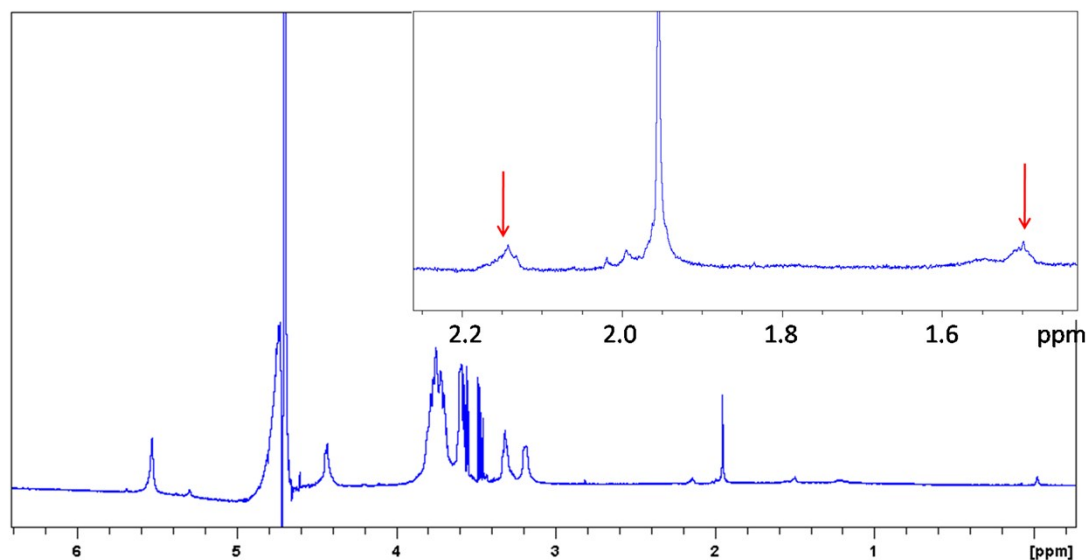


Figure S11. ^1H NMR spectra of *N*-desulfated heparin-ADH. Two sets of peaks at 1.51 and 2.16 ppm in the ^1H NMR spectra (insert) correspond to dihydrazide adipic linker.

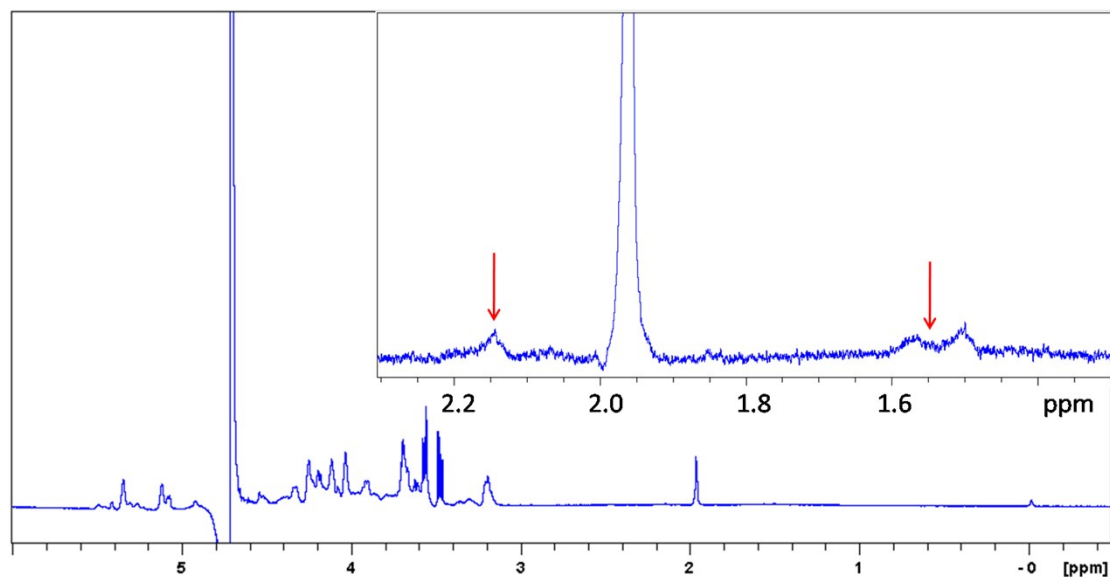


Figure S12. ^1H NMR spectra of LMW heparin-ADH. Two sets of peaks at 1.51 and 2.16 ppm in the ^1H NMR spectra (insert) correspond to dihydrazide adipic linker.

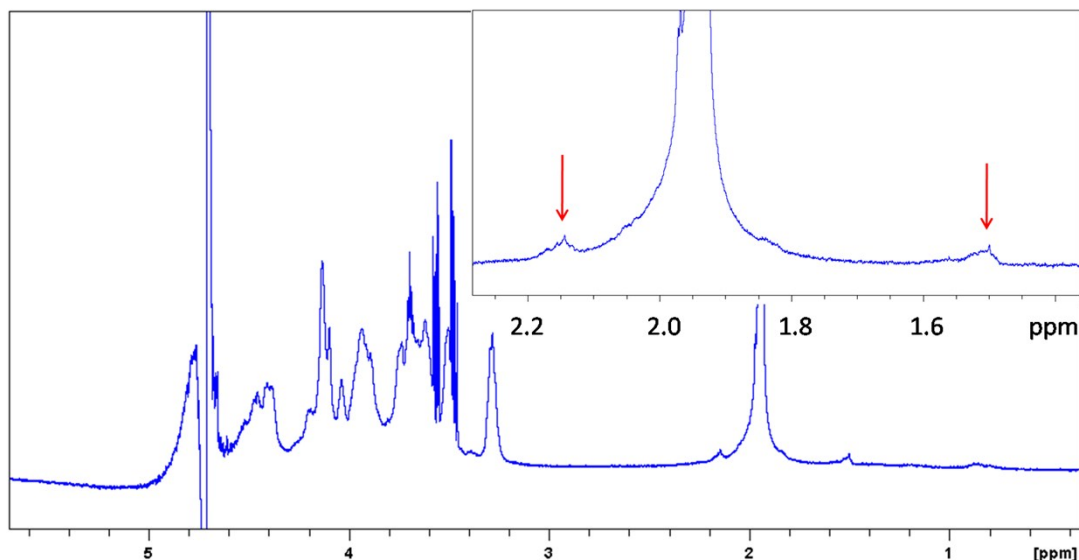


Figure S13. ^1H NMR spectra of chondroitin sulfate E-ADH. Two sets of peaks at 1.51 and 2.16 ppm in the ^1H NMR spectra (insert) correspond to dihydrazide adipic linker.

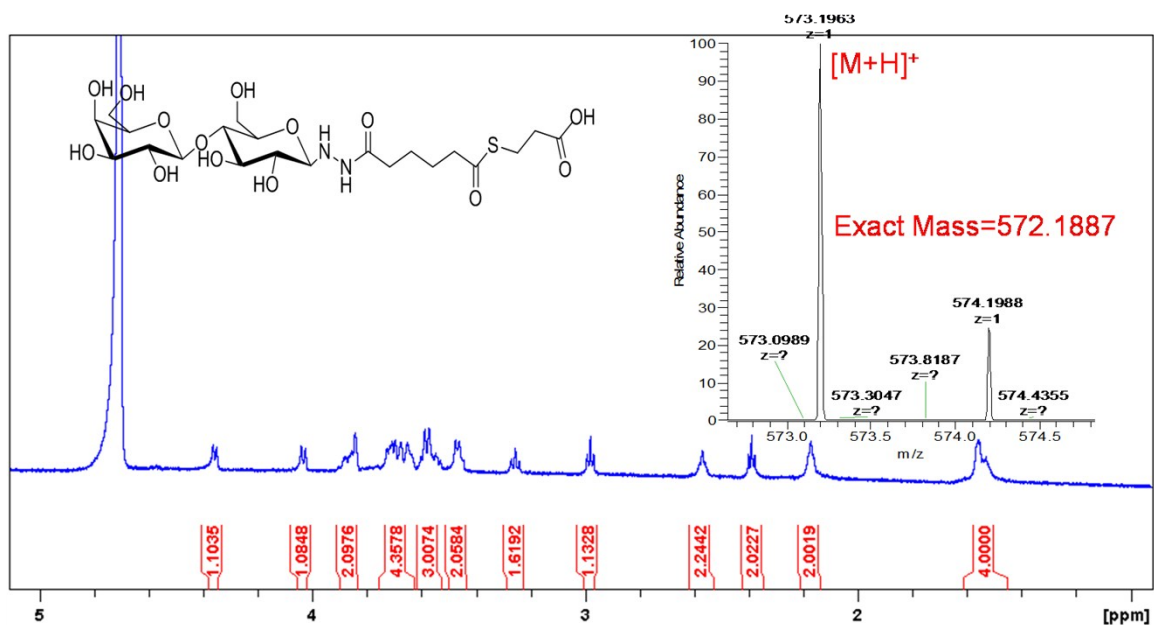


Figure S14. ^1H NMR spectrum and HRMS (insert, positive mode) of lactose-mercaptpropionic acid. White solid, ^1H NMR (600 MHz, D_2O) δ : 4.35 (d, $J = 7.9$, 1H), 4.04 (d, $J = 8.68$, 1H), 3.90-3.83 (m, 2H), 3.74-3.62 (m, 4H), 3.62-3.52 (m, 3H), 3.50-3.44 (m, 2H), 3.62 (t, $J = 8.68$, 1H), 2.98 (t, $J = 7.02$, 2H), 2.60-2.55 (m, 2H), 2.40 (t, $J = 7.21$, 2H), 2.20-2.14 (m, 2H), 1.59-1.49 (m, 4H).

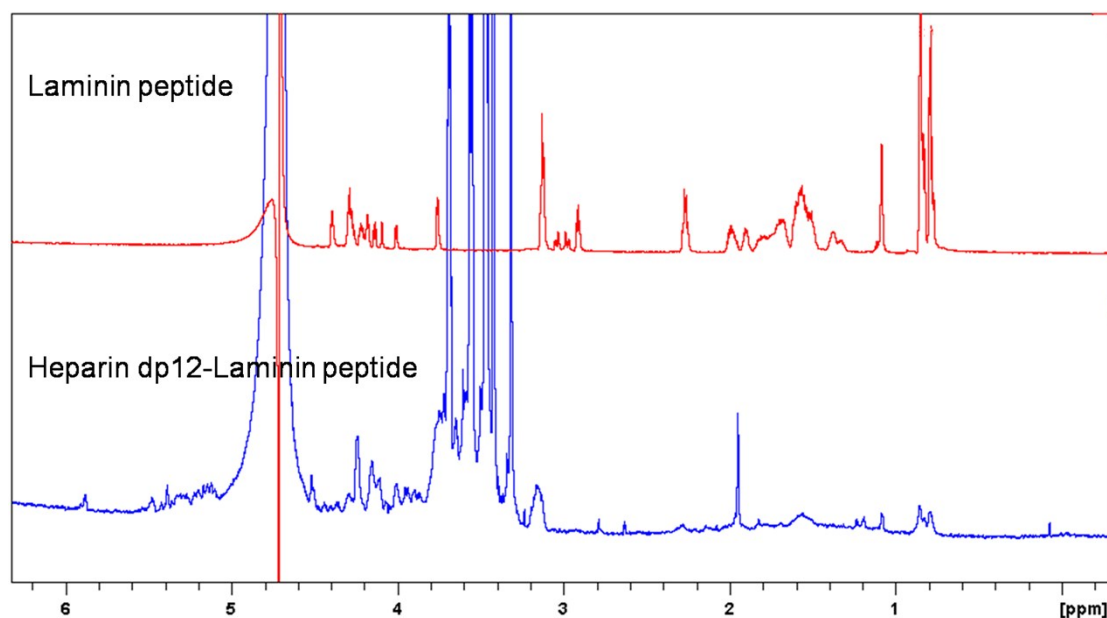


Figure S15. ^1H NMR spectra comparison of laminin peptide and heparin dp12-laminin peptide. Peaks between 3.4–3.7 ppm in the ^1H NMR spectrum of heparin dp12-laminin peptide were from glycerol that came from the size exclusion spin column membrane.
Laminin peptide: CRKRLQVQLSIRT

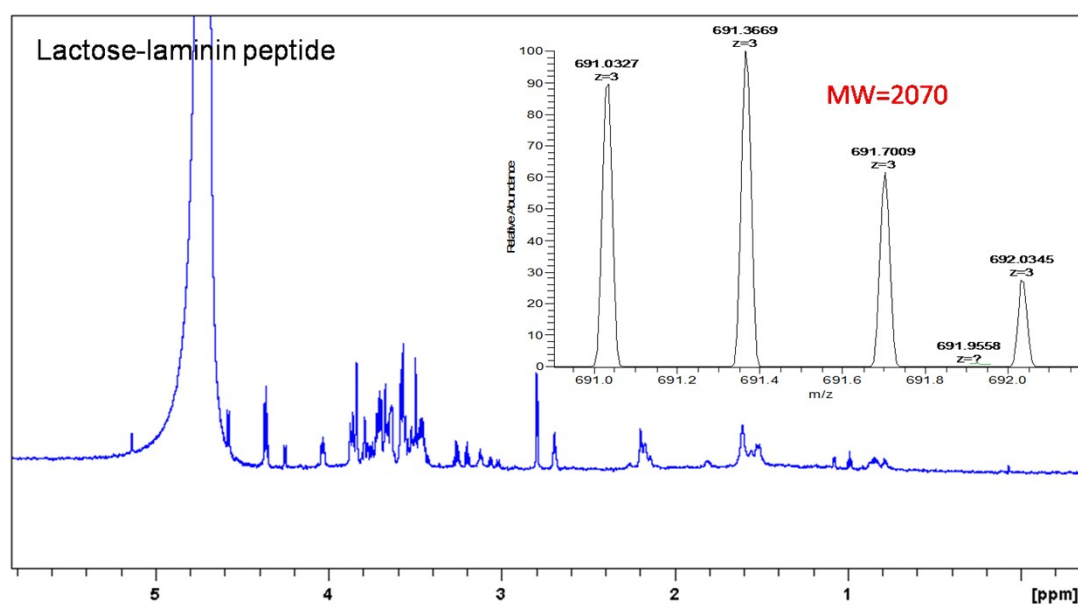


Figure S16. ^1H NMR spectra of lactose-laminin peptide. ESI MS: calcd. MW = 2070; found (m/z): 691.0327 $[\text{M}+3\text{H}]^{3+}$.

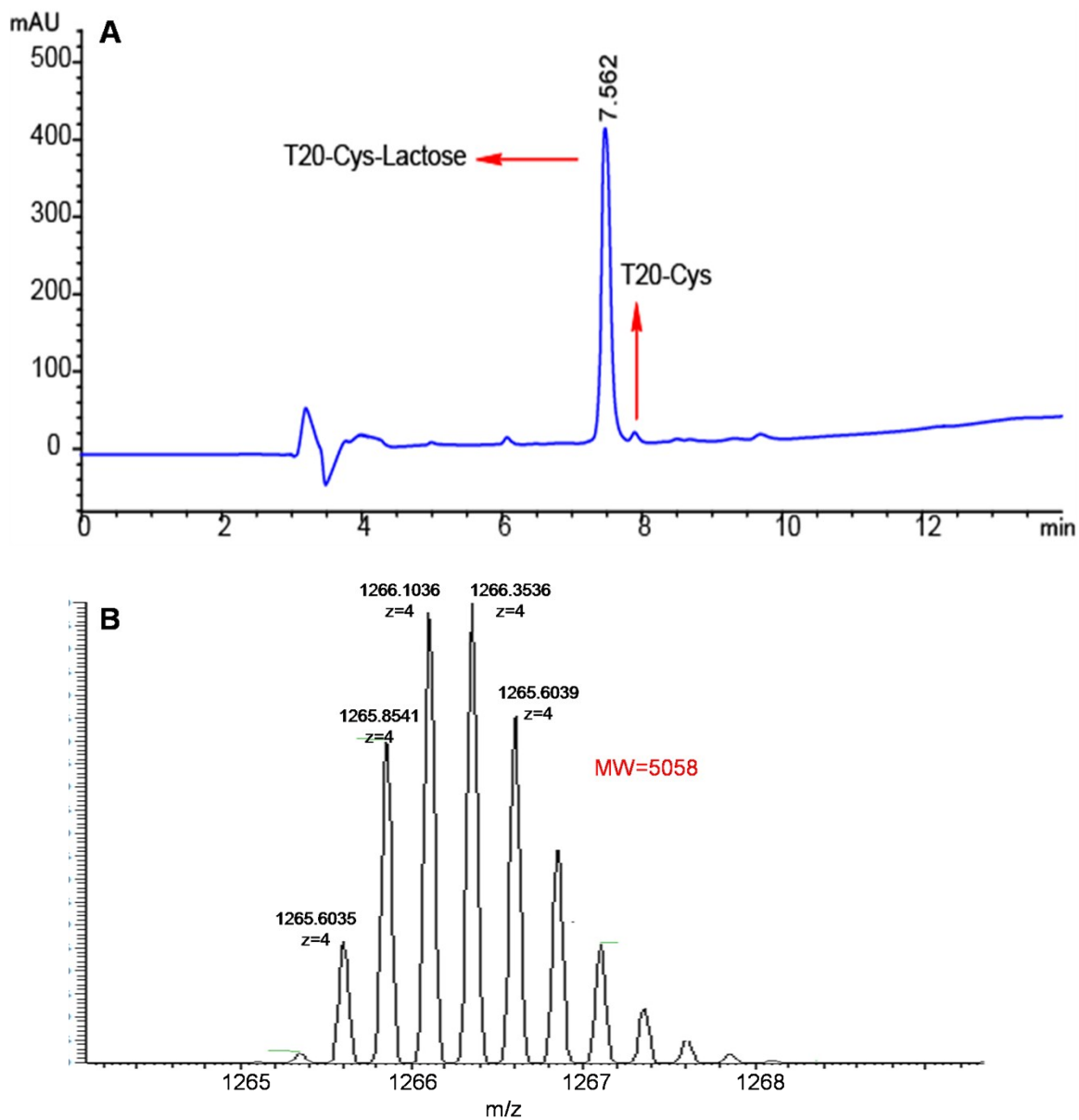


Figure S17. HPLC (A) and HRMS (B) analysis (positive mode) of T20C-lactose. HPLC chromatographic conditions: Phenomenex Synergi Hydro-RP C18 column (4 μ m, 10 \times 250 mm), 40 \rightarrow 70% CH₃CN/H₂O linear gradient over 11 min at 3 mL/min, UV detection at 220 nm. ESI MS: calcd. MW = 5058; found (m/z): 1265.6035 [M+4H]⁴⁺. T20C peptide: Ac-YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFC

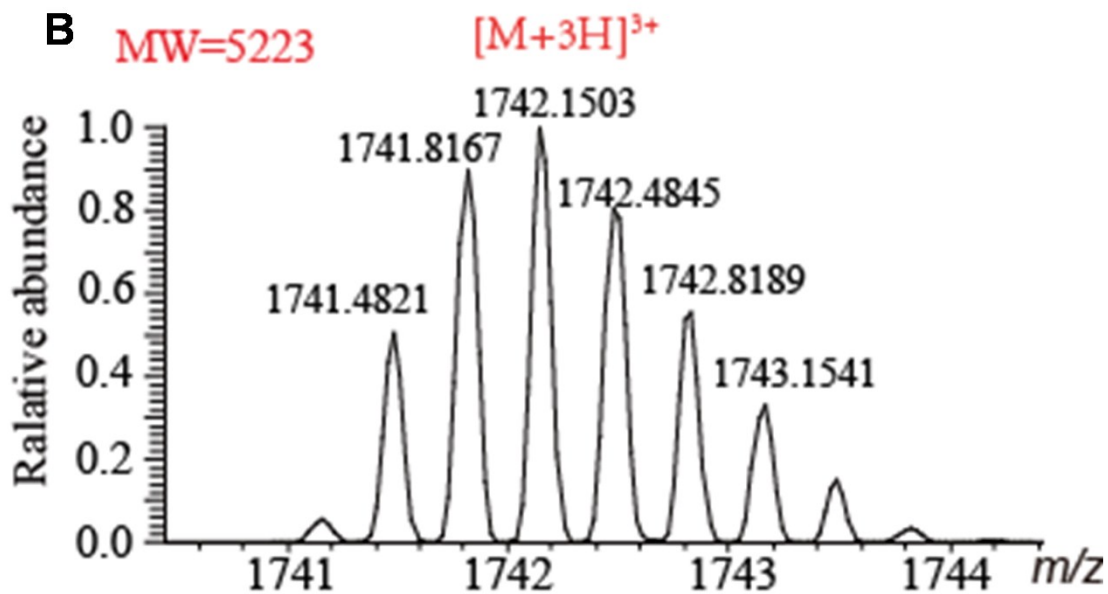
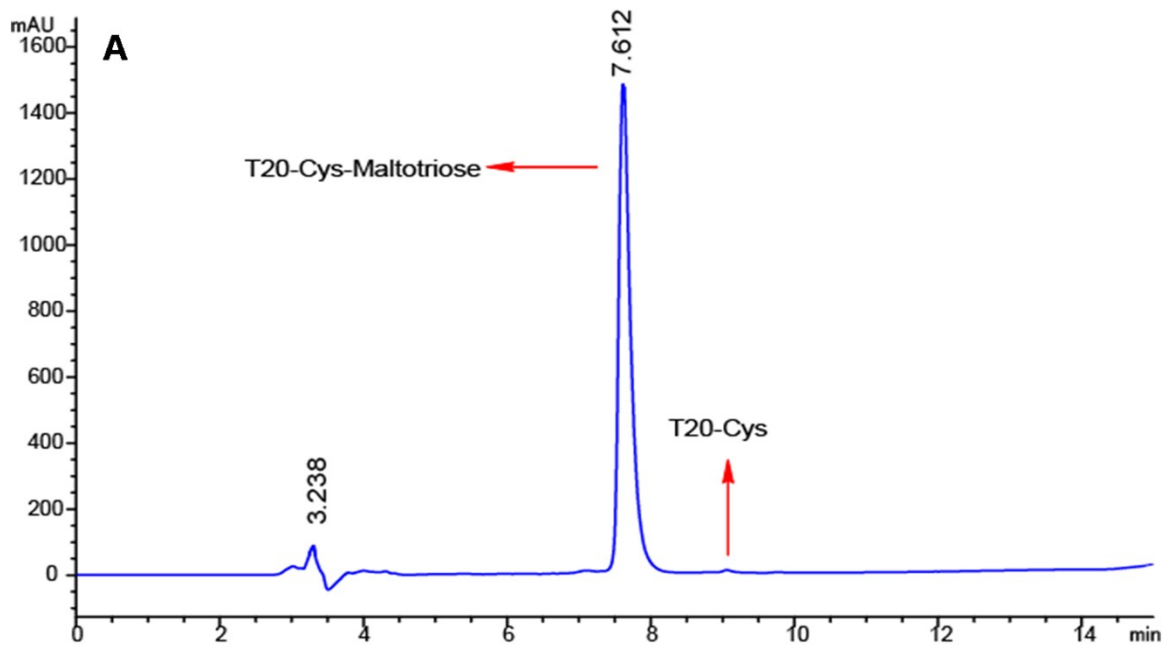


Figure S18. HPLC (A) and HRMS (B) analysis (positive mode) of T20C-maltotriose. HPLC chromatographic conditions: Phenomenex Synergi Hydro-RP C18 column (4 μ m, 10 \times 250 mm), 40 \rightarrow 70% CH₃CN/H₂O linear gradient over 11 min at 2.5 mL/min, UV detection at 220 nm. ESI MS: calcd. MW = 5223; found (m/z): 1741.4821 [M+3H]³⁺. T20C peptide: Ac-YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFC

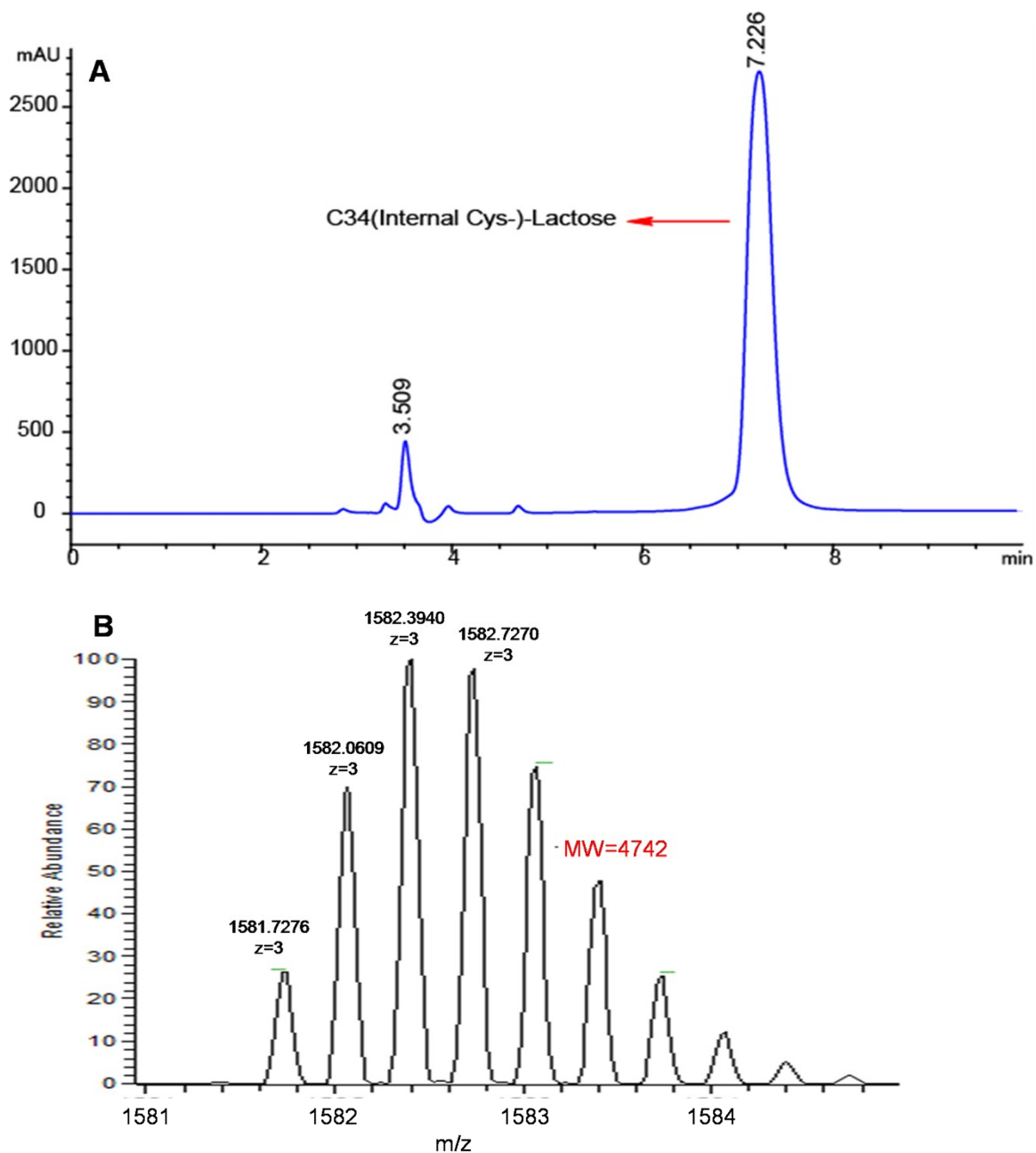


Figure S19. HPLC (A) and HRMS (B) analysis (positive mode) of C34NC-lactose. HPLC chromatographic conditions: Phenomenex Synergi Hydro-RP C18 column (4 μ m, 10 \times 250 mm), 40 \rightarrow 70% CH₃CN/H₂O linear gradient over 11 min at 2.5 mL/min, UV detection at 220 nm. ESI MS: calcd. MW = 4742; found (m/z): 1581.7276 [M+3H]³⁺. C34NC peptide: Ac-WMEWDREINCYTSLIHSLEESQNQQEKNEQELL

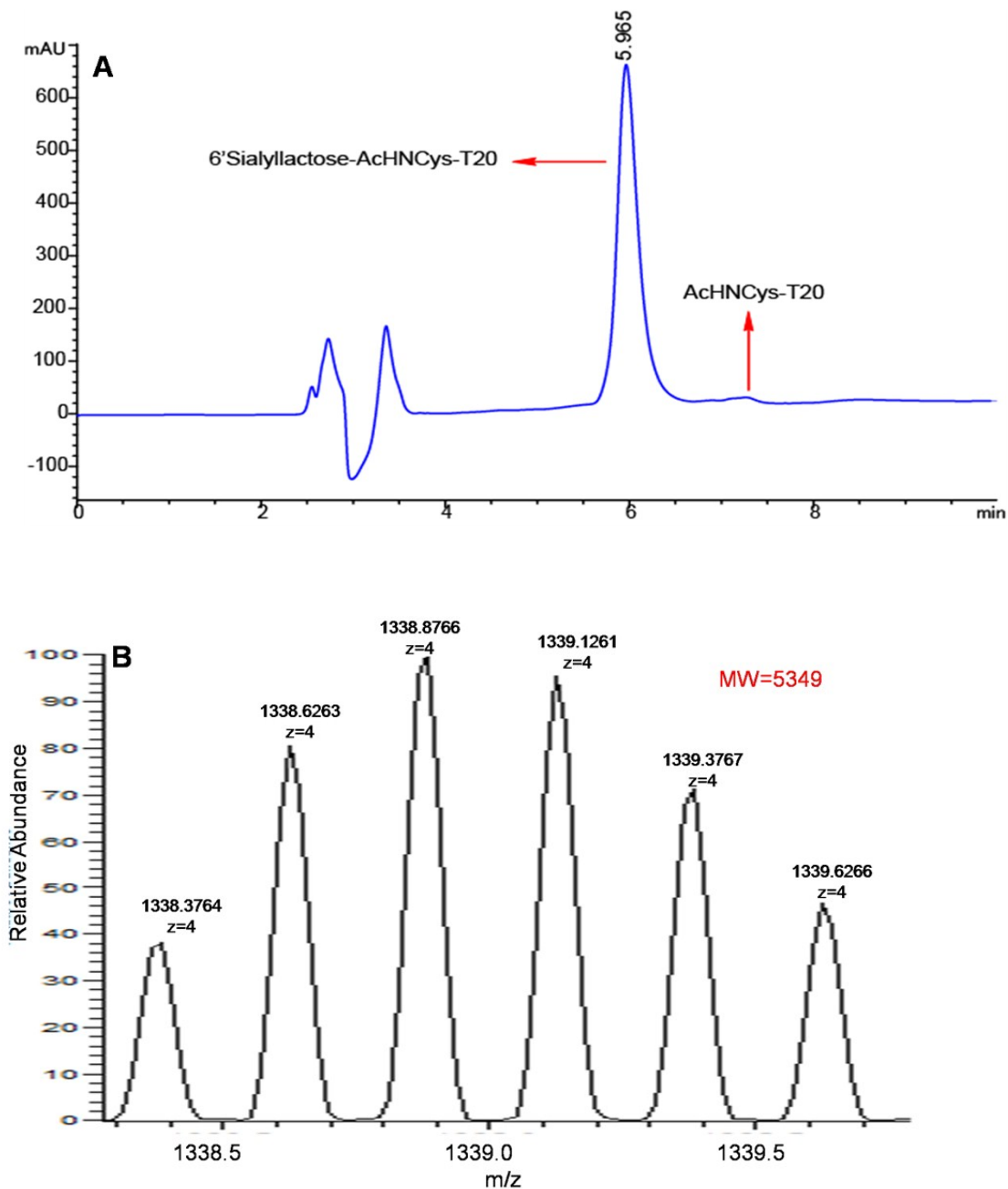


Figure S20. HPLC (A) and HRMS (B) analysis (positive mode) of 6'-Sialyllactose-CT20. HPLC chromatographic conditions: Phenomenex Synergi Hydro-RP C18 column (4 μ m, 10 \times 250 mm), 40 \rightarrow 70% CH₃CN/H₂O linear gradient over 11 min at 3 mL/min, UV detection at 220 nm. ESI MS: calcd. MW = 5349; found (m/z): 1338.3764 [M+4H]⁴⁺. CT20 peptide: Ac-CYTSLIHSLIEESQNNQKEKNEQELLELDKWASLWNWF

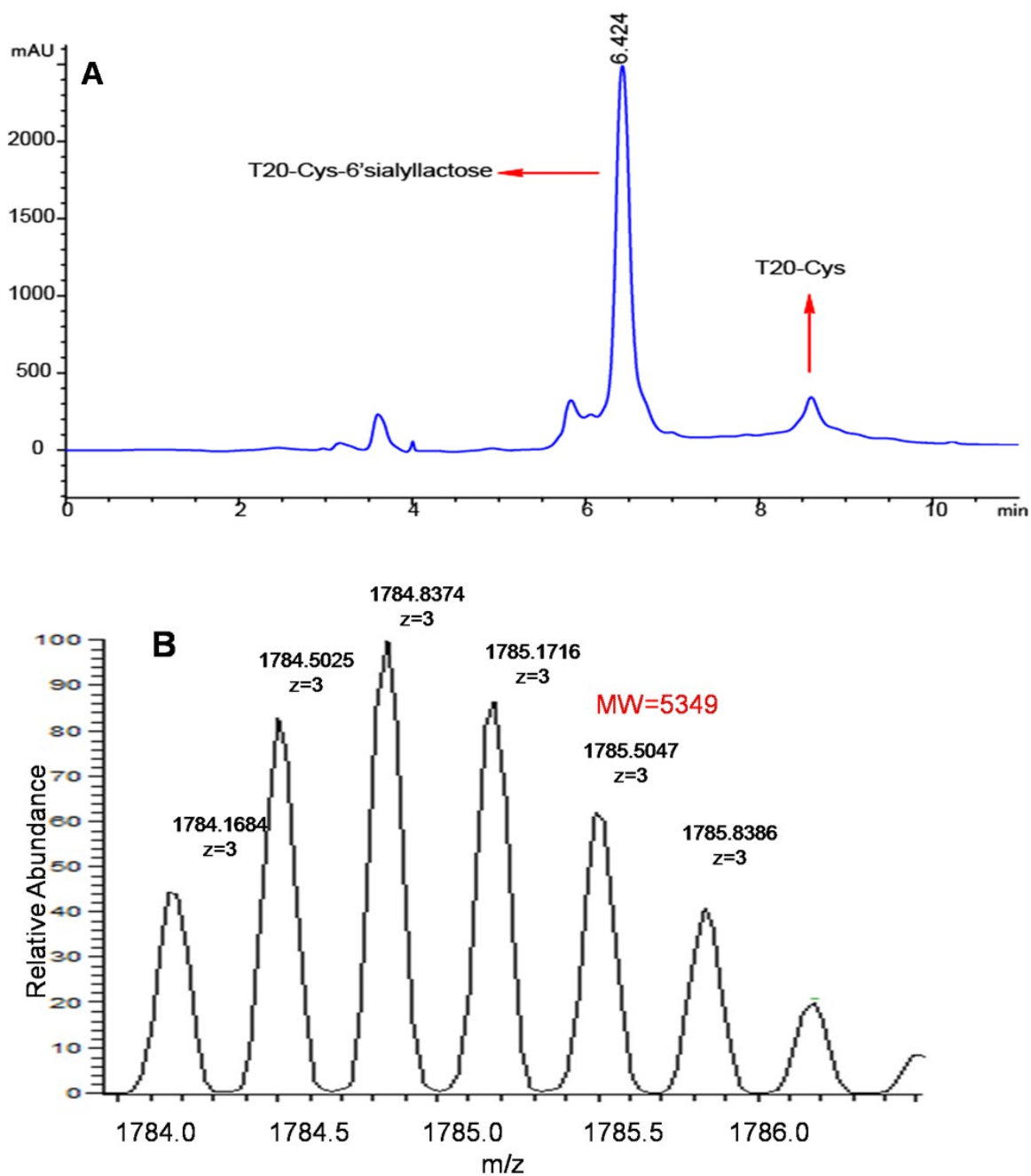


Figure S21. HPLC (A) and HRMS (B) analysis (positive mode) of T20C-6'-Sialyllactose. HPLC chromatographic conditions: Phenomenex Synergi Hydro-RP C18 column (4 μ m, 10 \times 250 mm), 40 \rightarrow 70% CH₃CN/H₂O linear gradient over 11 min at 2.5 mL/min, UV detection at 220 nm. ESI MS: calcd. MW = 5349; found (m/z): 1784.1684 [M+3H]³⁺. T20C peptide: Ac-YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFC

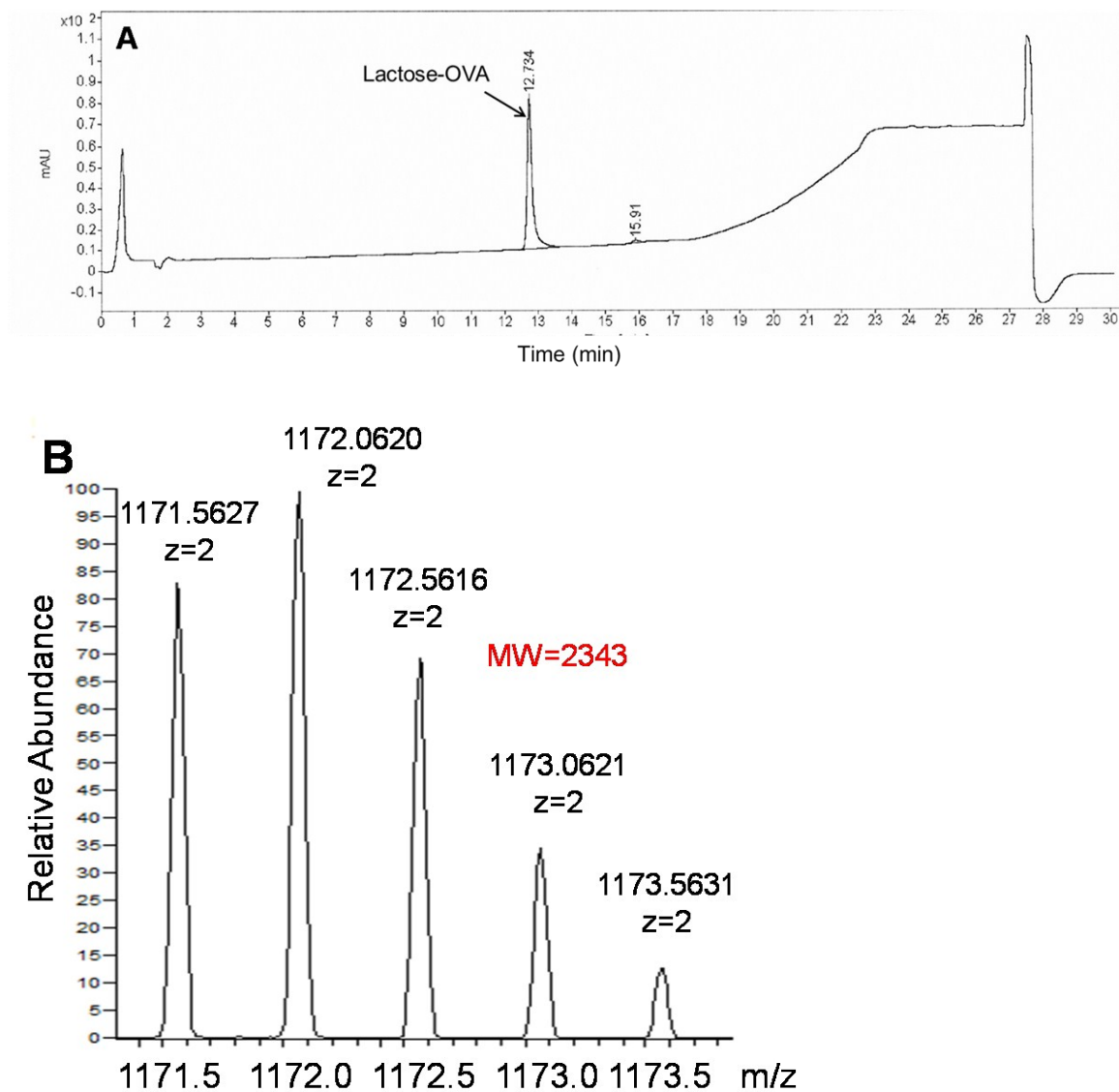


Figure S22. HPLC (A) and HRMS (B) analysis (positive mode) of lactose-OVA. HPLC chromatographic conditions: YMC-TriartC18 column (4.6 × 250 mm), 5→80% CH₃CN/H₂O (containing 0.075% trifluoroacetic acid), linear gradient over 30 min at 0.8 mL/min, UV detection at 220 nm. ESI MS: calcd. MW = 2343; found (m/z): 1171.5627 [M+2H]²⁺. OVA peptide: CISQAVHAAHAEINEAGR

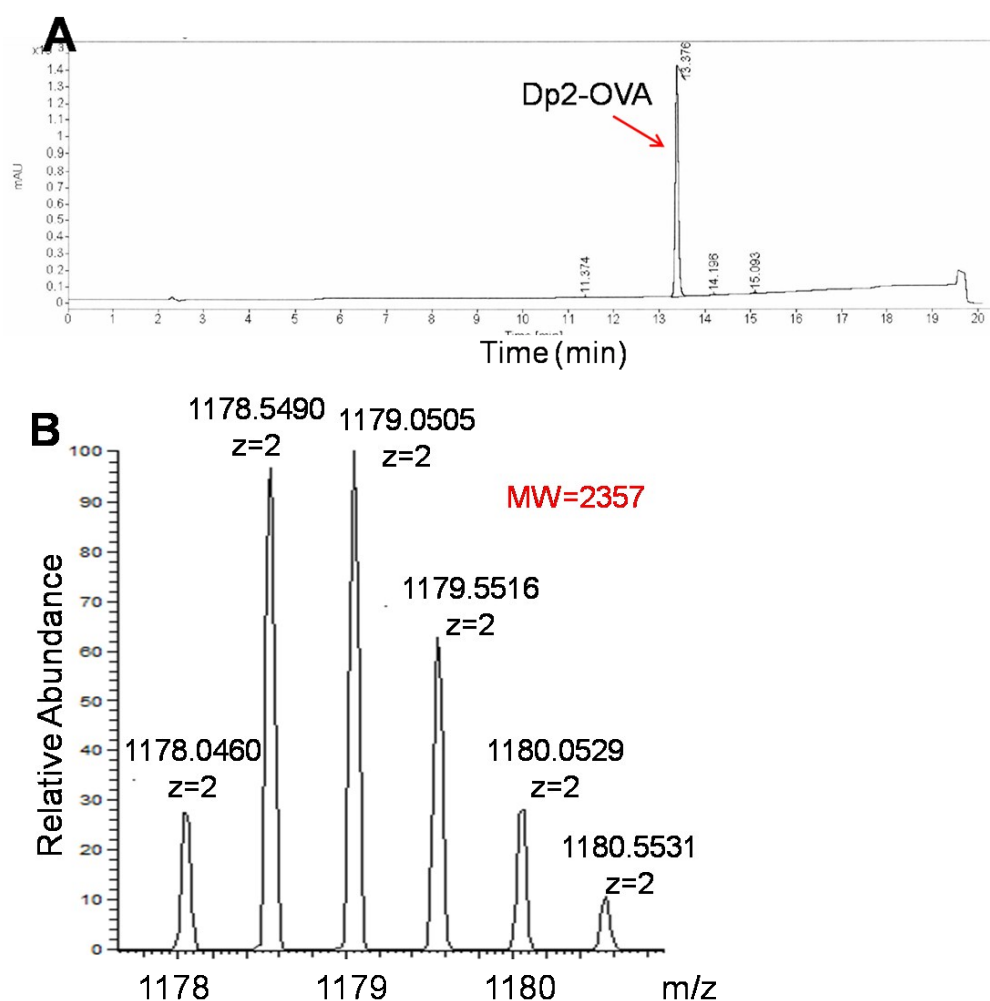


Figure S23. HPLC (A) and HRMS (B) analysis (positive mode) of dp2-OVA. HPLC chromatographic conditions: YMC-TriartC18 column (4.6 × 250 mm), 5→80% CH₃CN/H₂O (containing 0.075% trifluoroacetic acid), linear gradient over 30 min at 0.8 mL/min, UV detection at 220 nm. ESI MS: calcd. MW = 2357; found (m/z): 1178.0460 [M+2H]²⁺.

OVA peptide: CISQAVHAAHAEINEAGR

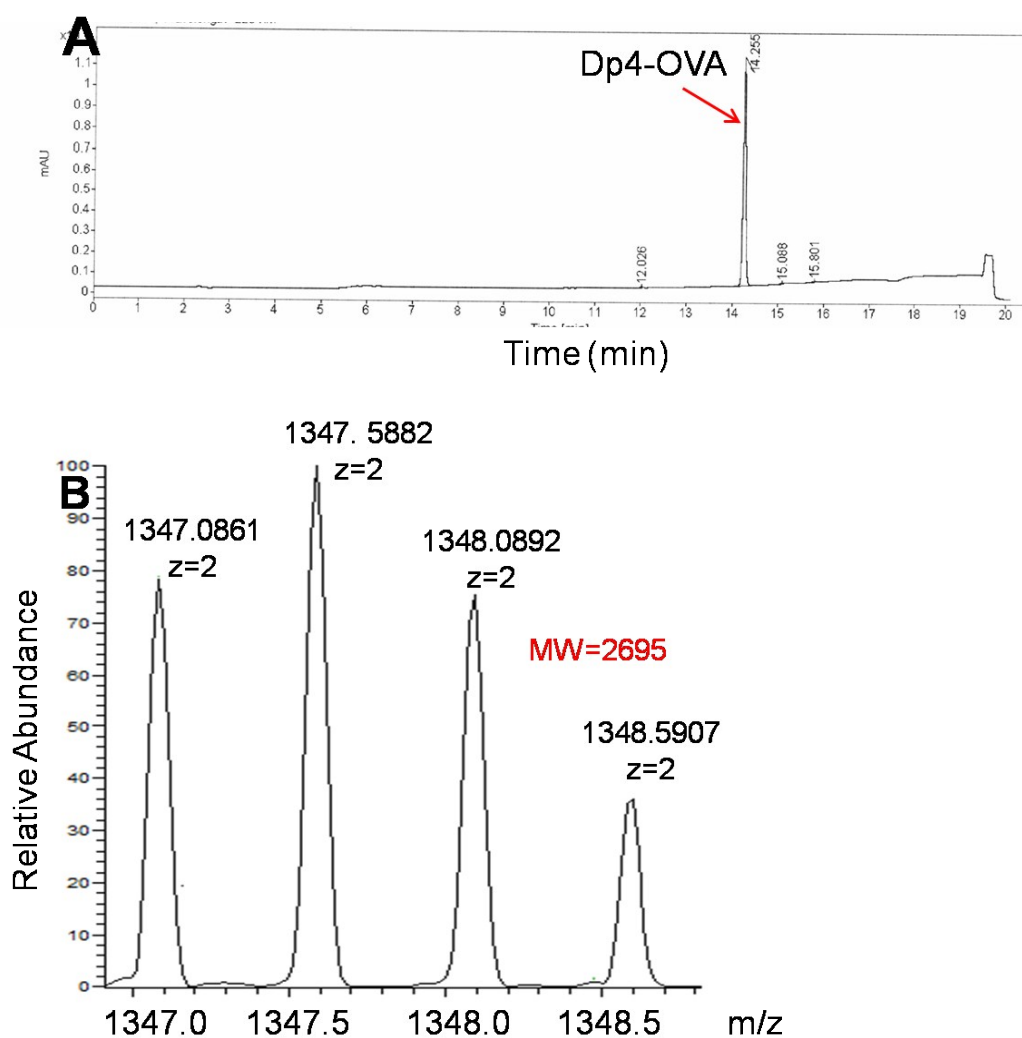


Figure S24. HPLC (A) and HRMS (B) analysis (positive mode) of dp4-OVA. HPLC chromatographic conditions: YMC-TriartC18 column (4.6 × 250 mm), 5→80% CH₃CN/H₂O (containing 0.075% trifluoroacetic acid), linear gradient over 30 min at 0.8 mL/min, UV detection at 220 nm. ESI MS: calcd. MW = 2695; found (m/z): 1347.0861 [M+2H]²⁺.

OVA peptide: CISQAVHAAHAEINEAGR

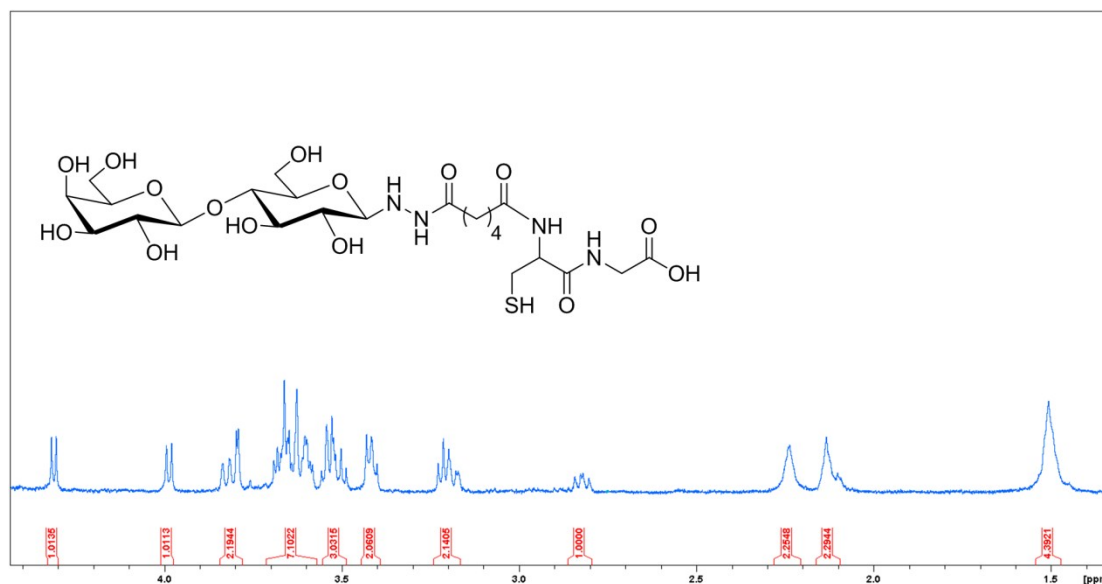


Figure S25. ¹H NMR spectrum of lactose-dipeptide (Cys-Gly). White solid. ¹H NMR (600 MHz, D₂O) δ : 4.35 (d, $J = 7.54$, 1H), 4.05 (d, $J = 8.84$, 1H), 3.90-3.83 (m, 2H), 3.74-3.62 (m, 7H), 3.61-3.53 (m, 3H), 3.49-3.44 (m, 2H), 3.28-3.21 (m, 2H), 2.90-2.84 (m, 1H), 2.32-2.25 (m, 2H), 2.20-2.12 (m, 2H), 1.62-1.50 (m, 4H).