Supporting Information

Orthogonal site-specific N-terminal biotinylation and C-terminal glycopeptide labelling of Cholera toxin B-subunit

Jonathan P. Dolan†, Darren C. Machin, Simone Dedola, Robert A. Field, Michael E. Webb*, W. Bruce Turnbull†

†School of Chemistry and Astbury Centre of Structural Biology, University of Leeds, Leeds, LS2 9JT, UK
2Iceni Glycoscience Ltd, Norwich Research Park, Norwich, NR4 7GJ, U.K.
3Department of Chemistry and Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, U.K.
†Present Address: School of Chemical and Physical Sciences & Centre for Glycoscience Research and Training, Keele University, Keele, Staffordshire, ST5 5BG, UK.

*Correspondence:
W. Bruce Turnbull m.e.webb@leeds.ac.uk
w.b.turnbull@leeds.ac.uk

Table of Contents
General Reagents & Methods ................................................................. 2
Synthesis of N-Fmoc-α-O-((AcO)₃GalNAc)-L-Serine and Threonine .................. 3
Solid Phase Peptide Synthesis (SPPS) .......................................................... 9
General Biochemical Methods and Equipment ........................................... 17
SDS-PAGE Gels of C-terminal Sortase Ligation Optimisation ......................... 22
SDS-PAGE Gels of CBD-Sortase7M Comparison and Optimisation .................. 26
Protein HRMS Spectra of Oxime Ligation .................................................. 28
References .................................................................................................... 30
NMR Spectra ................................................................................................. 31
HPLC Traces ................................................................................................. 48
**General Reagents & Methods**

Unless stated otherwise, all starting materials and reagents were purchased from commercial suppliers and used without further purification. All solvents used were dried prior to use, according to standard methods, unless otherwise stated. Reactions were performed under an N₂ atmosphere and within glassware which was oven dried. Completion of reactions was initially determined by TLC and visualized using shortwave ultraviolet light (254 nm) and/or charring with 5% H₂SO₄/MeOH. TLC plates used were Merck Silica-Gel 60 F₂₅₄ Aluminium backed. Silica chromatography columns prepared using Fisher 60Å 43-60 micron silica gel. Lyophilisation carried out using Virtis Benchtop K freeze dryer. Centrifugation was performed using an Eppendorf Centrifuge 5810.

NMR spectroscopy was recorded using Bruker AV3HD-400 (400 MHz, BBO Probe), Bruker AV4 NEO (500 MHz, BBO, TXI & TBO Probe) and Bruker AV4 NEO (500 MHz, C/H cryoprobe) spectrometers. NMR data is reported in parts per million (ppm) referenced to residual solvent signal at room temperature (Fulmer et al., 2010). The following abbreviations are used in ¹H NMR analysis: Ar = aromatic, s = singlet, d = doublet, t = triplet, q = quartet, m= multiplet, dd = double doublet, dt = doublet of triplets, td = triplet of doublets and ddd = double double doublet.

HRMS was performed using Bruker Daltonics MicroTOF mass spectrometer employing electrospray (ES+) ionisation. LC-MS analysis performed on Brucker AmaZon X series LC-MS spectrometer. IR spectroscopy was recorded using a Bruker Platinum ATR spectrometer. Optical rotations were measured with a Schmidt and Haensch Polartronic H-532 at the sodium D-line with the [α]D²₀ values given in the units 10⁻¹ deg cm² g⁻¹.

Standard numbering conventions for carbohydrates in pyranoside and furanoside systems is followed (i.e numbering starts from the anomeric centre). Numbering schemes of more complex structures will be demonstrated on the structure provided.
Synthesis of N-Fmoc-α-O-((AcO)₃GalNAc)-L-Serine and Threonine

Scheme S1: Synthesis of N-Fmoc-α-O-((AcO)₃GalNAc)-L-Serine and Threonine for use in solid phase peptide synthesis. a) NaN₃, EtOAc; b) Imidazole, H₂SO₄ (61%, 2 steps); c) d-Galactosamine, CuSO₄, K₂CO₃, MeOH, H₂O; d) Ac₂O, Pyridine, α/β=1/10 (62%, 2 steps); e) H₄N₂·AcOH, DMF; f) Cl₃CCN, DBU, DCM, α/β=10/1 (56%, 2 steps); g) TMSOTf, DCM, Et₂O, −30°C, α/β=1/1; h) Zn, CuSO₄, Ac₂O, AcOH, THF (S5i: 39%, S5ii: 40%, 2 steps); i) 95% TFA in H₂O (S6i: 74%, S6ii: 92%).

Imidazole-sulfonyl-azide hydrogen sulfate salt (S1)

Imidazole-1-sulfonyl azide is highly explosive in its neutral form. Reaction should only be performed behind Perspex blast shield and no attempt to concentrate solutions of imidazole-1-sulfonyl azide should be made. Sodium azide and imidazole were dried overnight (12-14 hrs) in a vacuum desiccator over phosphorous pentoxide prior to use. Extra dry EtOAc over molecular sieves was purchased from Acros Organics; a fresh bottle was used for each reaction.
Sodium azide (5.00 g, 77 mmol, 1.0 eq) was placed in a dry 500 mL three neck RBF with a dry stirrer bead. Extra dry EtOAc (77 mL) was added to the flask and the resulting suspension cooled to 0 °C. Sulfuryl chloride (6.2 mL, 77 mmol, 1.0 eq) was added dropwise over 5 minutes and the mixture allowed to warm to room temperature and stirred for 24 hrs. The suspension was recooled to 0 °C and imidazole (10.00 g, 146 mmol, 1.9 eq) was added continuously over 5 minutes. The suspension was stirred at RT for 3 hrs. The mixture was basified by the addition of sat. aq. NaHCO₃ solution (150 mL). Once bubbling had ceased, the mixture was separated, the organic portion washed with water (150 mL), dried over MgSO₄ and filtered. Filtrate was recooled to 0 °C and placed under N₂ atmosphere. Conc. H₂SO₄ (4.1 mL, 77 mmol, 1.0 eq) was added dropwise over the course of 5 minutes and gradually warmed to room temperature with vigorous stirring. Over the course of 30 minutes, a colourless precipitate formed and was collected by vacuum filtration. Precipitate was washed with a small amount of ice-cold EtOAc and the crystals were dried under high vacuum to yield imidazole-1-sulfonyl azide hydrogen sulfate salt, S1 (12.71 g, 46.8 mmol, 61%).

NMR data is in agreement with reported data (Potter et al., 2016).

\[ ^1H \text{NMR (400 MHz, DMSO): } \delta 14.29 \text{ (1 H, br S, NH\textsuperscript{+})}, 12.09 \text{ (1 H, br S, } \text{HSO}_4\text{)}, 9.07 \text{ (1 H, s, } H_1), 7.67 \text{ (2 H, s, } H_2, H_3) \text{;} \]

\[ ^13C \text{NMR (100 MHz, DMSO): } \delta 134.5 \text{ (C}_1\text{)}, 119.4 \text{ (C}_2\text{, C}_3\text{);} \]

\[ \text{IR (v\text{max/ cm}^{-1}): } 3083 \text{ (C=O stretch), 2860 (C-H stretch), 2177 (N=N=N stretch), 1586 (N-H bend), 1460 (S=O stretch);} \]

\[ \text{HRMS [ES+] found } [M-\text{HSO}_4\text{-}]^+ 174.0077, \text{ C}_3\text{H}_6\text{N}_2\text{O}_6\text{S requires 174.0080}. \]

\[ 1,3,4,6\text{-Tetra-O-acetyl-2-azido-2-deoxy-D-galactopyranose (S2)} \]

D-Galactosamine hydrochloride (1.00 g, 5.58 mmol, 1.0 eq). K₂CO₃ (2.16 g, 15.62 mmol, 2.8 eq) and CuSO₄·5H₂O (14 mg, 56 µmol, 1 mol%) were dissolved MeOH/H₂O (5:3, 29 mL). To this mixture imidazole-1-sulfonyl azide hydrogen sulfate, S1, (1.817 g, 6.70 mmol, 1.2 eq) was slowly added and left to stir at room temperature. After 3.5 hrs, the mixture was filtered through a pad of celite and co-evaporated with toluene. The water bath was not allowed to exceed 20 °C. Acetic anhydride (2.7 mL, 28.56 mmol, 5.1 eq) was added dropwise to the crude azido-galactose dissolved in pyridine (30 mL) and stirred overnight. Pyridine was removed in vacuo, co-evaporating with toluene twice. The residue dissolved in EtOAc (50 mL) and washed with sat. aq. NaHCO₃ solution (3 × 25 mL) and brine (1 × 25 mL). The organic extract was dried over MgSO₄, filtered, and concentrated to dryness. Crude product was purified using flash column chromatography (2:1 Hexane–EtOAc), to yield 1,3,4,6-tetra-O-acetyl-2-azido-2-deoxy-D-galactose, S2 (1.30 g, 3.48 mmol, 62%, 1:10 α/β) as a colourless oil.

NMR data is in agreement with reported data (Alper et al., 1996; De Silva et al., 2009).

\[ R_f \text{ 0.41 (2:1 Hexane–EtOAc); } ^1H \text{NMR (300 MHz, CDCl}_3\text{): } \delta 6.29 \text{ (1 H, d, } J = 3.5 \text{ Hz, } H_{1\alpha}\text{),}\]

\[ 5.54 \text{ (1 H, d, } J = 8.5 \text{ Hz, } H_{1\beta}\text{),} 5.44 \text{ (1 H, d, } J = 2.4 \text{ Hz, } H_{4\alpha}\text{),} 5.35 \text{ (1 H, d, } J = 3.1 \text{ Hz, } H_{4\beta}\text{),}\]

\[ 5.29 \text{ (1 H, dd, } J = 11.0, 2.9 \text{ Hz, } H_{3\alpha}\text{),} 4.89 \text{ (1 H, dd, } J = 10.8, 3.3 \text{ Hz, } H_{3\beta}\text{),} 4.13–3.95 \text{ (3 H, m,}\]

\[ H_5, H_{6a}, H_{6b}\text{),} 3.91 \text{ (1 H, dd, } J = 11.0, 3.6 \text{ Hz, } H_{2\alpha}\text{),} 3.81 \text{ (1 H, dd, } J = 10.3, 9.0 \text{ Hz, } H_{2\beta}\text{),} 2.17 \]
(3 H, s, Me), 2.13 (3 H, s, Me), 2.03 (3 H, s, Me), 2.00 (3 H, s, Me); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 170.2 (C=O), 169.8 (C=O), 169.5 (C=O), 168.4 (C=O), 92.7 (C$_{19}$), 90.3 (C$_{1a}$), 71.6 (C$_6$), 71.2 (C$_{3b}$), 68.6 (C$_{3a}$), 66.6 (C$_{4a}$), 66.1 (C$_{4b}$), 61.0 (C$_{5a}$), 60.9 (C$_{5b}$), 59.6 (C$_{2a}$), 20.9 (Me), 20.7 (Me), 20.5 (Me); IR ($\nu_{\text{max}}$/cm$^{-1}$): 2939 (C–H stretch), 2112 (N=N=N stretch) and 1743 (C=O stretch); HRMS [ES$^+$] found [M+N$^+$_H]$^+$ 391.1458, C$_{14}$H$_{23}$N$_4$O$_9$ requires 391.1460.

3,4,6-Tri-O-acetyl-2-azido-2-deoxy-D-galactopyranosyl trichloroacetimidate (S3)

1,3,4,6-Tetra-O-acetyl-2-azido-2-deoxy-galactopyranose, S2 (1.299 g, 3.48 mmol, 1.0 eq) and hydrazine acetate (385 mg, 4.18 mmol, 1.2 eq) were stirred in dry DMF (6 mL) and heated to 60 °C. After 1 hour, TLC showed reaction had reached completion and reaction was cooled to room temperature. Reaction was diluted with EtOAc (50 mL) and washed with water (2 × 50 mL) and brine (1 × 50 mL). Organic portion was dried over MgSO$_4$ and concentrated to dryness to yield the crude hemiacetal. Crude hemiacetal was dissolved in dry DCM (12 mL) and cooled to 0 °C. Trichloroacetonitrile (7.00 mL, 69.81 mmol, 20.0 eq) and DBU (260 µL, 1.74 mmol, 0.5 eq) were added and mixture stirred at 0 °C. After 2 hours, TLC showed no further reaction and reaction was diluted with DCM (40 mL) and washed with sat. aq. NH$_4$Cl solution (3 × 50 mL) and brine (1 × 50 mL), dried over MgSO$_4$, filtered and concentrated in vacuo. Crude trichloroacetimidate was purified by flash column chromatography (2:3 EtOAc–Hexane) to yield pure 3,4,6-tri-O-acetyl-2-azido-2-deoxy-D-galactopyranosyl trichloroacetimidate, S3 (927 mg, 1.95 mmol, 56%, 10:1 $\alpha$/β) as a yellow foam.

NMR data is in agreement with reported data for the $\alpha$-anomer (Koeller et al., 2000).

$R_f$ 0.38 (2:3 EtOAc–Hexane); $^1$H NMR (400 MHz, CDCl$_3$): $\delta$ 8.79 (1 H, br. s, NH), 6.78 (1 H, d, $J = 3.6$ Hz, H$_1$), 5.50 (1 H, d, $J = 2.3$ Hz, H$_4$), 5.34 (1 H, dd, $J = 3.2$, 11.1 Hz, H$_3$), 4.38 (1 H, t, $J = 6.7$, 6.7 Hz, H$_5$), 4.15–3.98 (3 H, m, H$_2$, H$_{3a}$, H$_{3b}$), 2.13 (3 H, s, Me), 2.04 (3 H, s, Me), 1.97 (3 H, s, Me); $^{13}$C NMR (100 MHz, CDCl$_3$): $\delta$ 170.3 (C=O), 170.0 (C=O), 170.0 (C=O), 160.7 (C=N), 94.5 (C$_1$), 90.7 (CCl$_3$), 69.2 (C$_3$), 68.7 (C$_3$), 67.0 (C$_4$), 61.2 (C$_6$), 57.1 (C$_2$), 20.7 (Me), 20.6 (Me), 20.6 (Me); IR ($\nu_{\text{max}}$/cm$^{-1}$): 3337 (C–H stretch), 3136 (C–H stretch), 2968 (N–H stretch), 2112 (N=N=N stretch), 1744 (C=O stretch) and 1676 (C=N stretch); HRMS [ES$^+$] found [M+Na]$^+$ 497.0012, C$_{14}$H$_{23}$N$_4$O$_9$Na requires 497.0004.
N-Fmoc-O-(3,4,6-tri-O-acetyl-2-azido-2-deoxy-α-D-galactopyranosyl)-L-serine (S4i) and threonine tert-butyl ester (S4ii)

Trichloroacetimidate, S3, (927 mg, 1.95 mmol, 1.1 eq) and Fmoc-Ser-O'Bu (679 mg, 1.77 mmol, 1.0 eq) or Fmoc-Thr-O'Bu threonine derivative (704 mg, 1.77 mmol, 1.0 eq) were combined and dried overnight under vacuum, then dissolved in a mixture of dry DCM:Et2O (1:1, 10 mL). After cooling to −30 °C, TMSOTf (42 µL, 0.23 mmol, 0.13 eq) was added dropwise. The mixture was stirred at −30 °C for 3 hours. The reaction mixture was quenched with DIPEA (53 µL, 0.30 mmol, 0.17 eq). The reaction mixture was diluted with DCM (40 mL) and washed with water (2 × 50 mL) and brine (1 × 50 mL), the organic portion was filtered and concentrated in vacuo to yield N-Fmoc-O-(3,4,6-tri-O-acetyl-2-azido-2-deoxy-α-D-galactopyranosyl)-L-serine tert-butyl ester, S4i (1.564 g, 1:1 α/β) or threonine derivative, S4ii (1.537 g, 1:1 α/β), both as yellow oils. Both S4i and S4ii were carried through to the next step without further purification.

NMR data is in agreement with reported data (Koeller et al., 2000).

**S4i:** \(R_f\): 0.26 (3:7 EtOAc–Hexane); \(^1\)H NMR (400 MHz, CDCl\(_3\)): δ 7.70 (4 H, d, \(J = 7.6\) Hz, ArH), 7.52 (4 H, d, \(J = 7.2\) Hz, ArH), 7.34 (4 H, t, \(J = 7.5\) Hz, ArH), 7.25 (4 H, t, \(J = 7.5\) Hz, ArH), 5.60 (2 H, d, \(J = 6.3\) Hz, NH(Thr)), 5.47 (1 H, d, \(J = 8.5\) Hz, \(H_1\)), 5.31 (1 H, d, \(J = 3.4\) Hz, \(H_2\)), 4.82 (1 H, dd, \(J = 10.7, 4.1\) Hz, \(H_3\)), 4.35 (4 H, d, \(J = 7.3\) Hz, \(H_6\)), 4.29–4.22 (2 H, m, \(H_9\)), 4.16 (2 H, t, \(J = 7.3\) Hz, \(H_8\)), 4.11–4.00 (4 H, m, \(H_{10}\)), 3.93 (1 H, t, \(J = 6.6\) Hz, \(H_{11}\)), 3.90–3.82 (4 H, m, \(H_7\)), 3.77 (1 H, dd, \(J = 11.6, 8.1\) Hz, \(H_2\)), 2.29 (6 H, s, Me), 2.12 (6 H, s, Me), 1.99 (6 H, s, Me), 1.42 (18 H, s, Me); HRMS [ES+] found [M+H]\(^+\) 697.2712, \(C_{35}H_{43}NaO_{12}\) requires 697.2715.

**S4ii:** \(R_f\) 0.51 (2:3 EtOAc–Hexane); \(^1\)H NMR (400 MHz, CDCl\(_3\)): δ 7.69 (4 H, dd, \(J = 7.7, 3.9\) Hz, ArH), 7.55 (4 H, dt, \(J = 8.0, 2.7\) Hz, ArH), 7.31 (4 H, dt, \(J = 7.2, 2.4\) Hz, ArH), 7.10 (4 H, dd, \(J = 7.7, 3.2\) Hz, ArH), 5.61 (1 H, d, \(J = 9.8\) Hz, NH(Thr)), 5.50 (1 H, d, \(J = 9.8\) Hz, NH(Thr)), 5.40 (1 H, d, \(J = 3.0\) Hz, \(H_{4a}\)), 5.27 (1 H, dd, \(J = 10.4, 3.3\) Hz, \(H_{3a}\)), 5.24 (1 H, d, \(J = 3.0, \ H_{4b}\)), 5.04 (1 H, d, \(J = 3.8\) Hz, \(H_{4a}\)), 4.69 (1 H, dd, \(J = 10.7, 3.3\) Hz, \(H_{3b}\)), 4.51 (1 H, qd, \(J = 12.4, 5.9, 2.1\) Hz, \(H_{5a}\)), 4.41 (1 H, d, \(J = 8.0, \ H_{5b}\)), 4.38 (1 H, dd, \(J = 6.8, 1.4\) Hz, \(H_{7a}\)), 4.34 (1 H, dd, \(J = 7.7, 2.7\) Hz, \(H_{7b}\)), 4.34–4.27 (3 H, m, \(H_{5a}, H_{5b}\)), 4.27–4.13 (5 H, m, \(H_{9a}, H_{9b}, H_{6a}, H_{6b}\)), 4.04 (4 H, dd, \(J = 6.6, 2.7\) Hz, \(H_{10a}, H_{10b}\)), 3.78 (1 H, t, \(J = 6.8\) Hz, \(H_{11}\)), 3.57 (2 H, dd, \(J = 11.1, 3.4\) Hz, \(H_{2a}, H_{2b}\)), 2.28 (6 H, s, Me), 2.08 (6 H, s, Me), 2.00 (6 H, s, Me), 1.43 (18 H, s, t-Bu(Me)), 1.27 (6 H, dd, \(J = 12.8, 6.2\) Hz, Me-8); IR (\(v_{\text{max}}/\text{cm}^{-1}\)): 3354 (C-H stretch), 2979 (N-H stretch), 2111 (N=H=N stretch), 1721 (C=O stretch); HRMS [ES+] found [M+Na]\(^+\) 733.2704, \(C_{35}H_{42}NaO_{12}Na\) requires 733.2691.
**N-Fmoc-O-(2-acetimido-2-deoxy-3,4,6-Tri-O-acetyl-α-D-galactopyranosyl)-L-serine (S5i) and threonine tert-butyl ester (S5ii)**

![Diagram](image)

Zinc dust (15.3 g, 234 mmol, 17.5 eq) was activated in 2% aq. CuSO₄ (120 mL). After 10 minutes, the zinc/copper mixture was added to crude S4i (9.34 g, 13.4 mmol, 1.0 eq) or crude S4ii (9.51 g, 13.4 mmol, 1.0 eq) dissolved in a mixture of THF/AcOH/Ac₂O (3:2:1, 180 ml). The mixture was stirred at room temperature for 4 hours. Once evolution of gas had ceased and TLC showed the reaction had reached completion; the mixture was filtered through celite and the pad was washed with THF (60 mL). The filtrate was concentrated in vacuo and residue dissolved in DCM (300 mL) and washed with 0.1 M HCl (2 × 400 mL). The organic portion was dried over MgSO₄, filtered and concentrated in vacuo. Crude acetamide was purified by flash column chromatography (1% MeOH in DCM) to yield pure S5i (2.626 g, 3.68 mmol, 39%) or S5ii (3.113 g, 4.28 mmol, 40%) as glassy white foams.

NMR data is in agreement with reported data (Koeller et al., 2000).

**S5i: [a]_D^25 +59 (c 1, chloroform); R_f 0.65 (15:1 DCM–MeOH); ^1H NMR (400 MHz, DMSO-d₆): δ 7.76 (2 H, d, J = 7.5 Hz, ArH), 7.61 (2 H, d, J = 7.1 Hz, ArH), 7.40 (2 H, t, J = 7.4 Hz, ArH), 7.32 (2 H, t, J = 7.4 Hz, ArH), 5.80 (2 H, d, J = 7.9 Hz, NHAc, NH(Ser)). 5.37 (1 H, d, J = 2.8 Hz, H₄), 5.12 (1 H, dd, J = 11.2, 2.1 Hz, H₃), 4.83 (1 H, d, J = 3.3 Hz, H₁), 4.59 (1 H, td, J = 11.1, 3.3 Hz, H₂), 4.46–4.39 (3 H, m, H₉, H₁₀), 4.24 (1 H, t, J = 6.9 Hz, H₁₁), 4.15–3.99 (3 H, m, H₅, H₆), 3.96 (1 H, d, J = 9.5 Hz, H₇a), 3.84 (1 H, d, J = 8.5 Hz, H₇b), 2.15 (3 H, s, Me), 1.99 (3 H, s, Me), 1.99 (3 H, s, Me), 1.93 (3 H, s, Me), 1.48 (9 H, s, t-Bu(Me)). ^13C NMR (100 MHz, DMSO-d₆): δ 170.9 (C=O), 170.4 (C=O), 170.1 (C=O), 169.0 (C=O), 155.9 (C=O), 143.8 (Ar), 141.3 (Ar), 127.8 (Ar), 127.1 (Ar), 125.1 (Ar), 120.1 (Ar), 98.8 (C₁), 83.1 (tBu(quat. C)), 69.3 (C₃), 68.4 (C₄), 67.3 (C₂), 67.3 (C₇), 62.0 (C₈), 54.8 (C₉), 53.5 (C₁₀), 47.6 (C₂), 47.1 (C₁₁), 28.1 (tBu(Me)), 23.2 (Me), 20.8 (Me), 20.8 (Me), 20.6 (Me); HRMS [ES+] found [M+H]^+ 713.2925, C₃₆H₄₅N₂O₁₃ requires 713.2916.

**S5ii: [a]_D^25 +62 (c 1, chloroform); R_f 0.83 (15:1 DCM–MeOH); ^1H NMR (400 MHz, CDCl₃): δ 7.70 (2 H, d, J = 7.1 Hz, ArH), 7.56 (2 H, d, J = 7.2 Hz, ArH), 7.33 (2 H, t, J = 6.5, 6.5 Hz, ArH), 7.25 (2 H, t, J = 7.2, 7.2 Hz, ArH), 6.01 (1 H, d, J = 9.9 Hz, NH(Thr)), 5.62 (1 H, d, J = 9.4 Hz, NHAc), 5.31 (1 H, d, J = 2.5 Hz, H₂), 5.02 (1 H, dd, J = 11.3, 2.5 Hz, H₃), 4.82 (1 H, d, J = 3.1 Hz, H₁), 4.55 (1 H, td, J = 10.5, 10.5, 3.1 Hz, H₂), 4.45–4.31 (2 H, m, H₆a, H₆b), 4.22–4.12 (4 H, m, H₁₁, H₉, H₇a, H₇b), 4.04–3.97 (2 H, m, H₁₀), 2.08 (3 H, s, Me), 1.96 (3 H, s, Me), 1.92 (6 H, s, Me), 1.32 (9 H, s, OtBu), 1.24 (3 H, d, J = 6.1 Hz, H₈); ^13C NMR (100 MHz, CDCl₃): 170.8 (C=O), 170.3 (C=O), 170.2 (C=O), 169.9 (C=O), 156.4 (C=O), 143.6 (Ar), 141.2 (Ar), 127.7 (Ar), 127.0 (Ar), 125.0 (Ar), 119.9 (Ar), 99.8 (C₁), 83.1 (tBu(quat. C)), 76.8 (C₁₁), 68.6 (C₃), 67.3 (C₄), 67.2 (C₂), 67.1 (C₇), 62.1 (C₁₀), 58.9 (C₉), 47.2 (C₂), 47.1 (C₇), 28.0 (tBu(Me)), 23.1 (Me), 20.6 (Me), 20.5 (Me), 18.5 (C₈); IR (ν_max/cm⁻¹): 3256 (s, C-H stretch),
2980 (s, C-H stretch), 1748 (s, C=O (ester) stretch), 1651 (s, C=O (amide) stretch); HRMS [ES+] found [M+H]+ 727.3072, C37H47N2O13 requires 727.3073.

N-Fmoc-O-(2-acetimido-2-deoxy-3,4,6-Tri-O-acetyl-α-D-galactopyranosyl)-L-serine (S6i) and threonine (S6ii)

S5i (1.00 g, 1.38 mmol, 1.0 eq) or S5ii (1.00 g, 1.40 mmol, 1.0 eq) dissolved in 95% TFA (3.0 mL) and stirred at room temperature for 1 hour. Solvent was removed in vacuo and co-evaporated with toluene. The crude product purified by flash column chromatography (5:1 Toluene–EtOH) to yield S6i (671 mg, 1.02 mmol, 74%) or S6ii (854 mg, 1.27 mmol, 92%) as a glassy off-white foams.

NMR data is in agreement with reported data (Koeller et al., 2000).

S6i: [α]D +88 (c 1, chloroform); Rf 0.47 (5:1 Toluene–EtOH); 1H NMR (500 MHz, DMSO-d6): δ 7.90 (2 H, d, J = 7.5 Hz, ArH), 7.81 (1 H, d, J = 8.4 Hz, NHAc), 7.73 (1 H, d, J = 8.4 Hz, NH(Ser)), 7.72 (2 H, d, J = 7.1, ArH), 7.42 (2 H, t, J = 7.7 Hz, ArH), 7.33 (2 H, t, J = 7.8 Hz, ArH), 5.31 (1 H, d, J = 2.6 Hz, H4), 5.07 (1 H, dd, J = 11.6, 3.2 Hz, H3), 4.86 (1 H, d, J = 3.6 Hz, H2), 4.38 (2 H, dd, J = 7.4, 3.6 Hz, H6), 4.31–4.23 (3 H, m, H4, H9, H11), 4.20 (1 H, ddd, J = 11.7, 8.4, 3.6 Hz, H2), 4.05 (1 H, dd, J = 11.3, 5.8 Hz, H10a), 3.98 (1 H, dd, J = 10.6, 7.0 Hz, H10b), 3.83 (1 H, dd, J = 10.9, 4.1 Hz, H7a), 3.77 (1 H, dd, J = 10.9, 4.8 Hz, H7b), 2.10 (3H, s, Me), 1.95 (3H, s, Me), 1.90 (3H, s, Me), 1.81 (3H, s, Me); 13C NMR (125 MHz, DMSO-d6): δ 171.5 (C=O), 170.1 (C=O), 169.9 (C=O), 169.8 (C=O), 169.7 (C=O), 156.2 (C=O), 143.9 (Ar), 143.8 (Ar), 140.8 (Ar), 137.4 (Ar), 128.9 (Ar), 128.2 (Ar), 125.3 (Ar), 125.2, 120.2 (Ar), 97.9 (C1), 67.6 (C7), 67.5 (C3), 67.1 (C4), 66.3 (C5), 65.7 (C6), 61.6 (C10), 54.3 (C9), 47.1 (C2), 46.7 (C11), 22.5 (Me), 21.1 (Me), 20.6 (Me), 20.5 (Me); HRMS [ES+] found [M+H]+ 657.2291, C32H37N2O13 requires 657.2290.

S6ii: [α]D +72 (c 1, chloroform); Rf 0.31 (5:1 Toluene–EtOH); 1H NMR (400 MHz, DMSO-d6): δ 7.30 (2 H, d, J = 7.5 Hz, ArH), 7.74 (2 H, dd, J = 7.2, 3.6 Hz, ArH), 7.67 (1 H, d, J = 9.4 Hz, NHAc), 7.59 (1 H, d, J = 9.8 Hz, NH(Thr)), 7.42 (2 H, t, J = 7.4 Hz, ArH), 7.33 (2 H, t, J = 7.2 Hz, ArH), 5.32 (1 H, d, J = 2.8 Hz, H4), 5.05 (1 H, dd, J = 11.6, 3.2 Hz, H3), 4.81 (1 H, d, J = 3.8 Hz, H1), 4.45 (2 H, ddd, J = 25.8, 10.8, 7.0 Hz, H6), 4.34–4.27 (2 H, m, H7, H5), 4.26–4.19 (2 H, m, H11, H2), 4.15 (1 H, dd, J = 9.9, 1.2 Hz, H9), 4.03 (2 H, d, J = 6.1 Hz, H10), 2.11 (3H, s, Me), 1.99 (3H, s, Me), 1.91 (3H, s, Me), 1.84 (3H, s, Me), 1.18 (3H, d, J = 6.4 Hz, H8); 13C NMR (100 MHz, DMSO-d6): δ 170.7 (C=O), 170.1 (C=O), 170.0 (C=O), 169.9 (C=O), 169.5 (C=O), 156.9 (C=O), 143.8 (Ar), 140.8 (Ar), 127.7 (Ar), 127.1 (Ar), 125.2 (Ar), 120.2 (Ar), 98.8 (C1), 74.9 (C7), 67.6 (C3), 67.2 (C4), 66.4 (C11), 65.6 (C6), 62.0 (C10), 58.4
(C₉), 46.8 (C₅), 46.6 (C₂), 22.7 (Me), 20.5 (Me), 18.6 (C₆); IR (νmax/cm⁻¹): 3329 (s, C-H stretch), 2941 (s, C-H stretch), 1744 (s, C=O (ester) stretch), 1657 (s, C=O (amide) stretch); HRMS [ES+] found [M+H]^+ 671.2450, C₃₃H₃₉N₂O₁₃ requires 671.2447.

**Solid Phase Peptide Synthesis (SPPS)**

**General Reagents & Equipment for SPPS**

All amino acids, resins and coupling reagents were purchased from Novabiochem or Fluorochem and used without further purification. Fritted polypropylene tubes (2.5 mL) were purchased from Supelco (Merck) and were used for all solid phase reactions. Agitation of solid phase reaction mixture was achieved by rotation on a Stuart blood rotator at room temperature.

Peptides and glycopeptides were synthesised as C-terminal amides using Rink Amide MBHA LL resin (50 µM, loading 0.35 mmol/g).

Peptides were analysed by HPLC using an Agilent 1290 affinity LC system equipped with an Ascentis Express 10 cm × 2.1 mm, 2.7 µm ES-C18 peptide column (0.5 ml min⁻¹) and ultraviolet (UV) detection at 220-280 nm. The peptide column was run with a gradient from 0.1% TFA/5% MeCN (v/v) in H₂O to 0.1% TFA/95% MeCN (v/v) in H₂O over 10 minutes.

**General procedure for Solid Phase Glycopeptide Synthesis**

Low loading Rink Amide MBHA resin (loading: 0.33-0.35 mmol/g, 100 µmol) was shaken in DMF (5 mL) for 2 hours, filtered and washed with DMF. **Fmoc deprotection** – resin bound peptide intermediates were shaken in 20% piperidine in DMF (3 × 5 mL × 3 min) and washed with DMF, CH₂Cl₂ then DMF. **Couplings** – the resin was treated with a 5-fold excess of the Fmoc-amino acid (except for the Fmoc-Ser/Thr(GalNAc)-OH which was employed in a 2-fold excess) in 2 mL DMF containing HCTU (5 equiv.) and DIPEA (10 equiv). The mixture was shaken and coupling times were 40 mins (except for the Fmoc-Ser/Thr(GalNAc)-OH and the subsequent coupling each of which were 12 hours in duration). The excess reagents removed by filtration and the resin washed with DMF, CH₂Cl₂ then DMF. **O-Acetyl Deprotection** – Following coupling of the final amino acid, resin was washed with methanol, shaken in 70% hydrazine hydrate in methanol (3 × 5 mL × 5 min) and washed with methanol. **Glycopeptide Cleavage** – Following Fmoc and O-Acetyl deprotection, the resin was washed with CH₂Cl₂ and then methanol, before drying under vacuum. The resin was suspended in TFA:H₂O:TIS (95:2.5:2.5, 5 mL) and shaken for a maximum of 2 hours. The mixture was filtered and washed with TFA. The solution was concentrated under a stream of N₂ to <1 mL. Glycopeptide obtained by ether precipitation, isolated by centrifugation and pellet further washed with ether. The white precipitate was dissolved in H₂O and lyophilized to give white foams.
**Synthesised Peptides & Glycopeptides**

*H₂N-GVAPDTRP-CONH₂* (*1*)

![Chemical structure](image1)

Yield: 30 mg, 58%.

**HRMS [ES+]** found [M+H]⁺ 811.4433, C₃₄H₅₉N₁₂O₁₁ requires 811.4421; **LCMS [ES+]** found [M+H]⁺ 811.55, C₃₄H₅₉N₁₂O₁₁ requires 811.44.

*H₂N-GVAPDT(GalNAc)RP-CONH₂* (*2*)

![Chemical structure](image2)

Yield: 28 mg, 41%.

**HRMS [ES+]** found [M+H]⁺ 1014.5224, C₄₀H₇₁N₁₃O₁₆ requires 1014.5214; **LCMS [ES+]** found [M+H]⁺ 1014.63, C₄₀H₇₁N₁₃O₁₆ requires 1014.52.

*H₂N-GVAPGSTAPPA-CONH₂* (*3*)

![Chemical structure](image3)

Yield: 30 mg, 65%.

**HRMS [ES+]** found [M+H]⁺ 923.4960, C₄₀H₆₆N₁₂O₁₃ requires 923.4945; **LCMS [ES+]** found [M+H]⁺ 923.61, C₄₀H₆₆N₁₂O₁₃ requires 923.49.
H₂N-GVAPGST(GalNAc)APPA-CONH₂ (4)

Yield: 45 mg, 80%.

HRMS [ES+] found [M+H]+ 1126.5759, C₄₈H₇₉N₁₃O₁₈ requires 1126.5739; LCMS [ES+] found [M+H]+ 1126.69, C₄₈H₇₉N₁₃O₁₈ requires 1126.57.

H₂N-GVAPGS(GalNAc)TAPPA-CONH₂ (5)

Yield: 47 mg, 75%.

HRMS [ES+] found [M+H]+ 1126.5765, C₄₈H₇₉N₁₃O₁₈ requires 1126.5739; LCMS [ES+] found [M+H]+ 1126.70, C₄₈H₇₉N₁₃O₁₈ requires 1126.57.

H₂N-GVAPGS(GalNAc)T(GalNAc)APPA-CONH₂ (6)

Yield: 40 mg, 55%.

HRMS [ES+] found [M+H]+ 1329.6536, C₅₆H₉₂N₁₄O₂₃ requires 1329.6533; LCMS [ES+] found [M+H]+ 1329.78, C₅₆H₉₂N₁₄O₂₃ requires 1329.65.
H₂N-GVAADTRP-CONH₂ (7)

Yield: 32 mg, 82%.

HRMS [ES+] found [M+H]⁺ 785.4274, C₃₂H₅₇N₁₂O₁₁ requires 785.4264; LCMS [ES+] found [M+H]⁺ 785.41, C₃₂H₅₇N₁₂O₁₁ requires 785.43.

H₂N-GVAADTGalNAcRP-CONH₂ (8)

Yield: 58 mg, 77%.

HRMS [ES+] found [M+H]⁺ 988.5061, C₄₀H₇₀N₁₃O₁₆ requires 988.5058; LCMS [ES+] found [M+H]⁺ 988.50, C₄₀H₇₀N₁₃O₁₆ requires 988.51.

H₂N-GV(AEEAc)APDT(GalNAc)RP-CONH₂ (9)

Yield: 41 mg, 71%.

HRMS [ES+] found [M+H]⁺ 1159.5946, C₄₈H₈₃N₁₄O₁₉ requires 1159.5953; LCMS [ES+] found [M+H]⁺ 1160.12, C₄₈H₈₃N₁₄O₁₉ requires 1159.60.
**H₂N-GV(AEEAc)₃APDT(GalNAc)RP-CONH₂ (10)**

Yield: 42 mg, 64%.

**HRMS [ES+]** found [M+H]+ 1304.6702, C₅₄H₉₄N₁₅O₂₂ requires 1304.6692; **LCMS [ES+]** found [M+H]+ 1305.24, C₅₄H₉₄N₁₅O₂₂ requires 1304.67.

**H₂N-GV(AEEAc)₃APDT(GalNAc)RP-CONH₂ (11)**

Yield: 50 mg, 69%.

**HRMS [ES+]** found [M+2H]²⁺ 725.3766, C₆₀H₁₀₆N₁₆O₂₅ requires 1450.7504; **LCMS [ES+]** found [M+2H]²⁺ 725.76, C₆₀H₁₀₆N₁₆O₂₅ requires 1450.75.

**H₂N-GGGAPDT(GalNAc)RP-CONH₂ (12)**

Yield: 48 mg, 93%.
HRMS [ES+] found [M+H]$^+$ 1029.4956, C$_{41}$H$_{69}$N$_{14}$O$_{17}$ requires 1029.4960; LCMS [ES+] found [M+H]$^+$ 1030.04, C$_{41}$H$_{69}$N$_{14}$O$_{17}$ requires 1029.50.

$H_2N$-GV[PEG]$_6$APGSTAPPA-CNH$_2$ (13)

Yield: 42 mg, 84%.

HRMS [ES+] found [M+H]$^+$ 1213.6423, C$_{52}$H$_{89}$N$_{14}$O$_{19}$ requires 1213.6423; LCMS [ES+] found [M+H]$^+$ 1214.29, C$_{52}$H$_{89}$N$_{14}$O$_{19}$ requires 1213.64.

$H_2N$-GV[PEG]$_6$APGST(GalNAc)APPA-CNH$_2$ (14)

Yield: 48 mg, 76%.

HRMS [ES+] found [M+H]$^+$ 1416.7214, C$_{60}$H$_{102}$N$_{15}$O$_{24}$ requires 1416.7217; LCMS [ES+] found [M+H]$^+$ 1417.40, C$_{60}$H$_{102}$N$_{15}$O$_{24}$ requires 1416.72.
Yield: 43 mg, 68%.

**HRMS [ES+]** found [M+H]^+ 1416.7222, \( \text{C}_{60}\text{H}_{102}\text{N}_{15}\text{O}_{24} \) requires 1416.7217; **LCMS [ES+]** found [M+H]^+ 1417.42, \( \text{C}_{60}\text{H}_{102}\text{N}_{15}\text{O}_{24} \) requires 1416.72.

\[ \text{H}_2\text{N-GV(PEG)}_3\text{APGPS(GalNAc)T(GalNAc)APPA-CNH}_2 \ (16) \]

Yield: 55 mg, 75%.

**HRMS [ES+]** found [M+2H]^2+ 810.4036, \( \text{C}_{68}\text{H}_{116}\text{N}_{16}\text{O}_{29} \) requires 1620.8094; **LCMS [ES+]** found [M+H]^+ 1620.57, \( \text{C}_{68}\text{H}_{115}\text{N}_{16}\text{O}_{29} \) requires 1619.80.

\[ \text{H}_2\text{N-GV(PEG)}_3\text{APPATSGPA-CNH}_2 \ (17) \]

Yield: 41 mg, 77%.

**HRMS [ES+]** found [M+H]^+ 1213.6418, \( \text{C}_{52}\text{H}_{89}\text{N}_{14}\text{O}_{19} \) requires 1213.6423; **LCMS [ES+]** found [M+H]^+ 1214.31, \( \text{C}_{52}\text{H}_{89}\text{N}_{14}\text{O}_{19} \) requires 1213.64.
H₂N-GV(PEG)₆APDTRP-CONH₂(18)

Yield: 37 mg, 77%.

HRMS [ES+] found [M+H]⁺ 1101.5903, C₄₆H₈₁N₁₄O₁₇ requires 1101.5899; LCMS [ES+] found [M+H]⁺ 1102.20, C₄₆H₈₁N₁₄O₁₇ requires 1101.59.
General Biochemical Methods and Equipment
Sterilisation of media and equipment was carried out in either a Prestige Medical bench top autoclave or an LTE Touchclave-R autoclave at 121 °C for 15 minutes. Thermo Electron Corporation SAFE 2010 Class II laminar flow cabinet or bench top bunsen burner were used to maintain a sterile environment. Bacterial cultures were incubated in a Kuhner ShakerX ISF1-X or Stuart Orbital incubator. LB-agar plates and enzymatic reactions were incubated in a Binder BD23 incubator.

Centrifugation was performed using either a Beckman Coulter™ Avanti™ JXN-30 centrifuge, Heraeus mufuge 3S-R centrifuge or Heraeus pico centrifuge. Gel filtration chromatography was achieved using a GE Pharmacia AKTA FPLC system or BioRad NGC FPLC system. Spectrophotometric readings were recorded using a Thermoscientific Nanodrop 2000. SDS-PAGE was carried out using Bio-Rad mini protean 3 apparatus, and a Bio-Rad imager Gel Doc™ XR was used to visualise polyacrylamide gels. Proteins were concentrated either using 10k or 30k MWCO Amicon® Ultra-15 centrifugal filter device or Amicon® Stirred Ultrafiltration Cell (10k MWCO membrane). High resolution mass spectrometry of protein samples was performed using a Bruker Daltonics MicroTOF mass spectrometer. Protein samples were loaded at a concentration of 20-40 µM (made up with H₂O) into the instrument before being automatically diluted into 0.1% TFA/50% MeCN (v/v) in H₂O prior to analysis. Analytical grade reagents were supplied by commercial suppliers.

Media & Buffers
Buffers and media were made in-house unless otherwise stated using analytical grade reagents supplied by commercial suppliers. All common buffers and media were prepared with 18.2 MΩ water to the required volume. The pH of the solutions was adjusted using 5 M NaOH or 5 M HCl. Gel filtration buffers were filtered through a 0.22 µm membrane under reduced pressure. Media was sterilised by autoclave at 121 °C for 20 minutes.

Miller's lysogeny broth (LB) medium: 1.0% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 1.0% (w/v) NaCl.

Auto induction medium: 1.0% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 4% (v/v) 25 × salts, 0.1% (v/v) 1000 × metals, 4.3 mM MgCl₂. Solution autoclaved at 120 °C for 20 mins before addition of 20 mL L⁻¹ 50 × Sugar Mix by filter sterilisation (0.2 µm Sartorius Minisart).

25 × salts: 1.25 mM Na₂HPO₄, 1.25 mM KH₂PO₄, 2.5 mM NH₄Cl, 250 mM Na₂SO₄, pH 7.4.

1000 × metals: 50 mM FeCl₃·6H₂O, 20 mM CaCl₂·6H₂O, 100 mM MnCl₂·4H₂O, 100 mM ZnSO₄·7H₂O, 1.7 mM CoCl₂·6H₂O, 2 mM CuCl₂·2H₂O, 4.1 mM Na₂MoO₄·2H₂O, 2 mM Na₂SeO₃, 2 mM H₃BO₃, 2 mM NiSO₄·6H₂O, 50% (v/v) 0.1 M HCl.

50 × Sugar Mix: 20% (v/v) glycerol, 2.5% (v/v) glucose, 10% (w/v) lactose.

Sodium Phosphate Buffered Saline: 50 mM NaH₂PO₄, 100 mM NaCl, pH 7.4.

Tris Buffered Saline: 50 mM Tris, 150 mM NaCl, pH 7.2.
**Expression of CTB-LPETGA**

A glycerol stock of C41 cells containing the mutant plasmid pSAB2.2-LPETGA was used to inoculate 2 x 5 ml of LB media containing 100 µg/ml ampicillin. These cultures were incubated at 37 °C overnight, shaking at 200 rpm and then used to inoculate 4 x 1 L of LB media containing 100 µg/ml of ampicillin. The culture was incubated at 37 °C with shaking (200 rpm) and overexpression by the addition of IPTG (0.5 mM) when OD$_{600}$ = 0.6-0.8 was reached. The cells were incubated overnight at a 25 °C before being collected by centrifugation (10,000 × g, 20 minutes, 4 °C). The *E. coli* cell pellet was discarded and media retained. Solid ammonium sulfate was added and dissolved to a final concentration of 57% (w/v) to precipitate the protein of interest. The saturated solution was stirred overnight at 4 °C. The solution was centrifuged (17,600 × g, 1 hour, 4 °C), and the supernatant discarded. The protein pellet was resuspended in sodium phosphate buffered saline (10 mL per 1 L LB media) and centrifuged (17,600 × g, 10 minutes, 4 °C) to remove insoluble material and to allow the suspension to be filtered through a 0.8 µm Sartorius Minisart filter.

The protein was purified by Ni-NTA affinity chromatography washing with sodium phosphate buffered saline (no imidazole). CTB has native surface histidine residue meaning the pentamer is able to bind to Ni-NTA resin without the presence of a His-tag. The protein was then eluted from the column by washing with sodium phosphate buffered saline containing 200 mM imidazole, before further purification by size exclusion chromatography on Superdex S75 in sodium phosphate buffered saline.
Figure S1: Purification of CTB-LPETGA for use in labelling reactions. a) SDS-PAGE gel of Ni-NTA affinity chromatography flow through, wash and elution fractions (E1-3). b) SEC trace for CTB-LPETGA using Superdex® S200 16/60 with SDS-PAGE gel showing formation of a stable pentamer which dissociates upon boiling. Two bands appear for protomeric CTB showing a mixture of full length and truncated material. c) Deconvoluted ES-MS of purified CTB-LPETGA protomer (12554.4 Da) and the truncated CTB protomer (11528.9 Da).

Expression of Sortase 7M and CBD-Sortase 7M
The plasmid pET30b-7M SrtA was a gift from Hidde Ploegh (Addgene plasmid #51141; http://n2t.net/addgene:51141; RRID:Addgene_51141). The plasmid pET28a-CBD-Srt7M is a syntetic construct obtained from genscript encoding Sortase 7M with an N-terminal hexahistidine and chitin binding domain purification tags.

A glycerol stock of BL21(DE3) cells containing the mutant plasmid pET30b-Srt7M (Wuethrich et al., 2014) or pET28a-CBD-Srt7M was used to inoculate 5 ml of LB media containing 50 µg/ml kanamycin. These cultures were incubated at 37 °C overnight, shaking at 200 rpm and then 2 mL used to inoculate 1 L of autoinduction-LB medium (Studier, 2005) containing 100 µg/ml of kanamycin. The culture was incubated at 37 °C with shaking (200
rpm) for 5 hours then the incubator temperature was reduced to 25 °C. The cells were incubated overnight before being collected by centrifugation (13,000 × g, 20 minutes, 4 °C). Cells were either used immediately or stored at -80 °C for future lysis and purification.

_E. coli_ cells were resuspended in Tris buffered saline at 4 °C and treated with DNase (10 µg/mL) and protease inhibitors. The cell mixture was lysed mechanically using a Constant Systems cell disruptor (20 kpsi). The cell debris was pelleted by centrifugation (30,000 × g, 45 mins, 4 °C) and the cell lysate was decanted off. Purification of the clarified lysate performed by Ni-NTA affinity chromatography washing with Tris buffered saline containing 50 mM imidazole. The protein was then eluted from the column by washing with Tris buffered saline containing 250 mM imidazole, before further purification by size exclusion chromatography on Superdex S75 in Tris buffered saline. Sortase 7M & CBD-Sortase 7M was stored at -80 °C containing 10% (v/v) glycerol.

**Figure S2:** Purification of Sortase 7M-His6 for use in labelling reactions. a) SDS-PAGE gel of Ni-NTA affinity chromatography column flow through (F1), wash (W1-2) and elution fractions (E1-3). b) SEC trace for Sortase 7M-His6 using Superdex® S75 26/60. c) Deconvoluted ES-MS of purified Sortase 7M-His6 (17849.5458 Da).
**Figure S3**: Purification of Chitin-binding Sortase 7M for use in preparative scale ligation reactions. 
a) SDS-PAGE gel of Ni-NTA affinity chromatography column flow through, wash and elution fractions (E1-8). 
b) SEC trace for Chitin-binding Sortase 7M using Superdex® S75 26/60. 
c) Deconvoluted ES-MS of purified Chitin-binding Sortase 7M (25166.7230 Da).
**SDS-PAGE Gels of C-terminal Sortase Ligation Optimisation**

*Figure S4*: SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Peptide 1 over 2 hours.

*Figure S5*: SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Glycopeptide 2 over 2 hours.
**Figure S6:** SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Peptide 1 in the presence of 10% (v/v) DMSO over 2 hours.

**Figure S7:** SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Glycopeptide 2 in the presence of 10% (v/v) DMSO over 2 hours.

**Figure S8:** SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Peptide 7 (P4A Variant of Peptide 1) over 2 hours.
Figure S9: SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Glycopeptide 8 (P4A Variant of Glycopeptide 2) over 2 hours.

Figure S10: SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Glycopeptide 9 in the presence of 10% (v/v) DMSO over 2 hours.

Figure S11: SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Glycopeptide 10 in the presence of 10% (v/v) DMSO over 2 hours.
**Figure S12:** SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Glycopeptide 11 in the presence of 10% (v/v) DMSO over 2 hours.

**Figure S13:** SDS-PAGE gel showing sortase 7M ligation of CTB-LPETGA and Glycopeptide 12 in the presence of 10% (v/v) DMSO over 2 hours.

**Figure S14:** SDS-PAGE gel showing sortase 7M ligation to CTB-LPETGA in the presence of 10% (v/v) DMSO after 2 hours. Lane 1: CTB-LPETGA (unlabelled); Lane 2: CTB-LPETGA (incubated with Sortase 7M); Lane 3: Glycopeptide 13; Lane 4: Glycopeptide 14; Lane 5:
Glycopeptide 15; Lane 6: Glycopeptide 16; Lane 7: Glycopeptide 17; Lane 8: Glycopeptide 18; Lane 9: Glycopeptide 10.

**SDS-PAGE Gels of CBD-Sortase7M Comparison and Optimisation**

![Image](image1.png)

**Figure S15:** SDS-PAGE gel showing Sortase 7M-His6 ligation of CTB-LPETGA and Glycopeptide 10 in the presence of 10% (v/v) DMSO over 2 hours.

![Image](image2.png)

**Figure S16:** SDS-PAGE gel showing His6-CBD-Sortase 7M ligation of CTB-LPETGA and Glycopeptide 10 in the presence of 10% (v/v) DMSO over 2 hours.
**Figure S17:** SDS-PAGE gel showing His6-CBD-Sortase 7M ligation to CTB-LPETGA in the presence of 10% (v/v) DMSO after 2 hours (Pre-purification of CBD-Sortase 7M with chitin resin and size-exclusion chromatography). Lane 1: CTB-LPETGA (unlabelled); Lane 2: CTB-Glycopeptide 13; Lane 3: CTB-Glycopeptide 14; Lane 4: CTB-Glycopeptide 15; Lane 5: CTB-Glycopeptide 16; Lane 6: CTB-Glycopeptide 17; Lane 7: CTB-Glycopeptide 18; Lane 8: CTB-Glycopeptide 10.

**Figure S18:** SDS-PAGE gel showing His6-CBD-Sortase 7M ligation to CTB-LPETGA in the presence of 10% (v/v) DMSO after 2 hours (Post-purification of CBD-Sortase 7M with chitin resin and size-exclusion chromatography). Lane 1: CTB-LPETGA (unlabelled); Lane 2: CTB-Glycopeptide 13; Lane 3: CTB-Glycopeptide 14; Lane 4: CTB-Glycopeptide 15; Lane 5: CTB-Glycopeptide 16; Lane 6: CTB-Glycopeptide 17; Lane 7: CTB-Glycopeptide 18; Lane 8: CTB-Glycopeptide 10.
Figure S19: SDS-PAGE gel showing His6-CBD-Sortase 7M ligation to CTB-LPETGA in the presence of 10% (v/v) DMSO after 2 hours (Post-purification of CBD-Sortase 7M with chitin resin and size-exclusion chromatography). Lane 1: CTB-LPETGA (unlabelled); Lane 2: CTB-LPETGA (oxidised); Lane 3: Biotin-CTB-LPETGA; Lane 4: Biotin-CTB-Glycopeptide 13; Lane 5: Biotin-CTB-Glycopeptide 14; Lane 6: Biotin-CTB-Glycopeptide 15; Lane 7: Biotin-CTB-Glycopeptide 16; Lane 8: Biotin-CTB-Glycopeptide 17; Lane 9: Biotin-CTB-Glycopeptide 18; Lane 10: Biotin-CTB-Glycopeptide 10.

Protein HRMS Spectra of Oxime Ligation

Figure S20: Deconvoluted high resolution protein mass spectrometry of CTB-LPETGA (12553.3809 Da) and truncated CTB (11528.9093 Da) before oxidation with NaIO₄.
**Figure S21:** Deconvoluted high resolution protein mass spectrometry of NaIO₄ oxidation reaction after 10 minutes showing quantitative oxidation of both substrates in their hydrate forms.

**Figure S22:** Deconvoluted high resolution protein mass spectrometry of oxime ligation reaction with alkoxyamine-PEG₄-biotin, showing near quantitative N-terminal biotinylation of both truncated CTB (11900.0620 Da) and CTB-LPETGA (12924.5271 Da). A small amount of oxidized-CTB-LPETGA that has cyclized to a diketopiperidine is present at 12508.3086 Da.
References

Alper, P. B., Hung, S.-C., and Wong, C.-H. (1996). Metal catalyzed diazo transfer for the synthesis of azides from amines. *Tetrahedron Lett.* 37, 6029–6032. doi:10.1016/0040-4039(96)01307-X.

De Silva, R. A., Wang, Q., Chidley, T., Appulage, D. K., and Andreana, P. R. (2009). Immunological response from an entirely carbohydrate antigen: Design of synthetic vaccines based on Tn-PS A1 conjugates. *J. Am. Chem. Soc.* 131, 9622–9623. doi:10.1021/ja902607a.

Fulmer, G. R., Miller, A. J. M., Sherden, N. H., Gottlieb, H. E., Nudelman, A., Stoltz, B. M., et al. (2010). NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* 29, 2176–2179. doi:10.1021/om100106e.

Koeller, K. M., Smith, M. E. B., and Wong, C. H. (2000). Chemoenzymatic synthesis of PSGL-1 glycopeptides: Sulfation on tyrosine affects glycosyltransferase-catalyzed synthesis of the O-glycan. *Bioorganic Med. Chem.* 8, 1017–1025. doi:10.1016/S0968-0896(00)00041-9.

Potter, G. T., Jayson, G. C., Miller, G. J., and Gardiner, J. M. (2016). An Updated Synthesis of the Diazo-Transfer Reagent Imidazole-1-sulfonyl Azide Hydrogen Sulfate. *J. Org. Chem.* 81, 3443–3446. doi:10.1021/acs.joc.6b00177.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* 41, 207–234. doi:10.1016/j.pec.2005.01.016.

Wuethrich, I., Peeters, J. G. C., Blom, A. E. M., Theile, C. S., Li, Z., Spooner, E., et al. (2014). Site-Specific Chemoenzymatic Labeling of Aerolysin Enables the Identification of New Aerolysin Receptors. *PLoS One* 9, e109883. doi:10.1371/journal.pone.0109883.
NMR Spectra

*Imidazole-1-sulfonyl azide hydrogen sulphate salt (S1)*

![NMR Spectra Diagram](image-url)

**Chemical Shift (ppm)**

- 1.07
- 7.68
- 7.89
- 1.87
- 1.31
- 1.00
- 1.95

**Normalized intensity**

- 1.0
- 0.9
- 0.8
- 0.7
- 0.6
- 0.5
- 0.4
- 0.3
- 0.2
- 0.1
- 0.0
1,3,4,6-Tetra-O-acetyl-2-azido-2-deoxy-D-galactose (S2)
3,4,6-Tri-O-acetyl-2-azido-2-deoxy-α-D-galactopyranosyl trichloroacetimidate (S3)
**N-Fmoc-O-(3,4,6-Tri-O-acetyl-2-azido-2-deoxy-α/β-D-galactopyranosyl)-L-serine tert-butyl ester (S4i)**

Chemical structure of N-Fmoc-O-(3,4,6-Tri-O-acetyl-2-azido-2-deoxy-α/β-D-galactopyranosyl)-L-serine tert-butyl ester (S4i) is shown in the image. The molecule contains a Fmoc group, an azido group, an acetyl group, and a galactopyranosyl group attached to an L-serine tert-butyl ester. The NMR spectrum below the structure shows the chemical shifts and normalized intensity for the molecule. The spectrum indicates the presence of various protons at different chemical shifts, which are characteristic of the functional groups present in the molecule.
N-Fmoc-\((3,4,6\text{-Tri-O-acetyl}-2\text{-azido-2-deoxy-\(\alpha/\beta\)-D-galactopyranosyl})\)-L-threonine tert-butyl ester (S4ii)
N-Fmoc-O-(2-acetimido-2-deoxy-3,4,6-Tri-O-acetyl-α-D-galactopyranosyl)-L-serine tert-butyl ester (S5i)
N-Fmoc-O-(2-acetimido-2-deoxy-3,4,6-Tri-O-acetyl-α-D-galactopyranosyl)-L-threonine tert-butyl ester (S5ii)
N-Fmoc-O-(2-acetimido-2-deoxy-3,4,6-Tri-O-acetyl-α-D-galactopyranosyl)-L-serine (S6i)
N-Fmoc-O-(2-acetimido-2-deoxy-3,4,6-Tri-O-acetyl-α-D-galactopyranosyl)-L-threonine (S6ii)
H₂N-GVAPGS(GalNAc)TAPPA-CONH₂ (5)

H₂N-GVAADTRP-CONH₂ (7)

H₂N-GVAPGS(GalNAc)T(GalNAc)APPA-CONH₂ (6)

H₂N-GVAADT(GalNAc)RP-CONH₂ (8)
$H_2N-GV(AEAc)_2APDT(GalNAc)RP-CONH_2$ (10)

$H_2N-GV(AEAc)_2APGSTAPPA-CONH_2$ (13)
H₂N-GV(AEEAc)₂APGST(GalNAc)APPA-CONH₂ (14)

H₂N-GV(AEEAc)₂APGS(GalNAc)TAPPA-CONH₂ (15)
H$_2$N-GV(AEEAc)$_2$APGS(GalNAc)T(GalNAc)APPA-CONH$_2$ (16)

H$_2$N-GV(AEEAc)$_2$APPATSGPA-CONH$_2$ (17)
$H_2N-GV(AEEAc)_{2}APDTRP-CONH_2$ (18)