Supplementary information

A novel enzymatic method for synthesis of glycopeptides carrying natural eukaryotic *N*-glycans

Yangyang Xu,[a#] Zhigang Wu,[b#] Peiru Zhang,[a] He Zhu,[b] Hailiang Zhu,[b] Qitao Song,[a] Li Wang,[a] Faxing Wang,[a] Peng George Wang[a*] and Jiansong Cheng[a*]

[a.] State Key Laboratory of Medicinal Chemical Biology and College of Pharmacy, Nankai University, Haihe Education Park, 38 Tongyan Road, Tianjin 300353, P. R. China.
[b.] Department of Chemistry, Georgia State University, Atlanta, Georgia 30303, United States.

Experimental procedure

General Information

Ac-CoA (>93%, CAS:102029-73-2) was purchased from Sigma-Aldrich (St. Louis, MO, USA). UDP-GlcNAc (98%, CAS:91183-98-1) and UDP-GalNAc (95%, CAS:108320-87-2) were purchased from J&K (Beijing, China). ^{1}H NMR spectra was recorded on a Bruker AV 400 MHz spectrometer at 400 MHz. High resolution mass spectra were performed on a 6520Q-TOF LC/MS instrument (Agilent) with ES ionization. MALDI-TOF MS were obtained with an Autoflex Speed mass spectrometer (Bruker) or a Varian 7.0T FTMS instrument using 2,5-dihydroxybenzoic acid (DHB) as matrix. Nano UHPLC-MS/MS assays were obtained on an LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher). Reverse-phase HPLC was performed on a Shimadzu SPD-20A instrument using a Shim-pack VP-ODS C18 column (5 μm, 4.6 mm × 250 mm).

Synthesis of UDP-GlcN, UDP-$GlcN_3$ and UDP-GalN

The method for synthesis of UDP-GlcN, UDP-$GlcN_3$ and UDP-GalN were described previously.[1, 2] For UDP-GlcN, reaction was monitored by Thin Layer Chromatography (TLC) on silica gel F254 (Merck), developed by isopropanol/NH_4OH=3:2(V/V), then stained with p-anisaldehyde solution (ethanol/sulfuric acid/p-anisaldehyde/aceticacid = 500:27:16:5.5, V/V/V/V). The crude products were purified by silica gel chromatography (isopropanol/

NH_4OH=3:1) and Bio-Gel P2 size exclusion chromatography to obtain pure UDP-GlcN (500 mg, 53%). ^1H NMR (400 MHz, D_2O) δ 7.88 (d, J = 8.1 Hz, 1H), 5.94–5.85 (m, 2H), 5.79 (dd, J = 6.9, 3.4 Hz, 1H), 4.32–4.27 (m, 2H), 4.24–4.10 (m, 3H), 3.93–3.82 (m, 2H), 3.82–3.71 (m, 2H), 3.50 (t, J = 9.6 Hz, 1H), 3.34 (dt, J = 10.8, 3.1 Hz, 1H). HRMS (ESI): calculated for $C_{15}H_{26}N_3O_{16}P_2$: 566.0788, found (*m/z*): $[M+H]^+$ = 566.0777.

Protein expression and purification

The *N*-Glycosyltransferase $ApNGT^{Q469A}$ and the glucosamine/glucosaminide *N*-acetyltransferase GlmA were cloned into pET22b and pET15b, respectively. The *Escherichia coli* BL21 (DE3) strains harboring the corresponding expression plasmids, were cultured in LB medium supplemented with ampicillin (50 μg mL^{-1}) under energetic sharking. When the optical density at 600 nm (OD_{600}) reached 0.8, 0.1 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) was supplied and followed by incubating at 16°C for 18 h. The cells were harvested by centrifugation (6,000×g, 4°C, 20 mins), and resuspended in lysis buffer (100 mM Tris-HCl, 5 mM $CaCl_2$, pH 8.0), and broken by ultrasonication. After centrifugation (14,000×g, 20 mins) of the cell lysate, the supernatant was loaded onto a Ni^{2+}-NTA affinity column (Qiagen). The column was pre-equilibrated with 10 column volumes of binding buffer (10 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 8.0) before the lysate was loaded. After washing with 10 column volumes of binding buffer and washing buffer (30 mM imidazole, 0.5 M NaCl, 5 mM $CaCl_2$, 20 mM Tris-HCl, pH 8.0), the target protein was eluted with elution buffer (250 mM imidazole, 0.5 M NaCl, 5 mM $CaCl_2$, 20 mM Tris-HCl, pH 8.0). Elution fractions were analyzed by 12% SDS-PAGE with Coomassie staining. The concentration of purified enzyme was quantified by the BCA Protein Assay Kit (Pierce Biotechnology, Rockford, IL) with bovine serum albumin as a protein standard.

Enzymatic reactions

(1) Synthesis of GlcN-peptides

Production of GlcN-peptides were carried out in the volume of 20 ml contains 20 μM ApNGT,10 mM UDP-GlcN (20 mM UDP-GlcN for peptide P15), 1 mM various peptides and with 100 mM Tris-HCl (pH8.0) at 30°C for overnight with a slight shaker. Reactions were stopped by boiling 5 mins and then analyzed by Shim-pack VP-ODS C18 column (5 μm, 4.6 mm × 250 mm) and purified through HYPERSIL GOLD C18 column (10 μm, 10 mm × 250 mm). The purified GlcN-peptides were further identified by HPLC and MALDI-TOF MS.

(2) Synthesis of GlcNAc-peptides

For synthesis of GlcNAc-peptides, 10 μM GlmA, 2 mM Ac-CoA (4 mM Ac-CoA for peptide P16),1 mM various GlcN-peptides with 100 mM Tris-HCl (pH7.5) were mixed at 37°C for 2

hours. Reactions were stopped by boiling 5 mins and then analyzed and purified by HPLC. The purified GlcNAc-peptides were further identified by MALDI-TOF MS and ESI-MS-MS analysis. MALDI-TOF: calculated M=1239.598; found (*m/z*): $[M+Na]^+$=1262.028. The purified GGAN(GlcNAc)LTYTIE was identified by ^{1}H-NMR. ^{1}H-NMR (400 MHz, D_2O, selected signals): δ 6.99 (d, *J* = 8.52 Hz, 2H, 2×Tyr Aromatic), 6.69 (d, *J* = 8.52 Hz, 2H, 2×Tyr Aromatic), 4.93 (d, *J* = 9.80 Hz, 1H, H-1 *β*-GlcNAc), 4.56 (q, *J* =7.55 Hz, 2H), 4.27-4.19 (m, 4H), 4.16 (d, *J* = 4.96 Hz, 1H), 4.05-3.95 (m, 3H), 3.90 (d, *J* = 2.56 Hz, 2H), 3.7744 (s, 2H), 3.75-3.74 (m, 2H), 3.64-3.60 (m, 1H), 3.50-3.46 (m, 1H), 3.36 (d, *J* = 5.64 Hz, 2H), 2.94 (dd, *J* = 6.36 Hz, *J* = 12.24 Hz, 1H), 2.86-2.74 (m, 2H), 2.62 (dd, *J* = 7.56 Hz, *J* = 15.9 Hz, 1H), 2.36-2.32 (m, 2H), 2.04-1.96 (m, 1H), 1.89 (s, 3H, 3×Ac GlcNAc), 1.86-1.81 (m, 1H), 1.79-1.72 (m, 1H), 1.54-1.34 (m, 4H), 1.26 (d, *J* = 7.0 Hz, 3H), 1.12-1.06 (m, 1H), 1.02 (d, *J* = 6.32 Hz, 6H), 0.81-0.73 (m, 12H).

(3) Synthesis of N-Glycan-peptides

The reaction mixtures contained 50 mM PBS (pH 6.5), 5 nM GlcNAc-peptide as acceptor, 20 nM sialylated complex-type glycans oxazoline (SCT-oxa) as donor, and 20 μg Endo M^{N175Q}, and was incubated at 30°C for 15 mins. After optimization of reaction condition for $EndoM^{N175Q}$, we chose to carry out 15 mins incubation for peptide transglycosylation. The reaction was terminated by boiling 5 mins and then analyzed by HPLC-C18 and ESI-MS.

Method of Electrospray ionization tandem mass spectrometry

GlcNAc-peptides were analyzed using Nano RP-HPLC-MS system as described before.[3] Briefly, samples were separated on EASY-Spray PepMap C18 Column (75 μm id × 15 cm, 3 μm) with an LTQ-Orbitrap Elite mass spectrometer using a linear gradient from 3% to 40% buffer B for 40 mins at a flow rate of 300 nL/min (mobile phase A: 1.95% ACN, 97.95% H_2O, 0.1% FA; mobile phase B: 79.95% ACN, 19.95% H_2O, 0.1% FA). Data-dependent mode was applied during the acquisition. A full-scan survey MS (*m/z* range from 400 to 2000; automatic gain control target, 1,000,000 ions; resolution at 200 *m/z*, 60,000; maximum ion accumulation time, 50 ms) was acquired by the spectrometer, and ten most intense ions were selected for HCD fragmentation with a collision energy of 27 ev.

Circular dichroism (CD) spectroscopy

Far-UV (190–260 nm) CD spectra was recorded on a J-810 spectropolarimeter (Jasco) at ambient temperature (25°C) using a 0.1-cm-pathlength quartz cuvette. The measurements of 50 μM mGLP-1 and GlcNAc-mGLP-1 were carried out in 10 mM PBS Buffer (pH 6.5). All spectra presented were averaged for 3 scans.

Table S1. Apparent kinetic parameters of ApNGT^{Q469A} and GlmA

Enzymes	Substrates	K_m (μM)	V_{max} (μM·min^{-1})	k_{cat} (S^{-1})	k_{cat}/K_m (M^{-1}·S^{-1})
Q469A	GGANLTYTIER	$(1.7\pm0.2)\times10^3$	3.8±0.2	6.3×10^{-3}	3.7
	UDP-GlcN	$(5.0\pm0.3)\times10^3$	2.1±0.1	3.5×10^{-3}	0.7
GlmA	GGAN(GlcN)LTYTIER	$(9.8\pm0.9)\times10^2$	11.4±0.5	9.5×10^{-2}	96.9
	Ac-CoA	19.9±2.9	2.6±0.1	2.2×10^{-2}	1.1×10^3

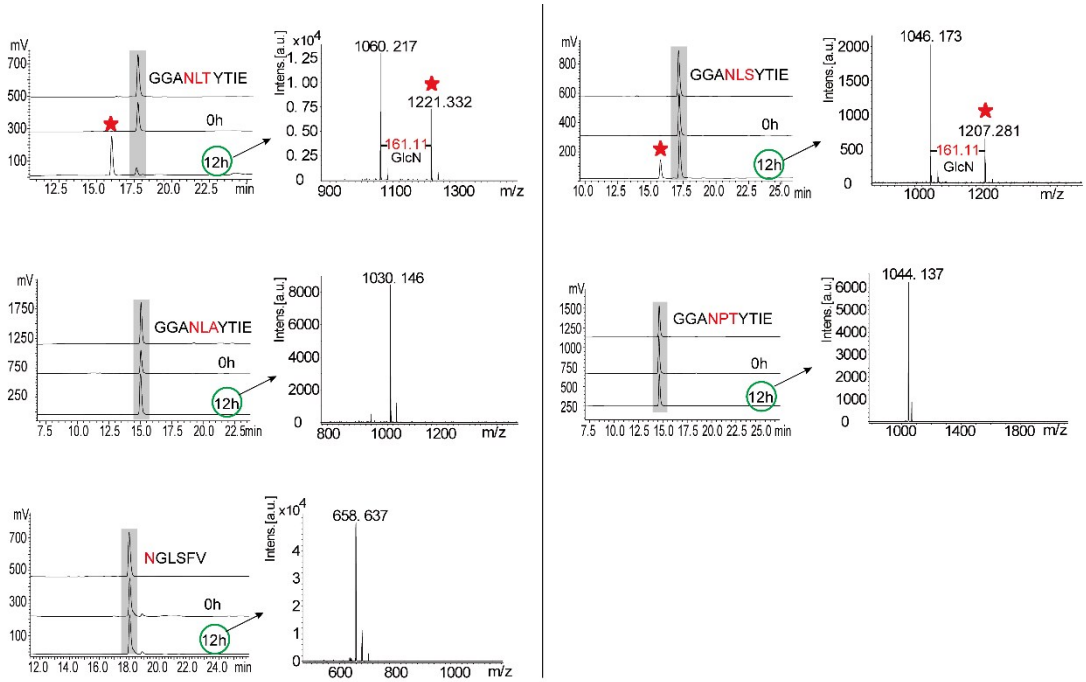


Fig. S1 The specificity of ApNGT^{Q469A} toward peptides. The reactions catalyzed by ApNGT^{Q469A} for 12 h were monitored by HPLC. The glycosylated peptides were confirmed by MALDI-TOF MS and labeled with red stars.

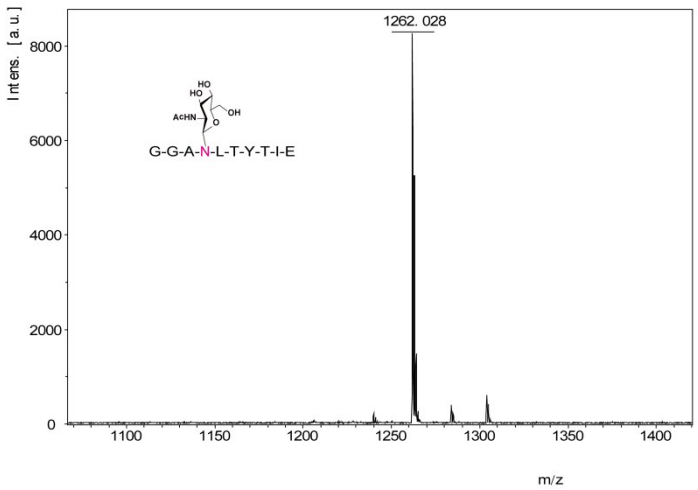
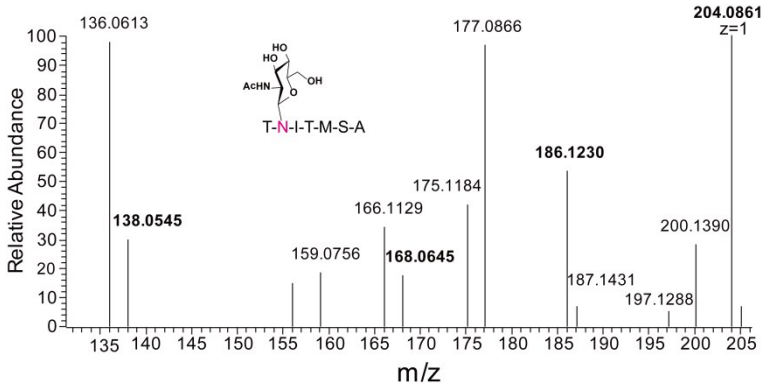


Fig. S2. MALDI-TOF MS profile of GGAN(GlcNAc)LTYTIE, calculated M=1239.598; found(*m/z*): $[M+Na]^+$ = 1262.028.

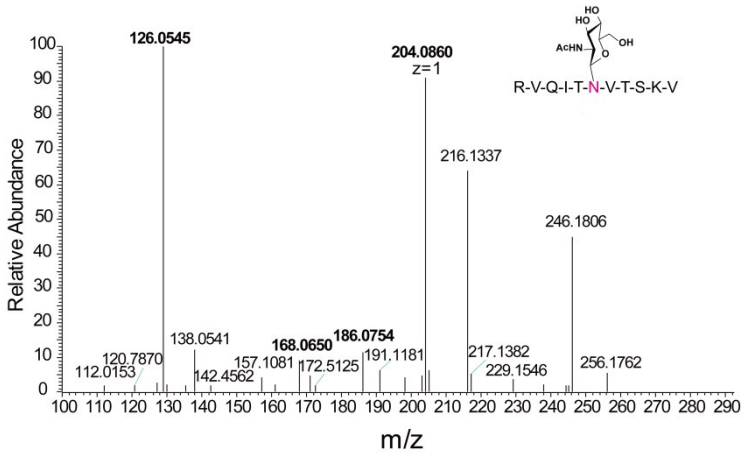
Fig. S3. HCD-MS2 spectra of glycopeptides GlcNAc-P11~GlcNAc-P19

GlcNAc-peptides were confirmed by the HCD spectrums with a series of diagnosis fragmentation ions for GlcNAc including 126.05, 138.05, 144.06, 168.06, 186.07 and 204.08.

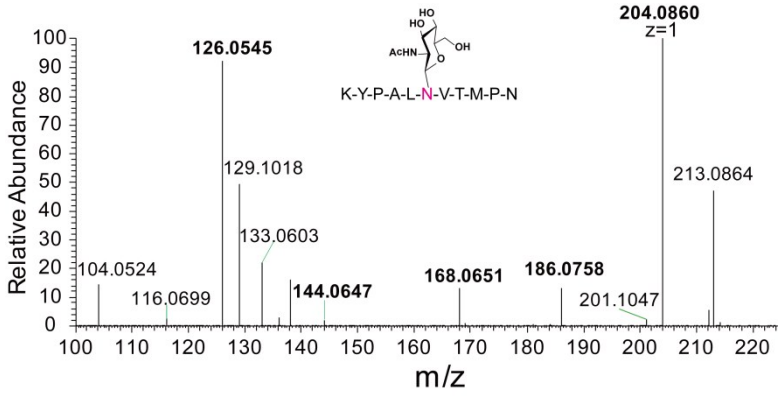
GlcNAc-P11:



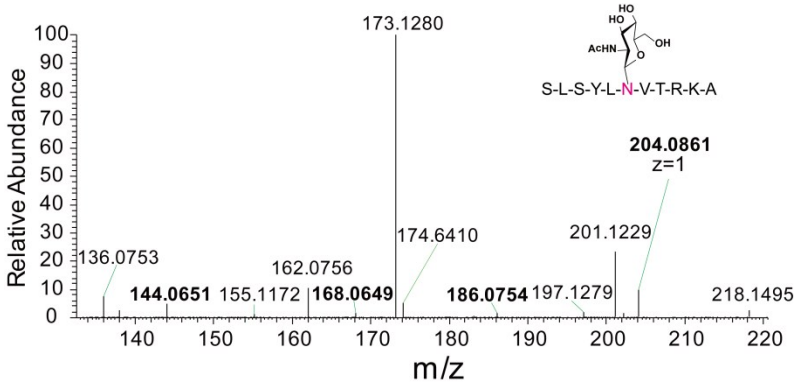
GlcNAc-P12:



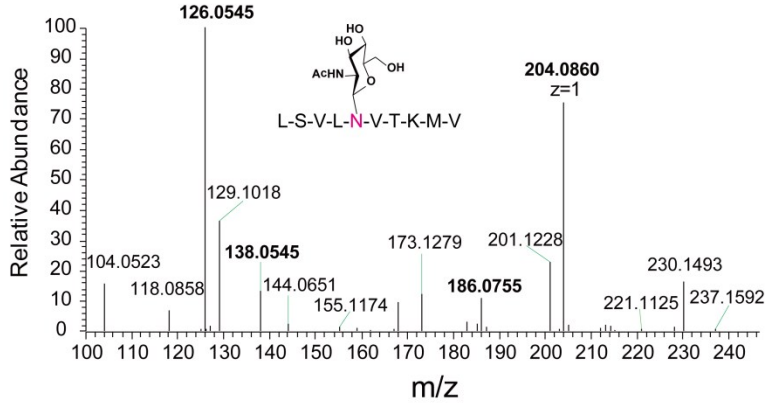
GlcNAc-P13:



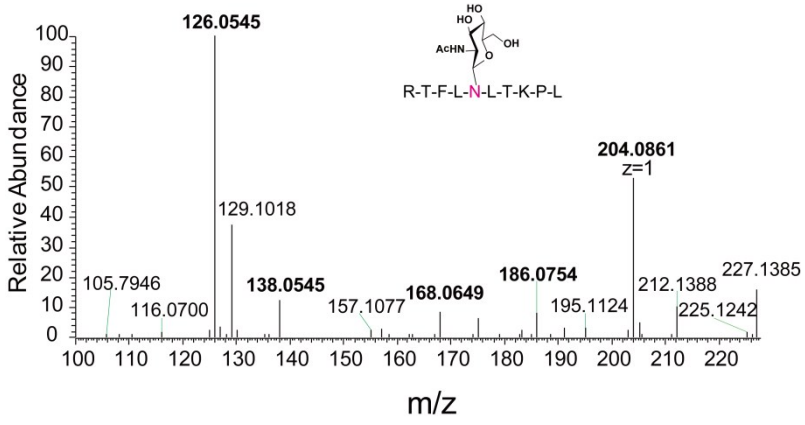
GlcNAc-P14:



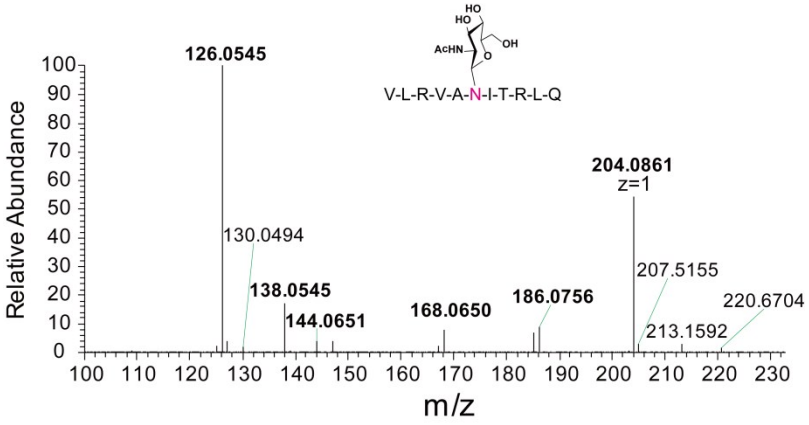
GlcNAc-P15:



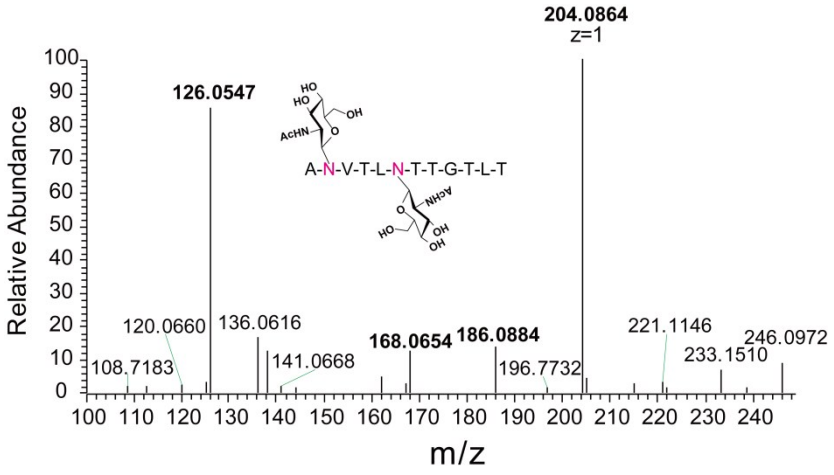
GlcNAc-P16:



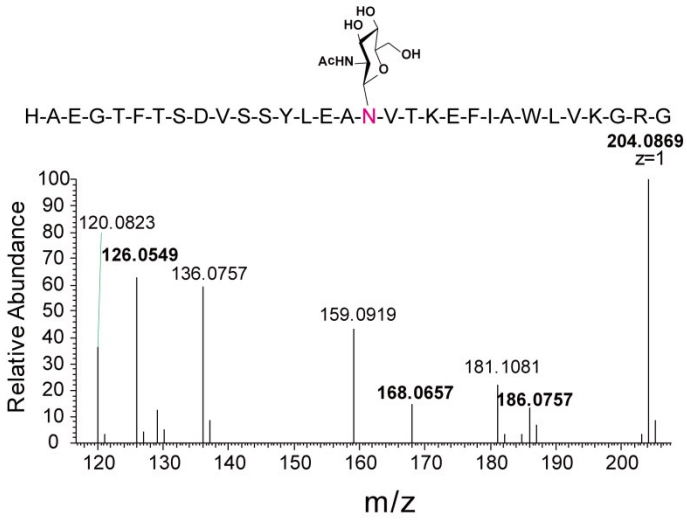
GlcNAc-P17:



GlcNAc-P18:



GlcNAc-P19:



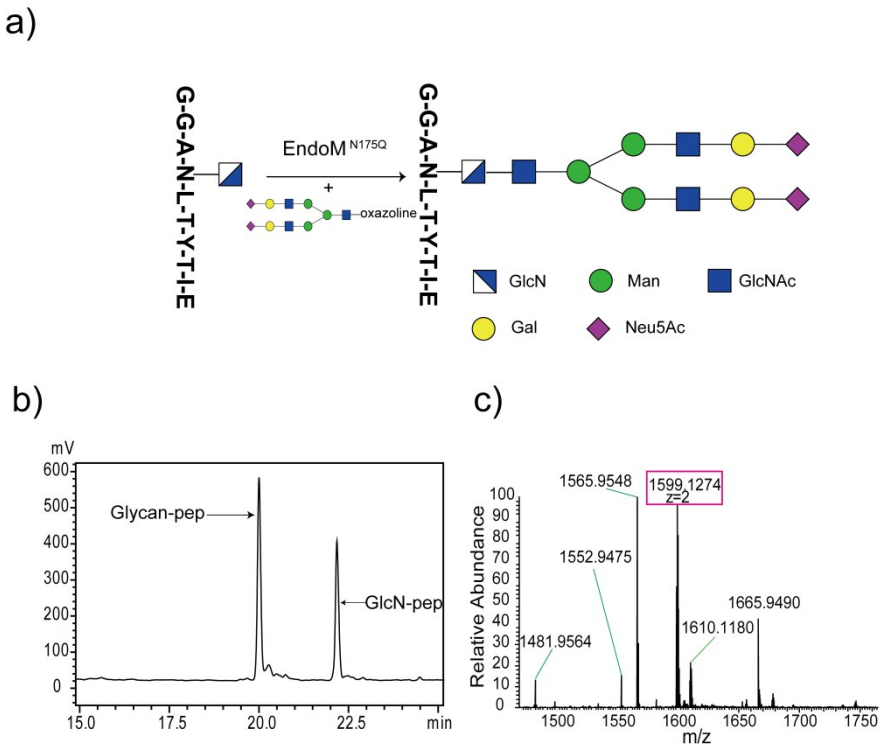


Fig. S4. Transglycosylation of GlcN-peptide. a) Scheme of transglycosylation of GlcN-peptide; b) Analytical HPLC monitoring of EndoM^{N175Q} catalyzed transglycosylation using the GlcN-peptide as the acceptor substrate; c) ESI-MS profile of desired glycan-peptide with an observed mass of 1599.127, z=2 (calculated: M = 3199.2824).

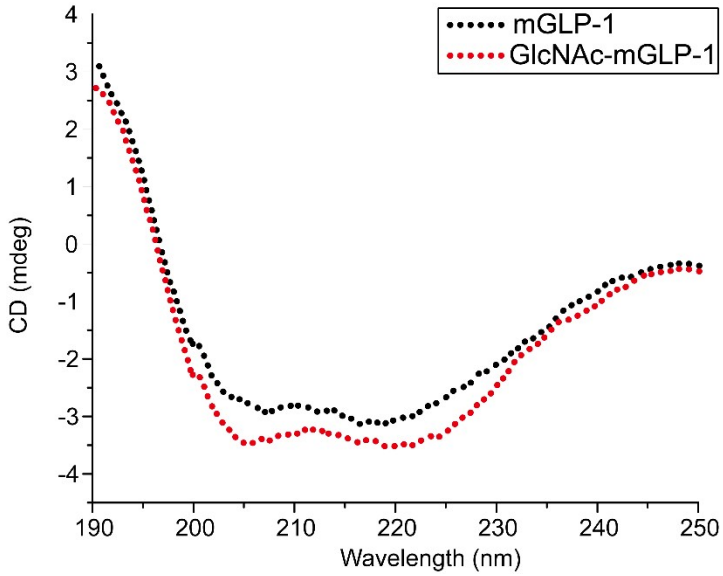
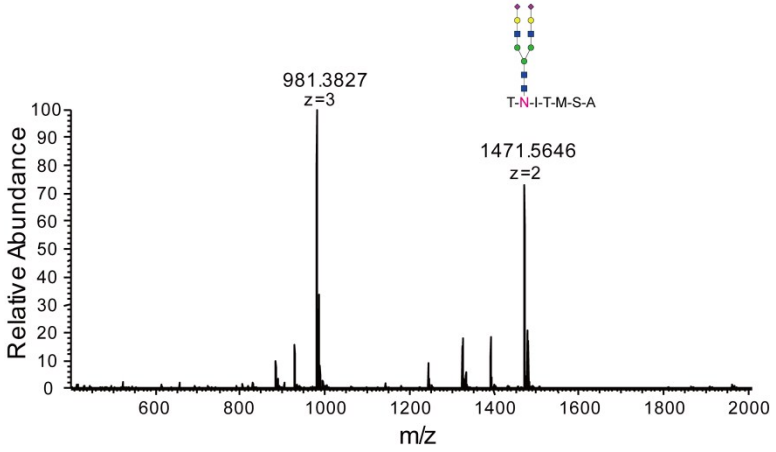


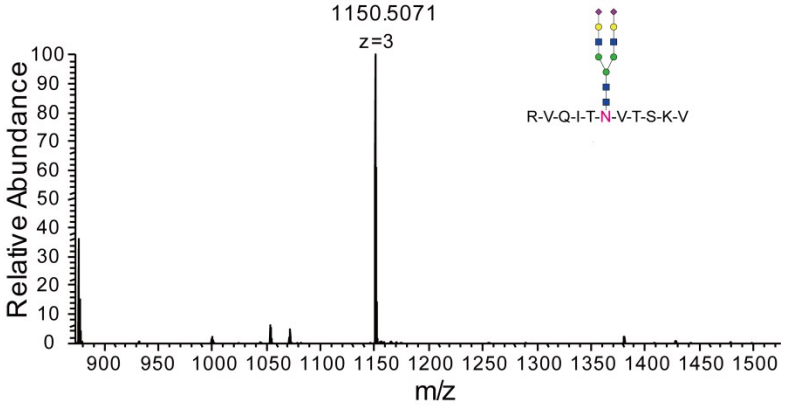
Fig. S5. Far-UV CD spectra of mGLP-1 and GlcNAc-mGLP-1.

Figure S6. ESI-MS spectra of glycopeptides Glycan-P11~Glycan-P19

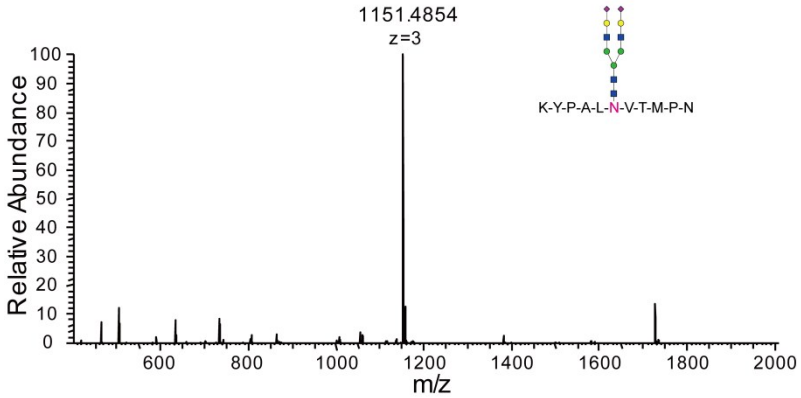
Glycan-P11: Calculated M = 2942.1183; found (*m/z*): $[M+3H]^{3+}$ = 981.3827, $[M+2H]^{2+}$ = 1471.5646.



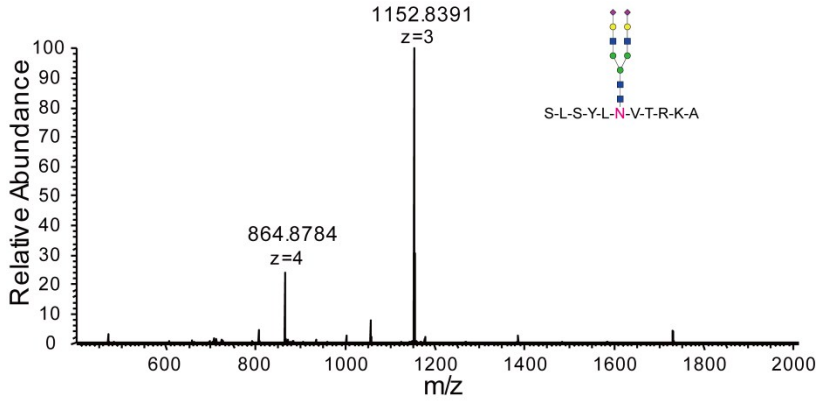
Glycan-P12: Calculated M = 3449.5006; found (*m/z*): $[M+3H]^{3+}$ = 1150.5071.



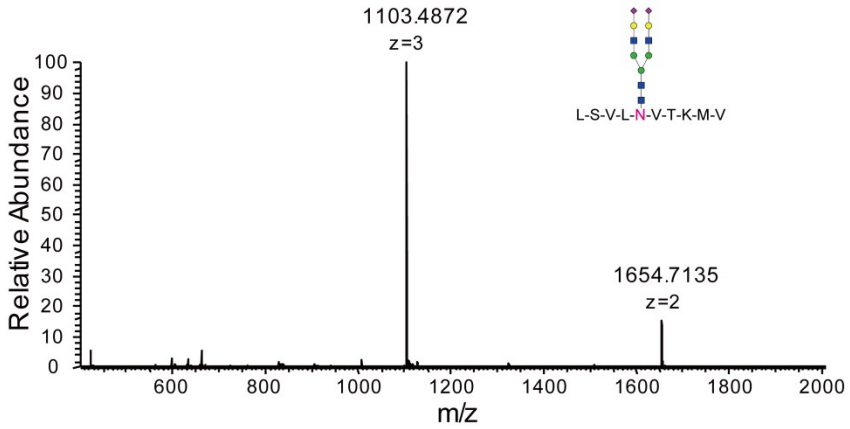
Glycan-P13: Calculated M = 3452.4138; found (*m/z*): $[M+3H]^{3+}$ = 1151.4854.



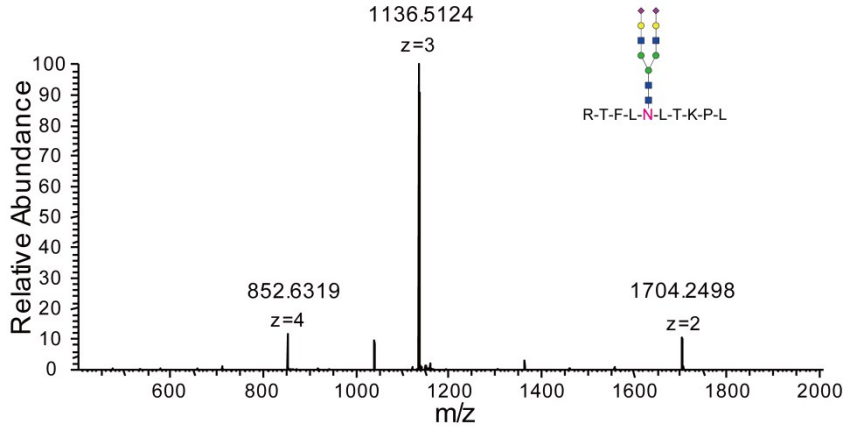
Glycan-P14: Calculated M = 3456.4741; found (*m/z*): $[M+4H]^{4+}$ = 864.8784, $[M+3H]^{3+}$ =1152.8391.



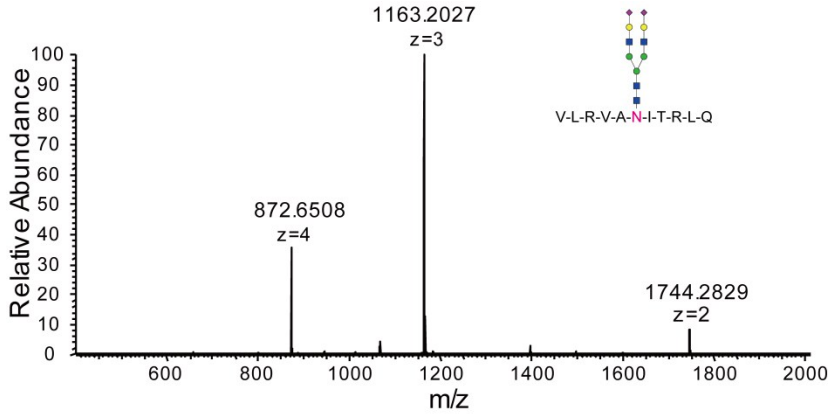
Glycan-P15: Calculated M = 3308.4178; found (*m/z*): $[M+3H]^{3+}$ = 1103.4872, $[M+2H]^{2+}$ =1654.7135.



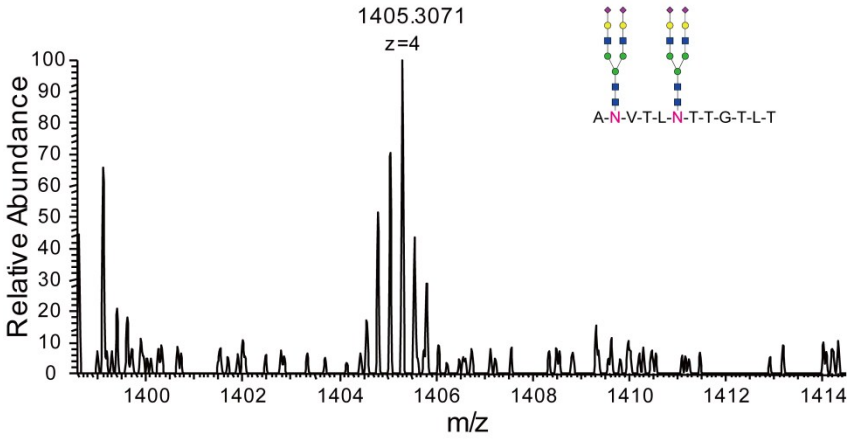
Glycan-P16: Calculated M = 3406.4907; found (*m/z*): $[M+4H]^{4+}$ = 852.6319, $[M+3H]^{3+}$ = 1136.5124, $[M+2H]^{2+}$ = 1704.2498.



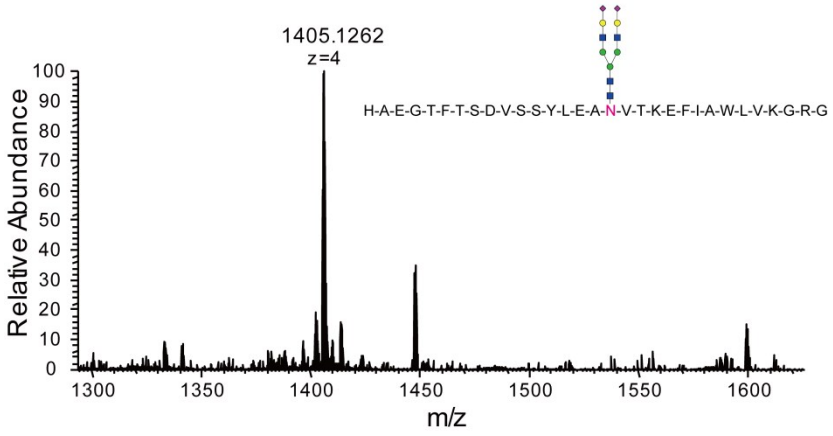
Glycan-P17: Calculated M = 3486.5605; found (*m/z*): $[M+4H]^{4+}$ =872.6508, $[M+3H]^{3+}$ =1163.2027, $[M+2H]^{2+}$ = 1744.2829.



diGlycan-P18: Calculated M = 5616.1815; found(*m/z*): $[M+4H]^{4+}$ = 1405.3071.



Glycan-P19: Calculated M = 5616.4824; found (*m/z*): $[M+4H]^{4+}$ = 1405.1262.



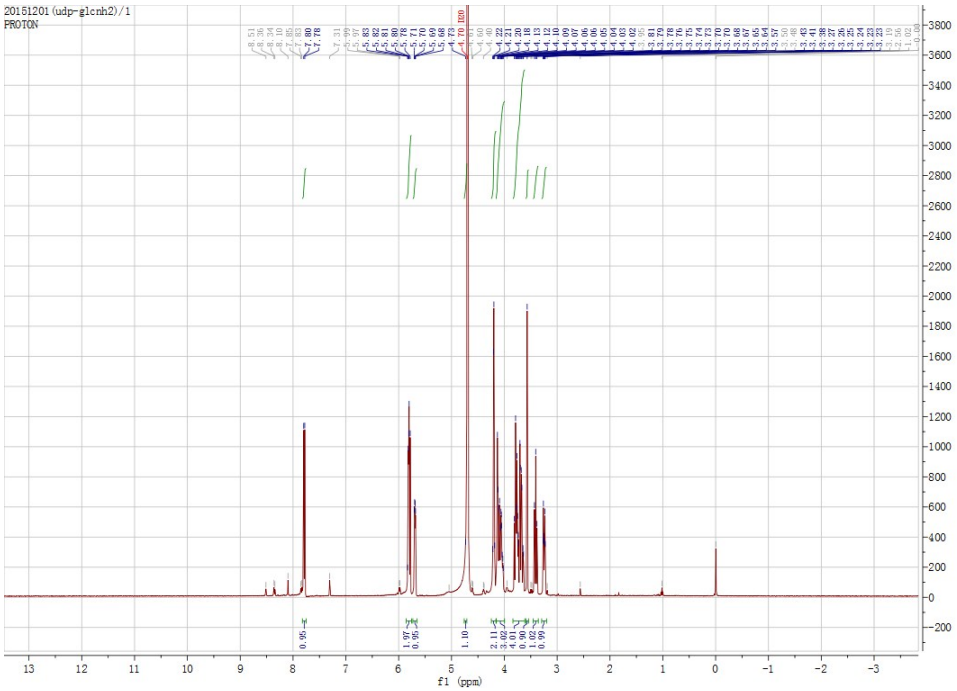


Fig. S7. ^{1}H NMR spectrum of UDP-GlcN.

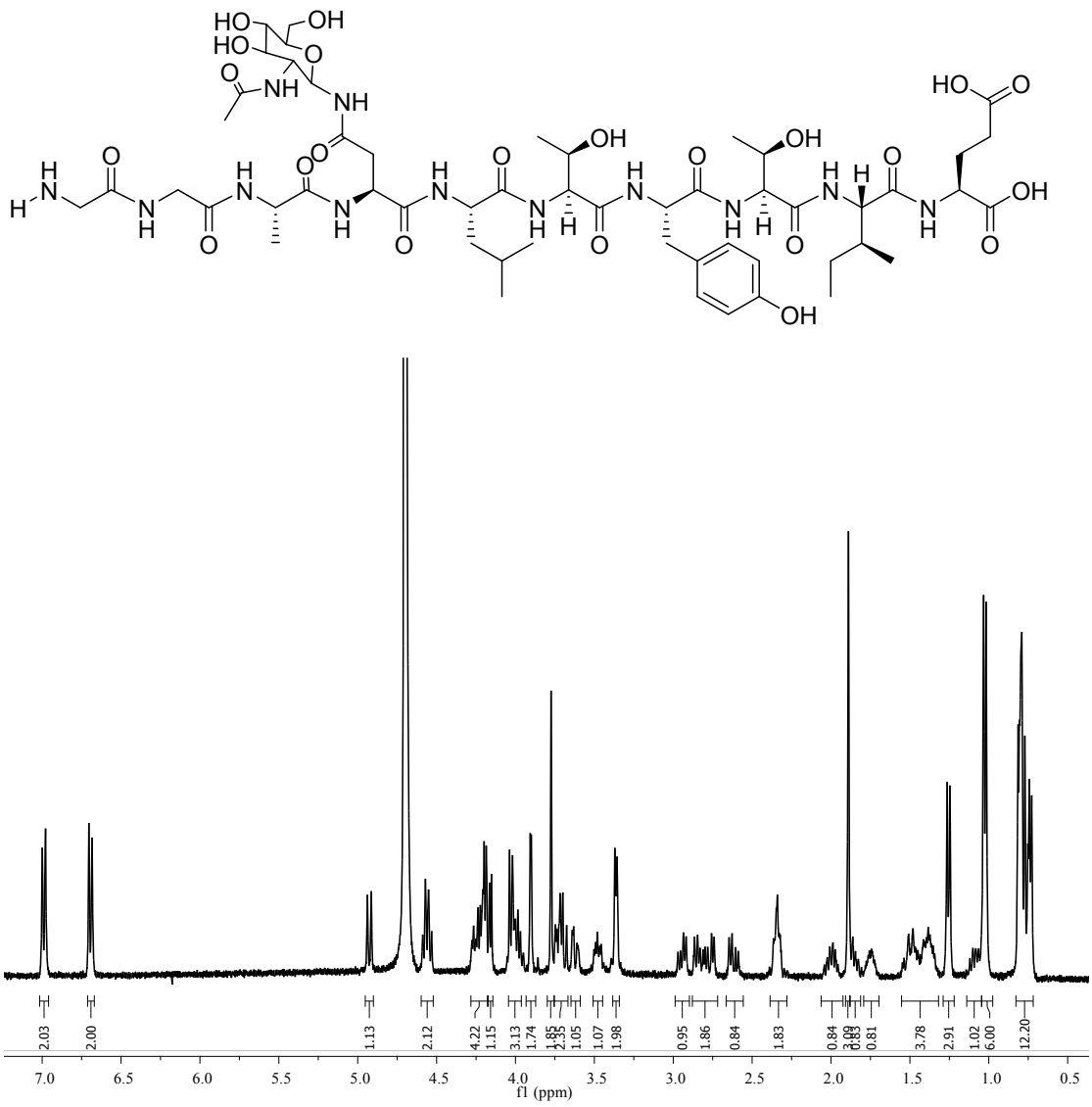


Figure S8. ^{1}H NMR spectra of GGAN(GlcNAc)LTYTIE.

Reference

1. M. M. Muthana, J. Qu, Y. Li, L. Zhang, H. Yu, L. Ding, H. Malekan and X. Chen, *Chem. Commun.*, 2012, **48**, 2728-2730.
2. Y. Zou, M. Xue, W. Wang, L. Cai, L. Chen, J. Liu, P. G. Wang, J. Shen and M. Chen, *Carbohydr. Res.*, 2013, **373**, 76-81.
3. C. Ma, Q. Zhang, J. Qu, X. Zhao, X. Li, Y. Liu and P. G. Wang, *J. Proteomics*, 2015, **114**, 61-70.