ChemBioChem

Supporting Information

N-Terminal Modification of Gly-His-Tagged Proteins with Azidogluconolactone

Karl D. Brune[+],* Ilva Liekniņa[+], Grigorij Sutov[+], Alexander R. Morris[+], Dejana Jovicevic, Gints Kalniņš, Andris Kazāks, Rihards Kluga, Sabine Kastaljana, Anna Zajakina, Juris Jansons, Dace Skrastiņa, Karīna Spunde, Alexander A. Cohen, Pamela J. Bjorkman, Howard R. Morris, Edgars Suna, and Kaspars Tārs*

Author Contributions

K.B. Conceptualization:Lead; Formal analysis:Equal; Investigation:Equal; Methodology:Equal; Project administration:Equal; Supervision:Equal; Validation:Equal; Visualization:Lead; Writing – original draft:Lead
I.L. Formal analysis:Equal; Investigation:Equal; Methodology:Equal; Visualization:Supporting
G.S. Conceptualization:Equal; Formal analysis:Equal; Funding acquisition:Equal; Investigation:Equal; Methodology:Equal; Validation:Equal
A.M. Formal analysis:Equal; Funding acquisition:Equal; Investigation:Equal; Methodology:Equal; Project administration:Equal; Resources:Equal; Validation:Equal
D.J. Investigation:Supporting; Methodology:Supporting
G.K. Investigation:Supporting; Methodology:Supporting; Visualization:Supporting
A.K. Investigation:Supporting; Methodology:Supporting
R.K. Investigation:Supporting; Methodology:Supporting
S.K. Investigation:Supporting; Methodology:Supporting
A.Z. Investigation:Supporting; Methodology:Supporting; Supervision:Supporting; Validation:Supporting
J.J. Investigation:Supporting; Methodology:Supporting
D.S. Investigation:Supporting; Methodology:Supporting
K.S. Investigation:Supporting; Methodology:Supporting
A.C. Investigation:Supporting; Methodology:Supporting
P.B. Funding acquisition:Supporting; Resources:Supporting; Supervision:Supporting
H.M. Formal analysis:Equal; Methodology:Equal; Resources:Equal; Supervision:Equal; Writing – review & editing:-Supporting
E.S. Conceptualization:Supporting; Formal analysis:Supporting; Funding acquisition:Equal; Methodology:Supporting; Project administration:Supporting; Resources:Supporting; Supervision:Supporting; Validation:Supporting
K.T. Conceptualization:Equal; Formal analysis:Equal; Funding acquisition:Equal; Methodology:Equal; Project administration:Equal; Resources:Equal; Supervision:Equal; Validation:Equal; Writing – review & editing:Lead

This PDF file includes:

Acknowledgements

We thank Alexander Leach, Christina Morris, Maria Panico, and the rest of the BioPharmaSpec staff for assistance with advanced MS instrumentation and discussion; George Fleet, Jan Staněk, Stephen Hanessian, Jacobo Cruces and Sergi Jorba for literature and azido-derivative synthesis advice; John Engen for discussion on Nef-1; Maria Elena Bottazzi and Adamo Biter for discussion on Tc24-C4; The Citizens of Jersey, the Durrell Wildlife Conservation Trust, and Digital Jersey for non-financial support; Imperial College London, Department of Bioengineering, Paul Chadderton and the Centre for Synthetic Biology, Department of Life Sciences, Steve Matthews for assistance with 6AGDL NMR spectroscopy; Ināra Akopjana and Jānis Bogans for assistance with yRBD production and purification; Mona Mohsen and Martin Bachmann for discussion on VLPs; Karen Polizzi, and Jeffrey Almond for antigen discussion. Miguel Chillón and the European infrastructure for translational medicine (EATRIS) for assistance with SARS-CoV-2 VNT.

Materials

Chemicals were ordered from Sigma Aldrich and used as received unless noted otherwise.

Solvents for 6AGDL synthesis were ordered from Alfa Aesar and used without further purification.

Milli-Q H_2O was obtained from a Purelab Option Q (Elga) water purification system.

TLC Silica gel 60 F_{254} plates were obtained from Merck KGaA.

Peptides were synthesized by GenScript at >85% purity and intact mass was confirmed by MALDI-TOF MS. Peptides were dissolved in ultrapure water. For Beltide-1 peptides, sometimes precipitation was observed at high concentration (>1 mM), which could be overcome by addition of acetonitrile (ACN) to 10% (v/v).

Cayman AN-2690 (Tavaborole) was obtained from Cambridge Bioscience (UK).

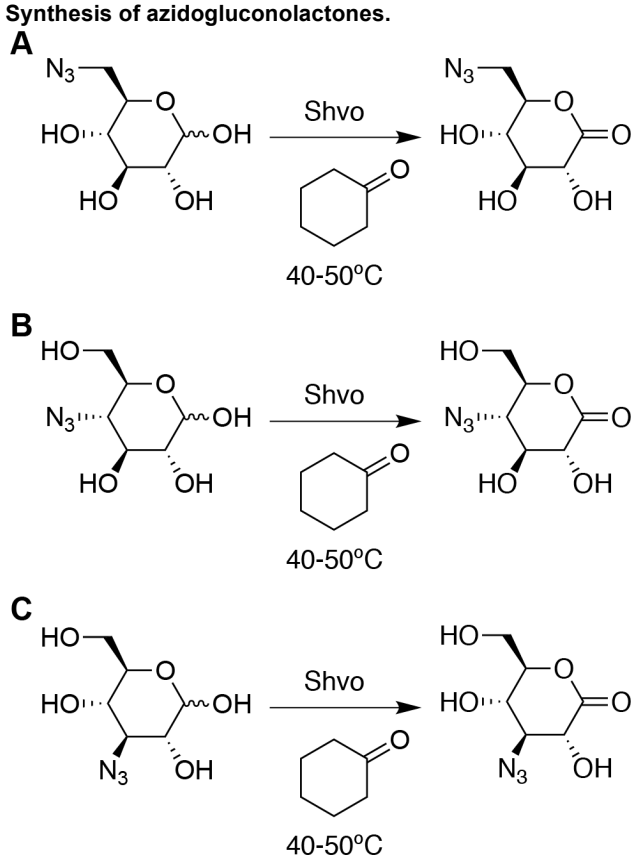
GH_6-heavy-chain-tagged atezolizumab was obtained from Absolute Antibodies (UK) in HEK cells, purified by Ni-NTA, then by size-exclusion chromatography into PBS.

hRBD was obtained from Absolute Antibodies in HEK cells, purified by Ni-NTA, then by size-exclusion chromatography into PBS.

Experimental Section

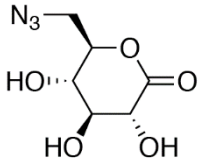
Chemical Synthesis

Synthesis of azidogluconolactones.



Scheme S1. **Chemical synthesis of (A) 6-, (B) 4-, and (C) 3-azido-a-deoxy-D-glucono-1,5-lactone.**

Several azido-glucopyranoses have been synthesized, including 6-[80–88], 4-[89–91], 3-[91,92], and 2-[93] n-azido-/n-deoxy-D-glucose. For oxidation, we resorted to a Shvo's transition metal catalyst[26,94,95](Scheme S1), as in our hands aqueous bromine oxidation of 6-azido-6-deoxy-D-glucose[81,87,96,97] always resulted in syrupy and inseparable mixtures of gamma- (1,4) and delta (1,5)-lactone, and often also the free acid.

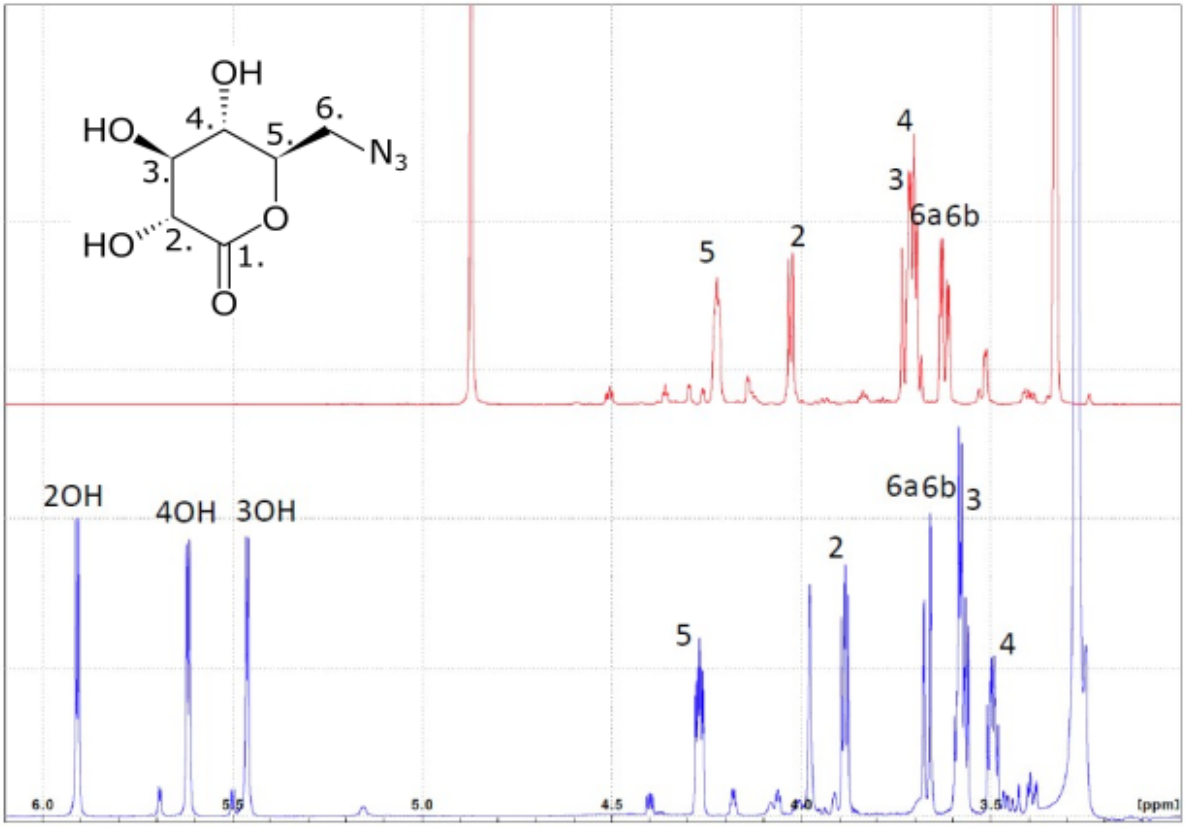


6-azido-6-deoxy-D-glucono-1,5-lactone (6AGDL) – small scale. The compound was synthesized according to a modified procedure for lactonization of aldonolactones[26,94,95] (Scheme S1A). 150 mg of powdered 6-azido-6-deoxy-D-glucose (0.732 mmol, 40 equivalents) were transferred into a conical flask. 20 mg (0.018 mmol, 1 equivalent) of Shvo's catalyst and 10 mL of cyclohexanone were added. The reaction mixture was degassed by bubbling with nitrogen gas for 15 min and then sealed tightly. The degassed reaction mixture was kept overnight constantly stirring at around 45 ºC. A color change from light yellow to dark orange was observed. The reaction mixture was transferred to polypropylene centrifuge tubes and spun for 15 min at 20,000g. A dark brown precipitate of small size was observed at the bottom. The supernatant was aspirated and transferred to a new tube. 3.5 volumes of hexane were added to the supernatant volume. Upon contact and mixing, a whitish precipitate was observed immediately. After carrying out the precipitation for 15 min at RT, the reaction mixture was spun for 5 min at 20,000g and a tight white pellet was observed at the bottom of the microcentrifuge tube. The supernatant was discarded and the wet pellet was resuspended redissolved in 5:1 (v/v) mixture of ethyl acetate/acetone by vortexing. Traces of remaining solids (red brownish colour) were precipitated by spinning for 20 min at 20,000g. The supernatant was transferred to a new tube and spun a second time for 20 min at 20,000g. The supernatant, a colourless liquid, was transferred once more in to a new receiving tube and was left to evaporate at atmospheric pressure and 22 ºC in a microcentrifuge tube in a laminar flow chemical hood at RT to give crystals. Yield (90 mg, 0.44 mmol, 60%). Purity >95 % as determined by $_1$H-NMR. M.p. (crystals) found 114-118ºC, literature 114-116 ºC[97], 124-130 ºC [81,97], and 138-140 ºC[96]; ^{1}H NMR (800 MHz, MeOD, in red, top): δ=4.22 (m, 1H, C5H), 4.02 (d, J=8.4, 1H, C2H), 3.72 (m, 1H, C6Ha), 3.62 (m, 1H, C6Hb), 3.71 (m, 1H, C4H), 3.70 (m, 1H, C3H); ^{1}H NMR (800 MHz, DMSO-d6, in blue, bottom): δ=4.26 (m, J=2.6, 5.5, 9.4, 1H, C5H), 3.90 (dd, J= 5.5, 8.1, 5.0, 1H, C2H), 3.67 (m, 1H, C6Ha), 3.57 (m, J=13.5, 2.5, 1H, C6Hb), 3.60 (m, 1H, C4H), 3.50 (m, 1H, C3H); ^{13}C NMR (200 MHz, MeOD): δ=173.03 (C1), 80.85 (C5), 75.21 (C4), 73.06 (C2), 69.88 (C3), 52.55 (C6); 2D ^{1}H-^{13}C HSQC (MeOD, red): ^{1}H δ=4.22 (m) and ^{13}C δ= 79.5 (C5H); ^{1}H δ=4.02 (d, J=8.4) and ^{13}C δ=71.5 (C2H); ^{1}H δ= 3.62,3.72 (m) and ^{13}C δ=51.1 ($C6H_2$); ^{1}H δ=3.70 (m) and ^{13}C δ=68.4 (C3H); ^{1}H δ=3.71 (m) and ^{13}C δ=73.8 (C4H); 2D ^{1}H-^{13}C HSQC (DMSO-d6, blue): ^{1}H δ=4.26 (m, J=2.6, 5.5, 9.4) and ^{13}C δ=78.1, (C5H); ^{1}H δ=3.90 (dd, J= 5.5, 8.1) and ^{13}C δ=71.5 (C2H); ^{1}H δ=3.67, 3.57 (m, J=13.5, 2.5) and ^{13}C δ=50.8 (C6H); ^{1}H δ=3.50 (m) and ^{13}C δ=68.6 (C3H); ^{1}H δ=3.60 (m) and ^{13}C δ=73.8 (C4H); UPLC-MS: *m/z* calcd for $C_7H_{10}N_3O_7$ [M+HCO^{2-}] 248.1, found 248.1.

Note: Powdered amorphous material with similar NMR spectra was obtained when the last evaporation step was quickly affected with a Concentrator Plus (Eppendorf, 30 ºC, V-AL setting). When syrupy 6-azido-6-deoxy-D-glucose was used as starting material, the resulting product was always of syrupy nature as well. 6AGDL ($C_6H_9N_3O_5$) satisfies the "rule of six" [carbon + oxygen (11) per functional azido group(1)][98], and has a ratio of ≥3 (3.67, sum of carbon + oxygen per azide nitrogen), and may thus considered to be stable. Reasonable care should be taken in the preparation of azidated molecules due to risk of explosion and toxic gas formation - see Presolksi et al. 2011 for practical details and guidance on safety precautions.[46]

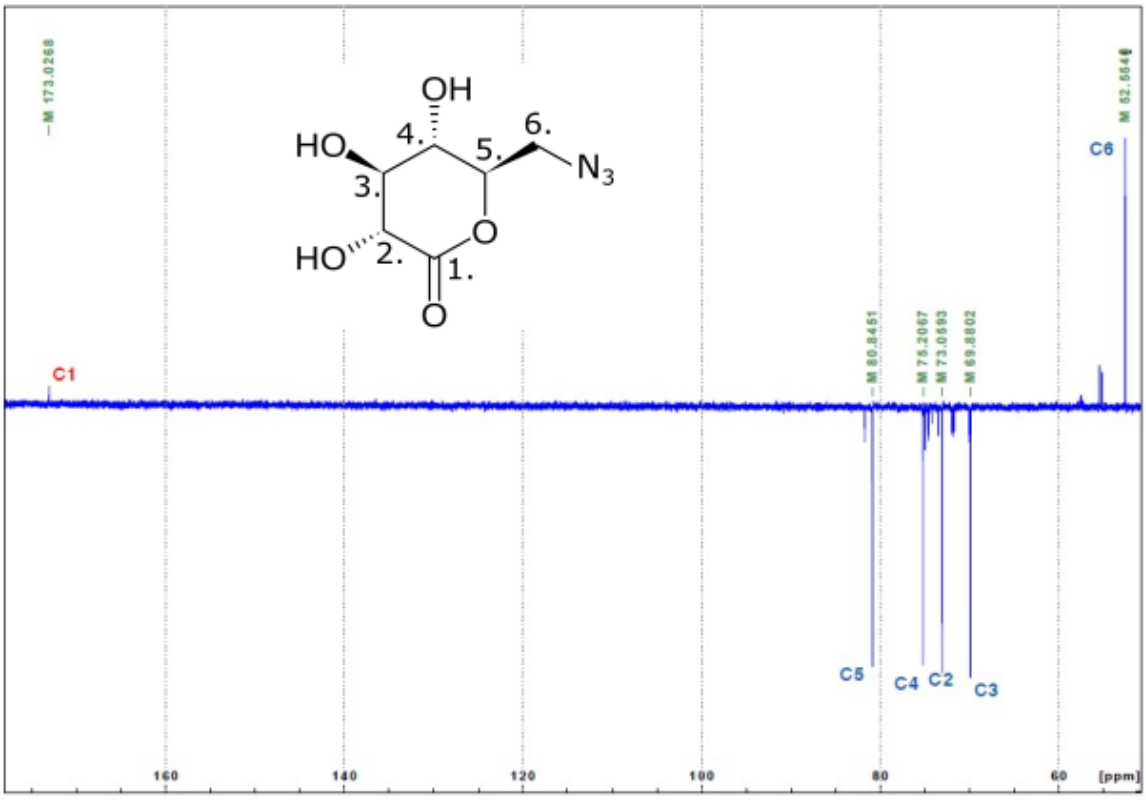
6-azido-6-deoxy-D-glucono-1,5-lactone

[^{1}H NMR, 800 MHz, MeOD, in red, top and DMSO (d6), blue, bottom]



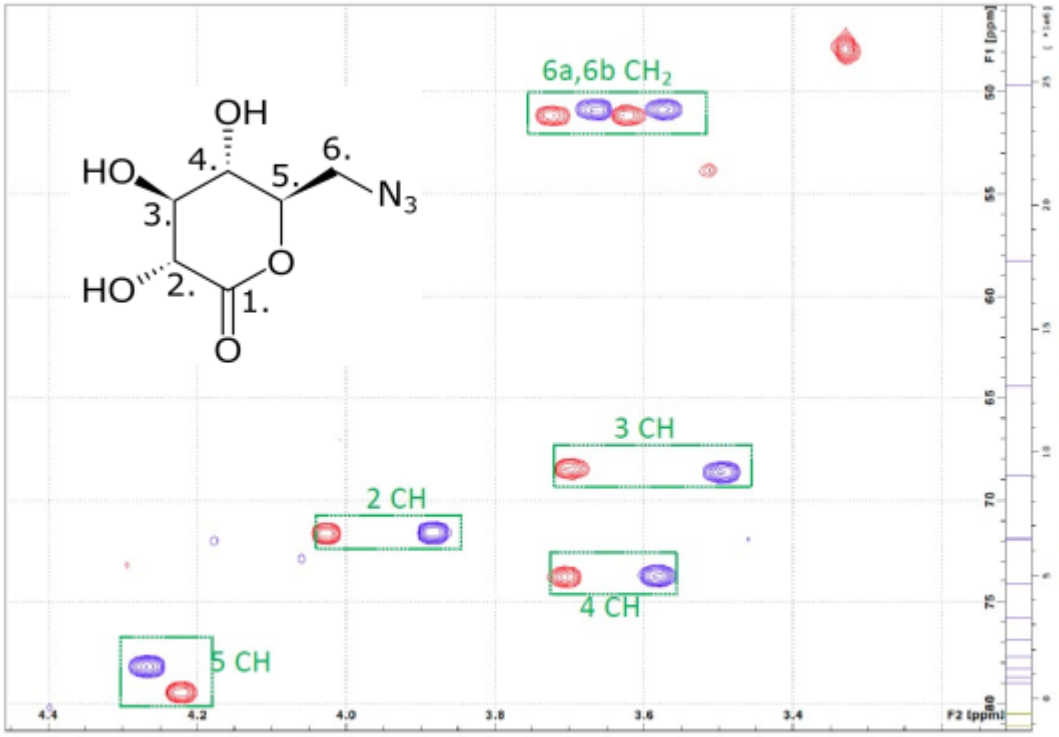
6-azido-6-deoxy-D-glucono-1,5-lactone

[^{13}C NMR, 200 MHz, MeOD]



6-azido-6-deoxy-D-glucono-1,5-lactone

[2D ^{1}H-^{13}C HSQC NMR, MeOD, in red and DMSO (d6), in blue]

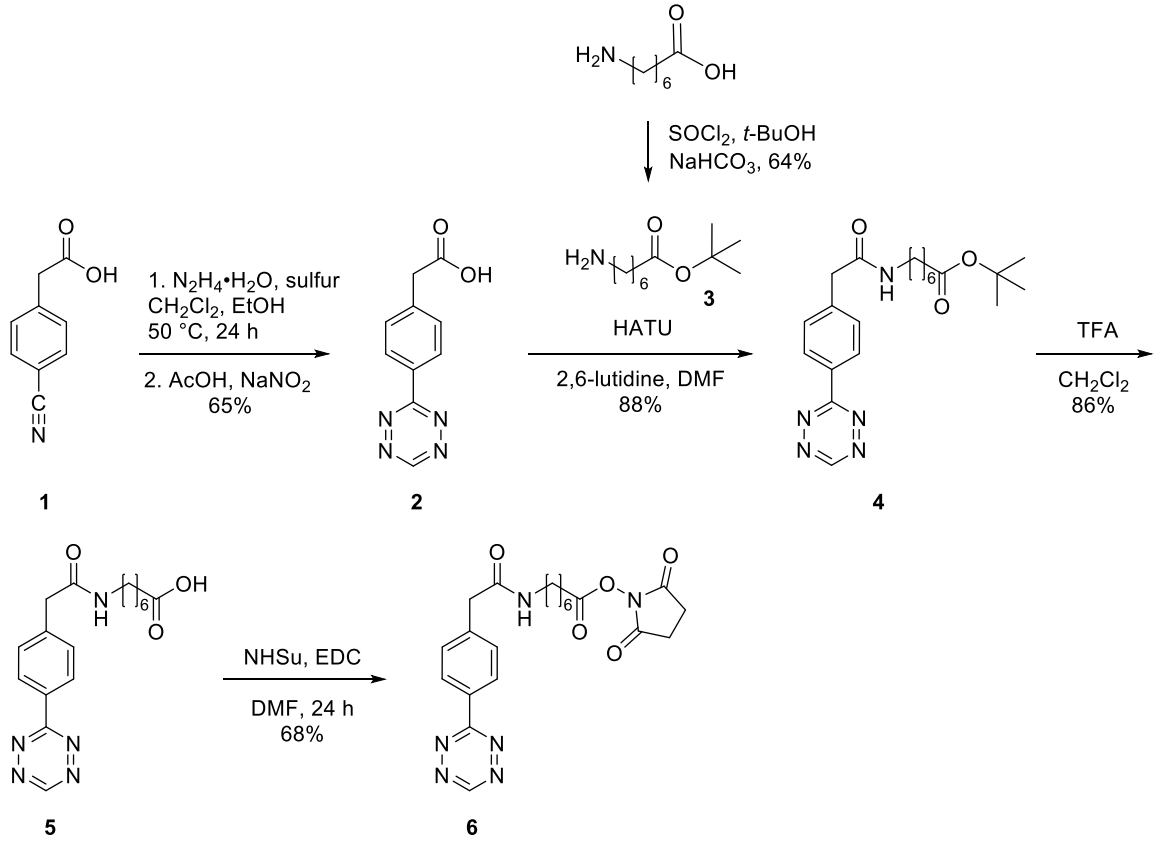


4-azido-4-deoxy-D-glucono-lactone (4AGL) and 3-azido-3-deoxy-D-glucono-lactone (3AGL).

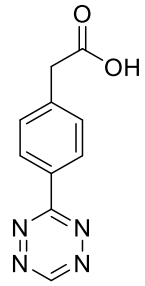
The synthesis of 4AGL and 3AGL was conducted as above for 6AGDL, by oxidation of the corresponding azido-glucoses with Shvo's catalyst in cyclohexanone (Scheme S1B,C). 3-Azido-3-deoxy-1,2:5,6-di-O-isopropylidene-alpha-D-glucofuranose was obtained from Carbosynth Ltd. (UK), deprotected with excess aqueous TFA (1%, v/v) overnight at 60 ºC, followed by lyophilisation to give a syrup. 4-Azido-4-deoxy-D-glucose was obtained from Carbosynth as a syrup. Both compounds were used without any further purification. In both cases syrups were obtained after Shvo mini-scale oxidation, possibly because the starting materials were syrups to begin with (see above for 6AGDL). Lactone formation was indirectly demonstrated by reacting with model peptides to give the expected masses by MALDI TOF MS. We do not know if the 1,5-, the 1,4- lactones, or free acid, or mixtures thereof were obtained. No NMR spectra were acquired for 4- or 3AGL.

Note: We tried to obtain 2-azido-2-deoxy-D-glucono-lactone from the corresponding azido-glucose, but were unable to obtain peptide-reactive material. We suggest that either the 2-azido-2-deoxy-D-glucose is either incompatible with Shvo's catalyst - yielding no lactone, or that 2-azido-2-deoxy-D-glucono-lactone does not react with GDL-reactive model peptides.

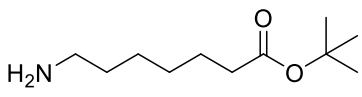
Synthesis of tetrazine-NHS ester.



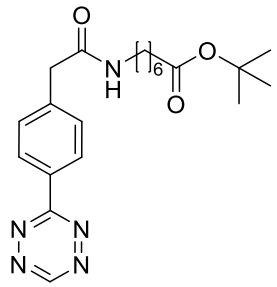
Scheme S2. Synthesis of 2,5-Dioxopyrrolidin-1-yl 7-(2-(4-(1,2,4,5-tetrazin-3-yl)phenyl)acetamido)heptanoate (6).



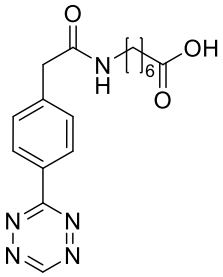
2-(4-(1,2,4,5-Tetrazin-3-yl)phenyl)acetic acid (2). Compound was synthetized following modified literature procedure.[99] Accordingly, a glass pressure tube (120 mL) was charged with 2-(4-cyanophenyl)acetic acid (500 mg, 3.1 mmol, 1 equiv), sulphur (199 mg, 0.78 mmol, 0.25 equiv), CH_2Cl_2 (259 µL, 4.03 mmol, 1.3 equiv) and ethanol (3.1 mL). Then hydrazine monohydrate (1.21 mL, 24.8 mmol, 8 equiv) was added slowly to the well-stirred yellow suspension, which upon addition of hydrazine turned orange. The pressure tube was sealed and heated at 50 °C for 24 h with stirring. The resulting brown mixture was cooled to room temperature, diluted with CH_2Cl_2 (9 mL) and a solution of sodium nitrite (3.21 g, 46.5 mmol, 15 equiv) in water (30 mL) was added. Vigorous stirring at room temperature was continued until the brown mixture turned light yellow. Excess of acetic acid (10.7 mL, 186 mmol, 60 equiv) was added slowly (*caution! vigorous exothermic reaction with the evolution of toxic gas; the reactions should be performed in well ventilated fume hood!*) whereupon the yellow suspension changed colour to bright magenta. The mixture was vigorously stirred until evolution of brown gas ceased and then the reaction mixture was extracted with CH_2Cl_2 (3×20 mL). The organic phase was washed with brine (30 mL), dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo*. The residue was purified by silica gel column chromatography using gradient elution from 1% AcOH in CH_2Cl_2 to 1% AcOH in 95:5 CH_2Cl_2:MeOH). The title compound was obtained as a bright pink solid (435 mg, 2.01 mmol, 65%). ^{1}H NMR (300 MHz, CD_3OD): δ=10.32 (s, 1H), 8.59 – 8.52 (m, 2H), 7.62 – 7.55 (m, 2H), 3.77 (s, 2H). ^{13}C NMR (75 MHz, CD_3OD): δ=174.7, 167.6, 159.2, 141.6, 132.1, 131.5, 129.2, 41.8.



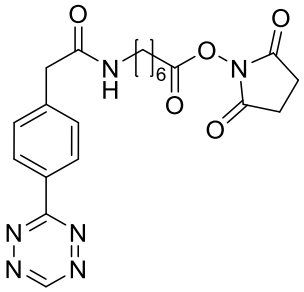
***tert*-Butyl 7-aminoheptanoate (3).** A solution of 7-aminoheptanoic acid (1.00 g, 6.89 mmol, 1 equiv) in $SOCl_2$ (5 mL, 68.9 mmol, 10 equiv) was stirred at room temperature for 1.5 hours, whereupon excess of $SOCl_2$ was removed *in vacuo*. A suspension of $NaHCO_3$ (1.74 g, 20.7 mmol, 3 eq) in *t*-BuOH (16.3 mL, 12.8 mmol, 25 eq) was added to the residue under stirring, and the reaction mixture was stirred overnight at room temperature. The volatiles were removed *in vacuo* and the residue was taken up in EtOAc (90 mL) and aqueous NaOH (1M solution, 67 mL). The organic layer was washed with water (2×40 mL), brine (20 mL), dried over anhydrous Na_2SO_4 and concentrated *in vacuo* to give the title compound as a yellow oil (881 mg, 4.38 mmol, 64%). ^{1}H NMR (300 MHz, $CDCl_3$): δ=2.67 (app. t, *J* = 6.9 Hz, 2H), 2.20 (app. t, *J* = 7.3 Hz, 2H), 1.65 – 1.52 (m, 2H), 1.50 – 1.38 (m, 11H), 1.37 – 1.28 (m, 4H), 1.19 (s, 2H). ^{13}C NMR (75 MHz, CDCl3): δ=173.2, 79.9, 42.1, 35.5, 33.6, 28.9, 28.1, 26.6, 25.0. HRMS (ESI/Q-TOF) (m/z) calcd for $C_{11}H_{24}NO_2$ $[M + H]^+$ 202.1807, found 202.1809.



tert-Butyl 7-(2-(4-(1,2,4,5-tetrazin-3-yl)phenyl)acetamido) heptanoate (4). To a solution of tetrazine **2** (300 mg, 1.39 mmol, 1 equiv) in anhydrous DMF (9 mL) under argon atmosphere at 0 ºC (crushed ice bath) was added 2,6-lutidine (194 µl, 1.66 mmol, 1.2 equiv). After stirring for 5 min, HATU (633 mg, 1.66 mmol, 1.2 equiv) was added and the resulting dark red solution was stirred for more 5 min under argon at 0 ºC, whereupon a solution of amine **3** (405 mg, 2.01 mmol, 1.45 equiv) in dry DMF (3 mL) was added. The resulting dark pink solution was stirred at 0 ºC for 1.5 h under argon atmosphere. Excess of DMF was removed in vacuo and the residue was diluted with EtOAc (20 mL). The suspension was centrifuged and the bright pink supernatant solution was decanted from white precipitates (HOAT). The centrifugation/decantation sequence was repeated 4 times. Combined supernatants were concentrated and product was purified by silica gel column chromatography using gradient elution from 0% EtOAc in hexanes to 100% EtOAc in hexanes. The title product **4** was obtained as a bright pink flakes (489 mg, 1.22 mmol, 88%). ^{1}H NMR (300 MHz, $CDCl_3$): δ= 10.22 (s, 1H), 8.66 – 8.58 (m, 2H), 7.57–7.49 (m, 2H), 5.52 – 5.37 (br s, 1H), 3.67 (s, 2H), 3.28 – 3.20 (m, 2H), 2.17 (t, J = 7.4 Hz, 2H), 1.61 – 1.39 (m, 13H), 1.33 – 1.24 (m, 4H). ^{13}C NMR (75 MHz, CDCl3): δ= 173.3, 169.9, 166.4, 158.0, 140.8, 130.8, 130.5, 129.0, 80.2, 44.0, 39.9, 35.5, 29.4, 28.7, 28.3, 26.6, 25.0. HRMS (ESI/Q-TOF) (m/z) calcd for $C_{21}H_{29}N_5O_3Na$ [M + Na]+ 422.2168, found 422.2177.



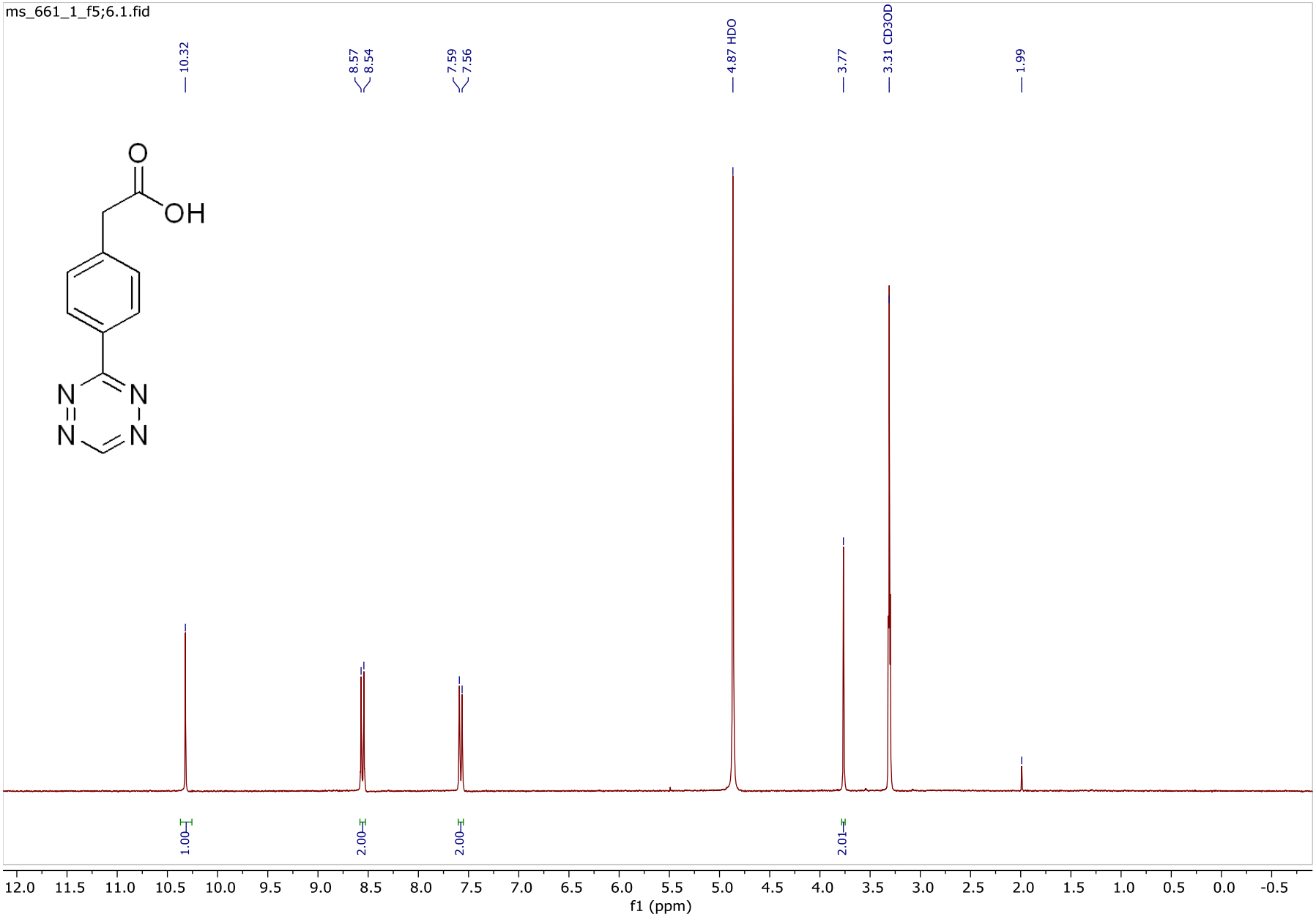
7-(2-(4-(1,2,4,5-Tetrazin-3-yl)phenyl)acetamido)heptanoic acid (5). *tert*-Butyl ester **4** (489 mg, 1.22 mmol, 1 equiv) was dissolved in anhydrous CH_2Cl_2 (25 mL) and TFA (940 µL, 12.2 mmol, 10 equiv) was added dropwise at room temperature. Full conversion of starting material was observed by TLC after stirring for 18 h (R_f=0.45, product **5**; R_f=0.6, starting material **4**; 9:1 $CHCl_3$:MeOH). The bright red mixture was diluted with H_2O (100 mL) and extracted with EtOAc (4× 40 mL). Combined organic extracts were washed with brine (30 mL), dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to give the title compound as light red amorphous solid (360 mg, 1.05 mmol, 86%). *Due to the low solubility, crude product was used in the next step without further purification.* ^{1}H NMR (300 MHz, CD_3OD): δ= 10.31 (s, 1H), 8.60 – 8.50 (m, 2H), 7.63 – 7.53 (m, 2H), 3.63 (s, 2H), 3.24 – 3.17 (m, 2H), 2.28 – 2.20 (m, 2H), 1.63 – 1.27 (m, 8H). ^{13}C NMR (75 MHz, DMSO-d_6): δ= 174.5, 169.2, 165.5, 158.1, 141.9, 130.1, 130.0, 127.7, 42.4, 33.6, 38.6 (overlaps with DMSO signal), 28.9, 28.2, 26.1, 24.4. HRMS (ESI/Q-TOF) (m/z) calcd for $C_{17}H_{22}N_5O_3$ $[M + H]^+$ 344.1723; Found 344.1722.



2,5-Dioxopyrrolidin-1-yl 7-(2-(4-(1,2,4,5-tetrazin-3-yl)phenyl)acetamido)heptanoate (6). 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (46 mg, 0.24 mmol, 1.5 equiv) was added to a solution of carboxylic acid **5** (55 mg, 0.16 mmol, 1 equiv) in anhydrous DMF (1 mL) under argon atmosphere. The resulting dark pink solution mixture was cooled to 0 °C (crushed ice bath), *N*-hydroxysuccinimide (28 mg, 0.24 mmol, 1.5 equiv) was added, the reaction mixture was warmed to room temperature and stirred overnight. To the dark red solution water (10 mL) was added until precipitation of the product was observed. The suspension was centrifuged and the supernatant was removed (the sequence was repeated 4 times). The solid residue was dissolved in $CHCl_3$ (20 mL), dried over anhydrous Na_2SO_4 and concentrated *in vacuo*. Pure product was obtained as red solid (48 mg, 0.11 mmol, 68 %) after flash column chromatography on silica gel using gradient elution from 0% MeOH in CH_2Cl_2 to 5% MeOH in CH_2Cl_2). ^{1}H NMR (300 MHz, $CDCl_3$): δ= 10.21 (s, 1H), 8.65 – 8.55 (m, 2H), 7.58 – 7.48 (m, 2H), 5.73 – 5.57 (br s, 1H), 3.66 (s, 2H), 3.25 (q, *J* = 6.6 Hz, 2H), 2.83 (s, 4H), 2.56 (t, *J* = 7.2 Hz, 2H), 1.77 – 1.64 (m, 2H), 1.54 – 1.24 (m, 6H). ^{13}C NMR (75 MHz, $CDCl_3$): δ= 170.0, 169.4, 168.7, 166.3, 157.9, 140.9, 130.6, 130.5, 128.9, 43.9, 39.6, 30.8, 29.2, 28.2, 26.2, 25.7, 24. 5. HRMS (ESI/Q-TOF) (m/z) calcd for $C_{21}H_{25}N_6O_5$ $[M + H]^+$ 441.1886; Found 441.1902.

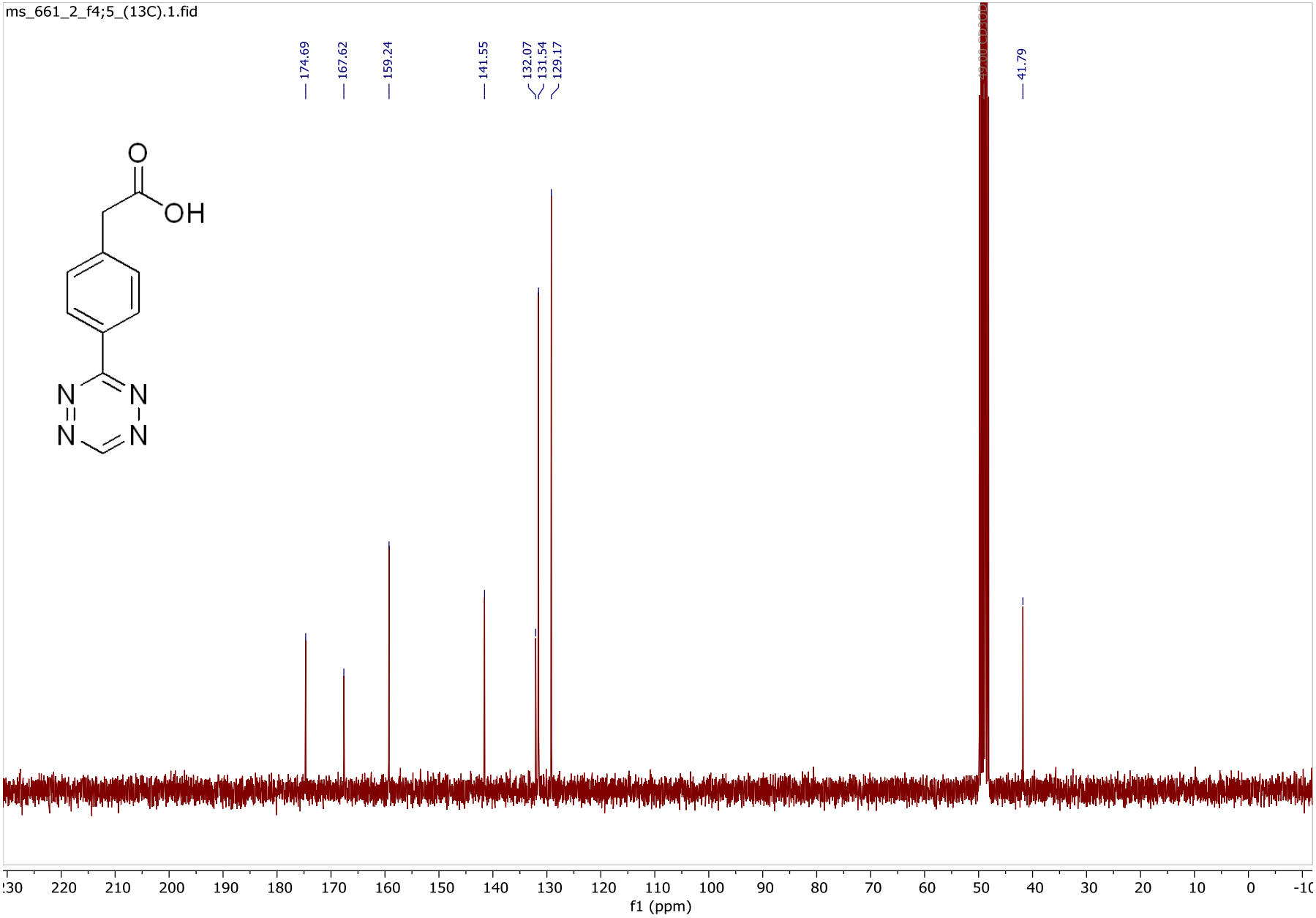
2-(4-(1,2,4,5-Tetrazin-3-yl)phenyl)acetic acid (2)

[^{1}H-NMR, 300 MHz, CD_3OD]



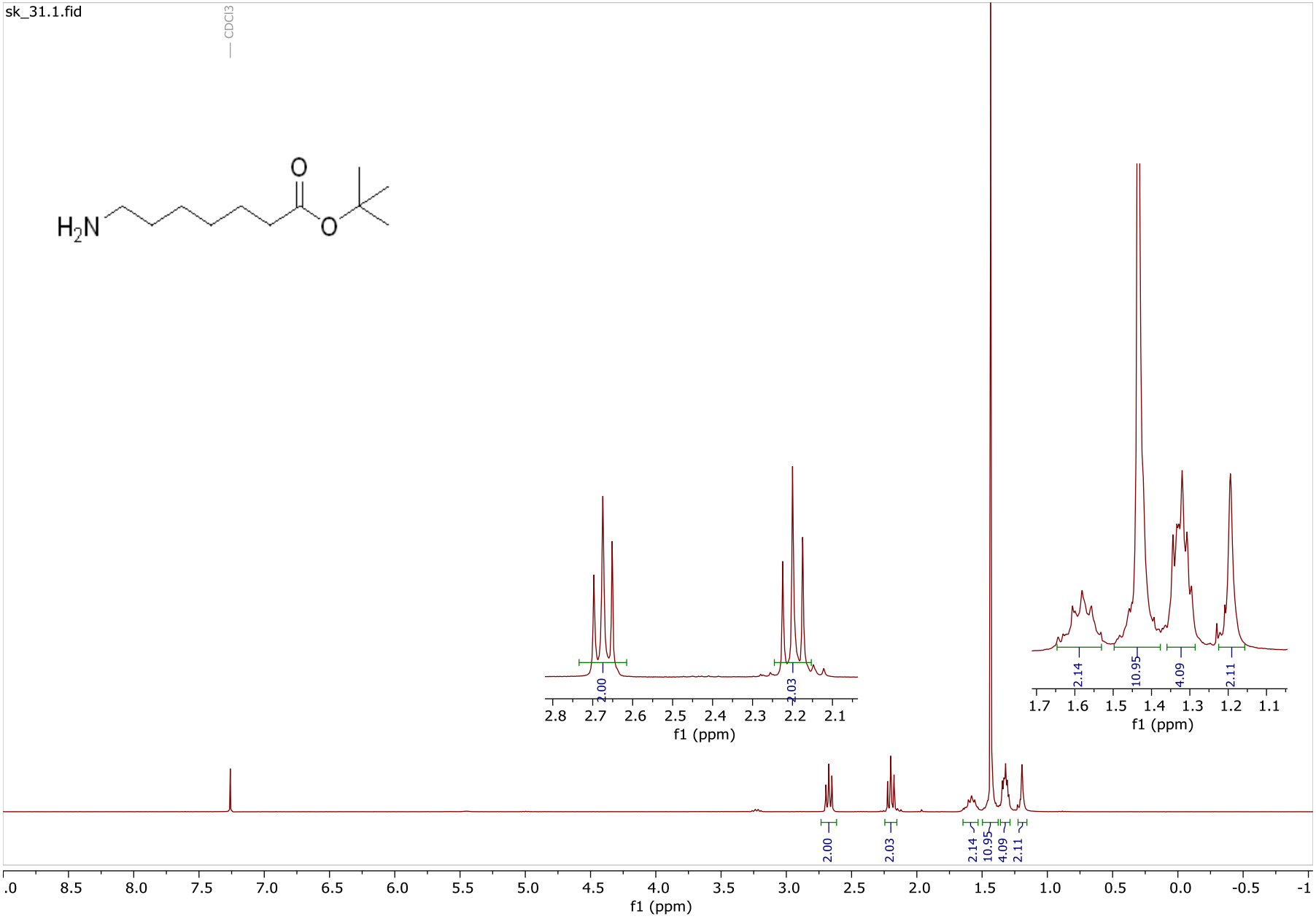
2-(4-(1,2,4,5-Tetrazin-3-yl)phenyl)acetic acid (2)

[^{13}C-NMR, 75 MHz, CD_3OD]



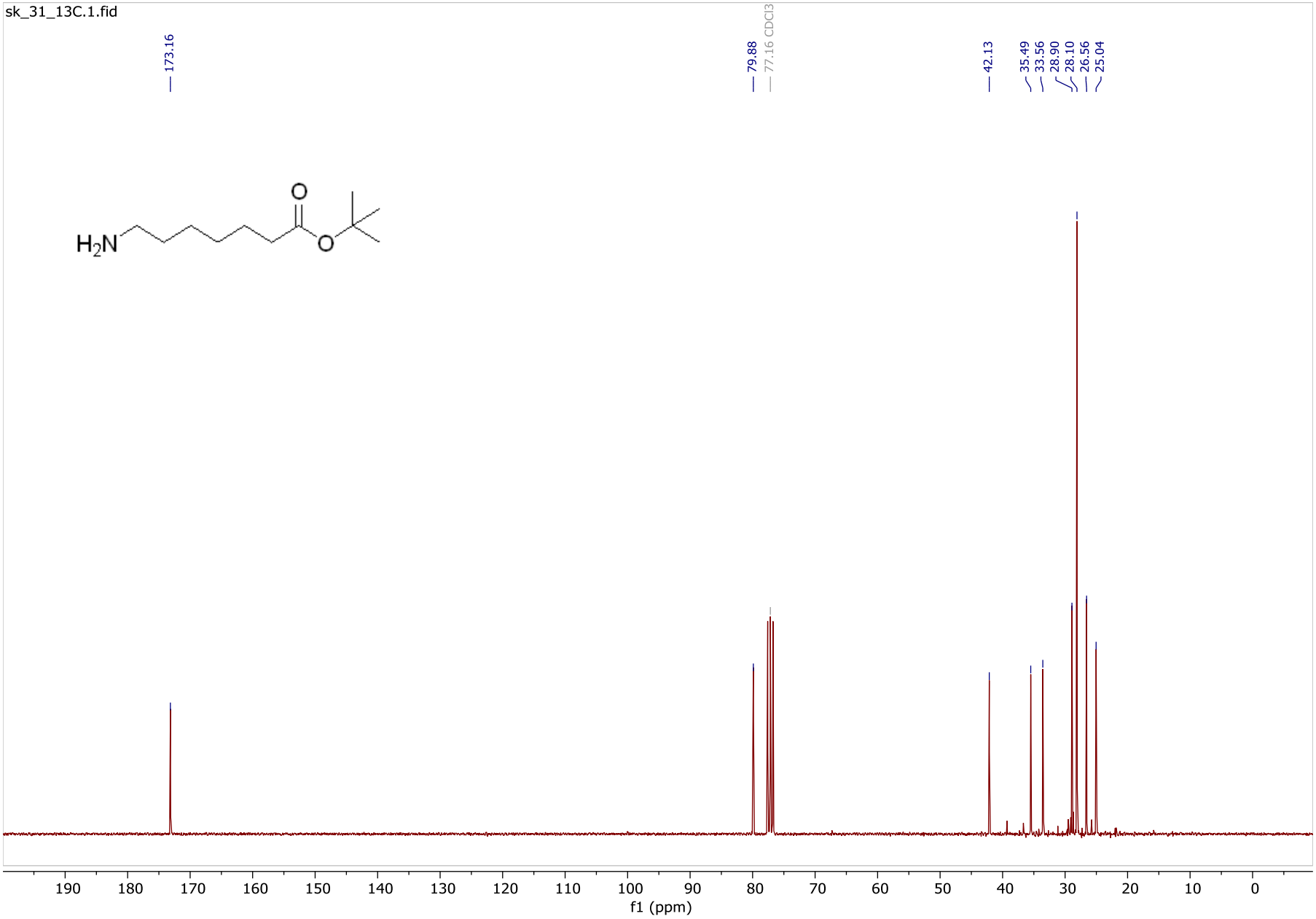
***tert*-Butyl 7-aminoheptanoate (3)**

[^{1}H-NMR, 300 MHz, $CDCl_3$]



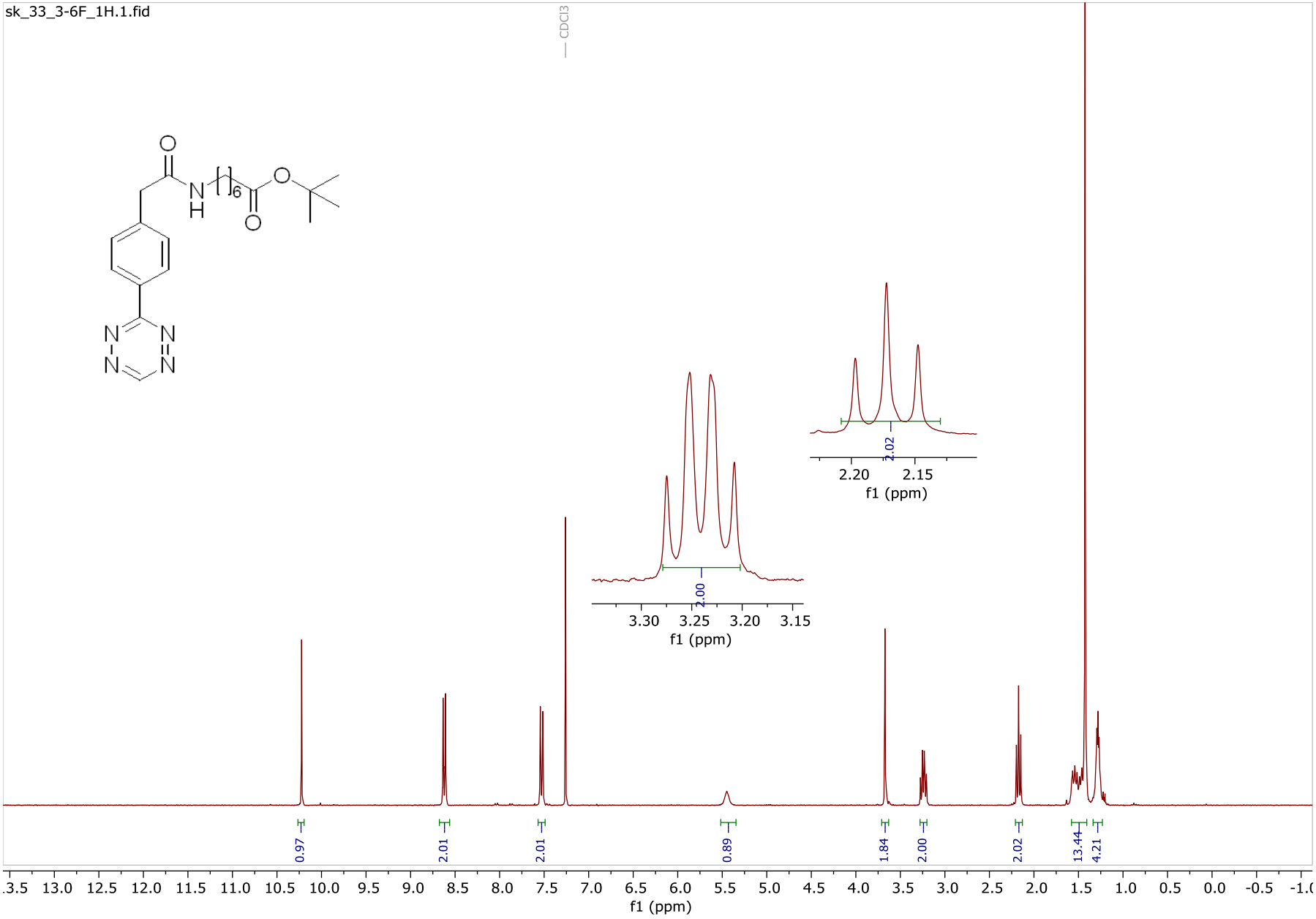
tert-Butyl 7-aminoheptanoate (3)

[^{13}C-NMR, 75 MHz, $CDCl_3$]



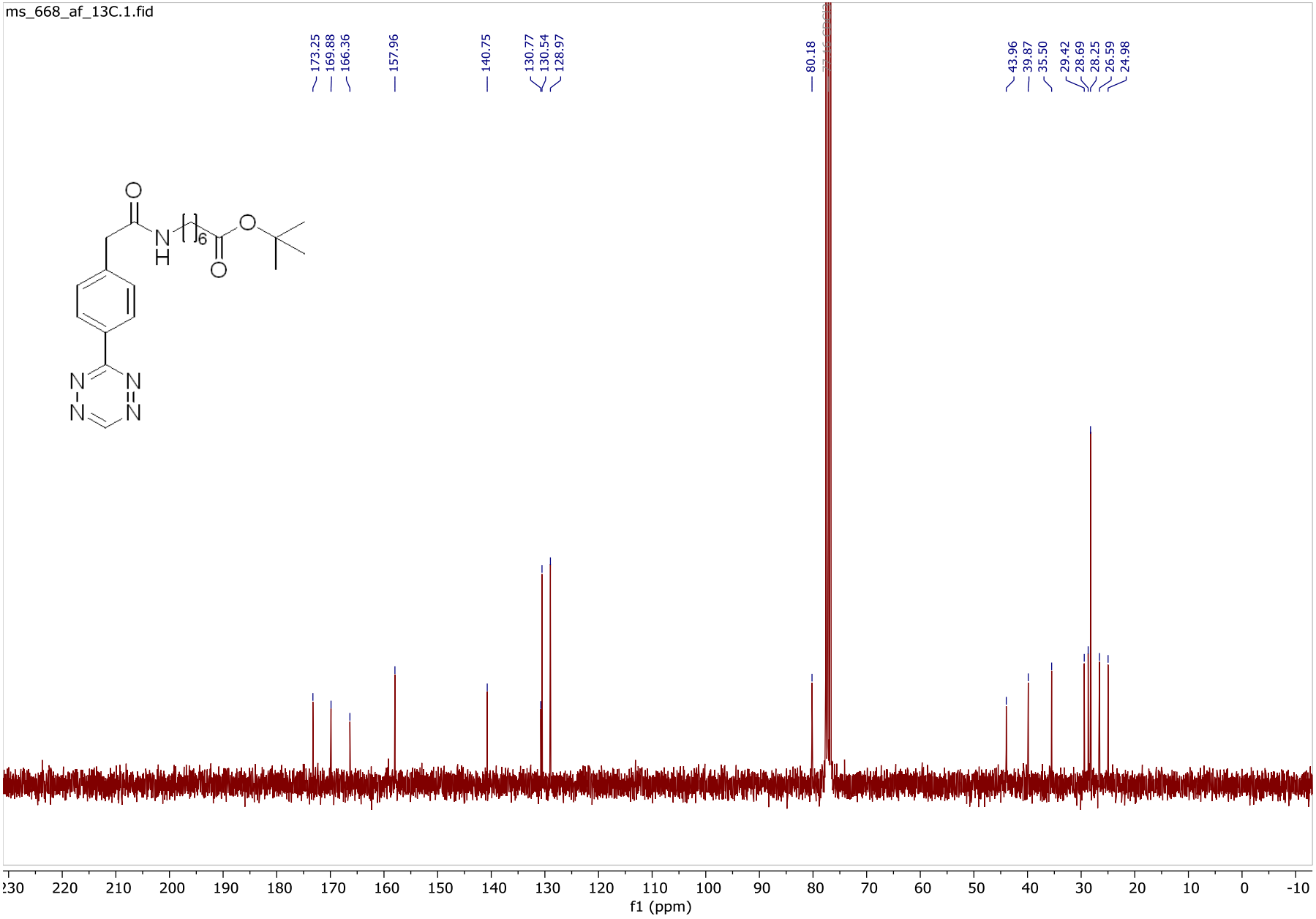
***tert*-Butyl 7-(2-(4-(1,2,4,5-tetrazin-3-yl)phenyl)acetamido)heptanoate (4)**

[^{1}H-NMR, 300 MHz, $CDCl_3$]



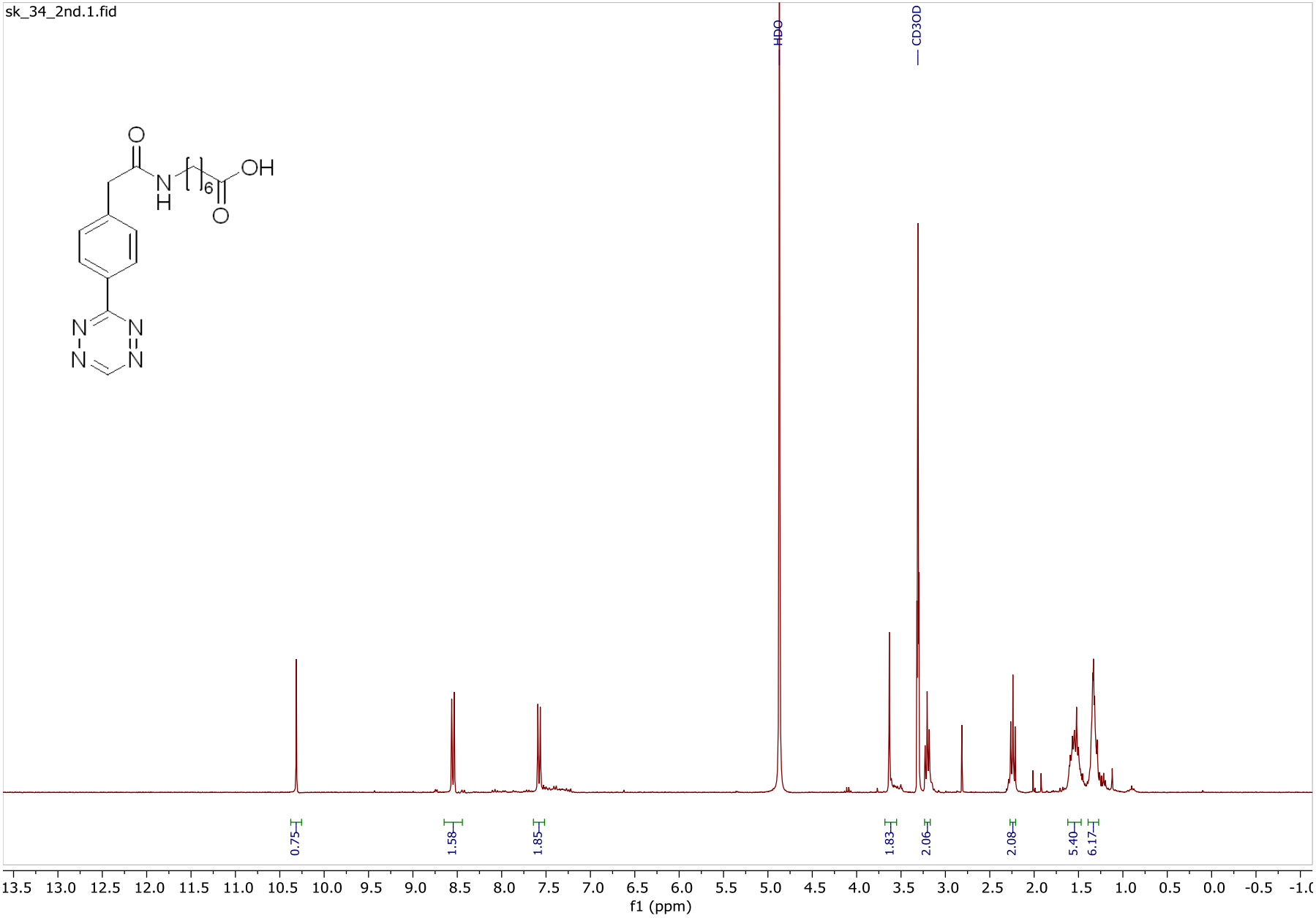
***tert*-Butyl 7-(2-(4-(1,2,4,5-tetrazin-3-yl)phenyl)acetamido)heptanoate (4)**

[^{13}C-NMR, 75 MHz, $CDCl_3$]



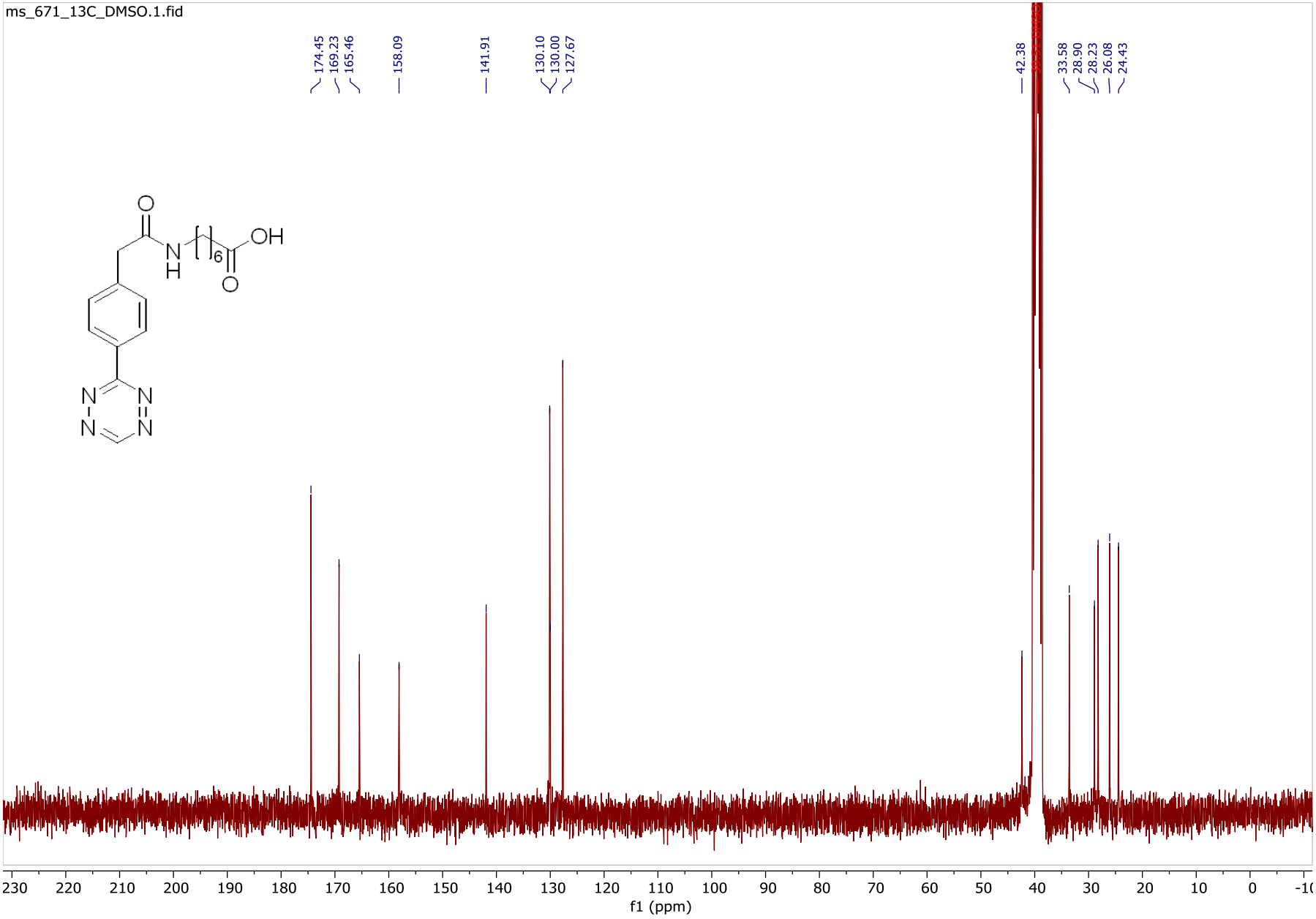
7-(2-(4-(1,2,4,5-Tetrazin-3-yl)phenyl)acetamido)heptanoic acid (5)

[^{1}H-NMR, 300 MHz, CD_3OD]



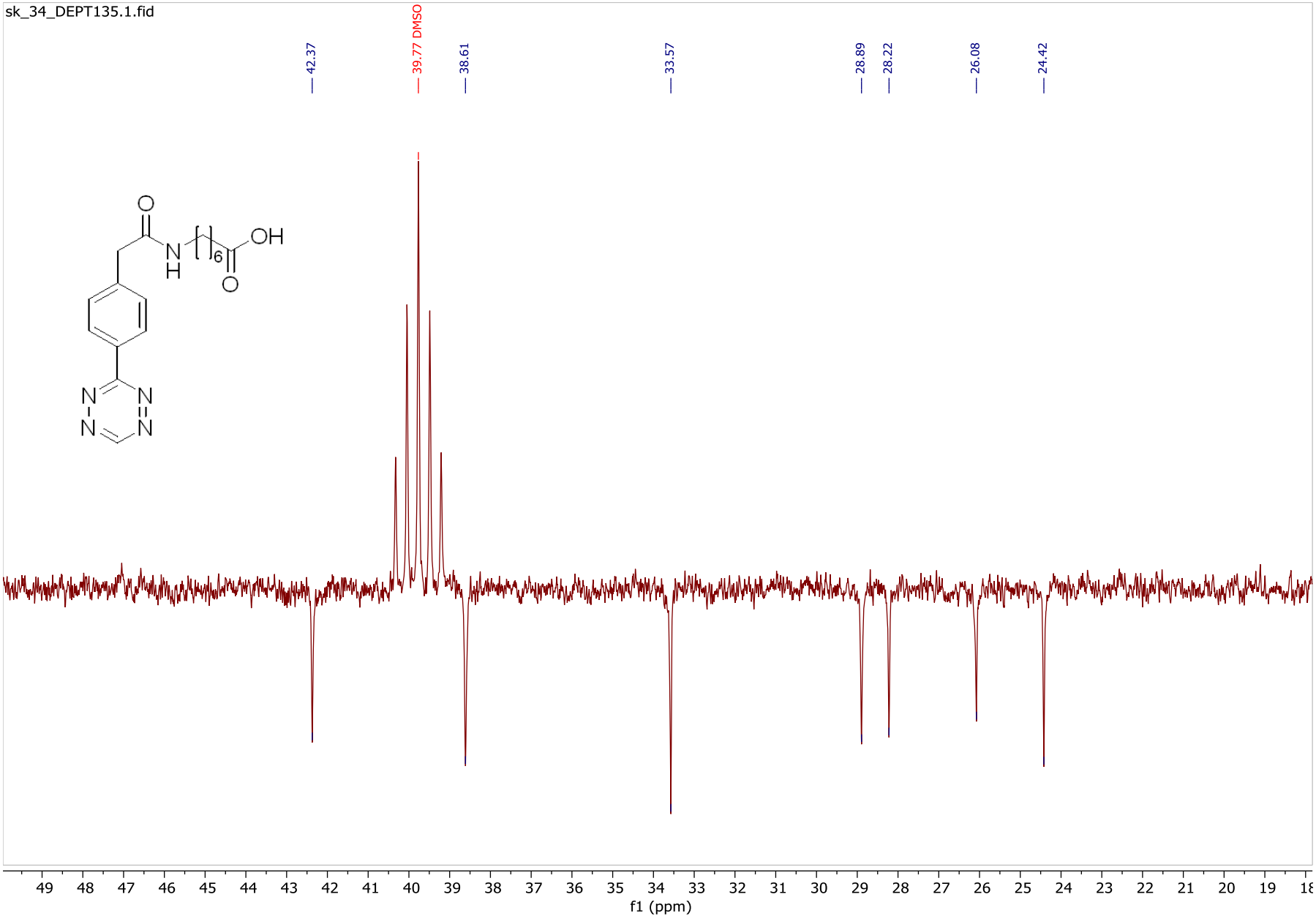
7-(2-(4-(1,2,4,5-Tetrazin-3-yl)phenyl)acetamido)heptanoic acid (5)

[^{13}C-NMR, 75 MHz, DMSO-d_6]



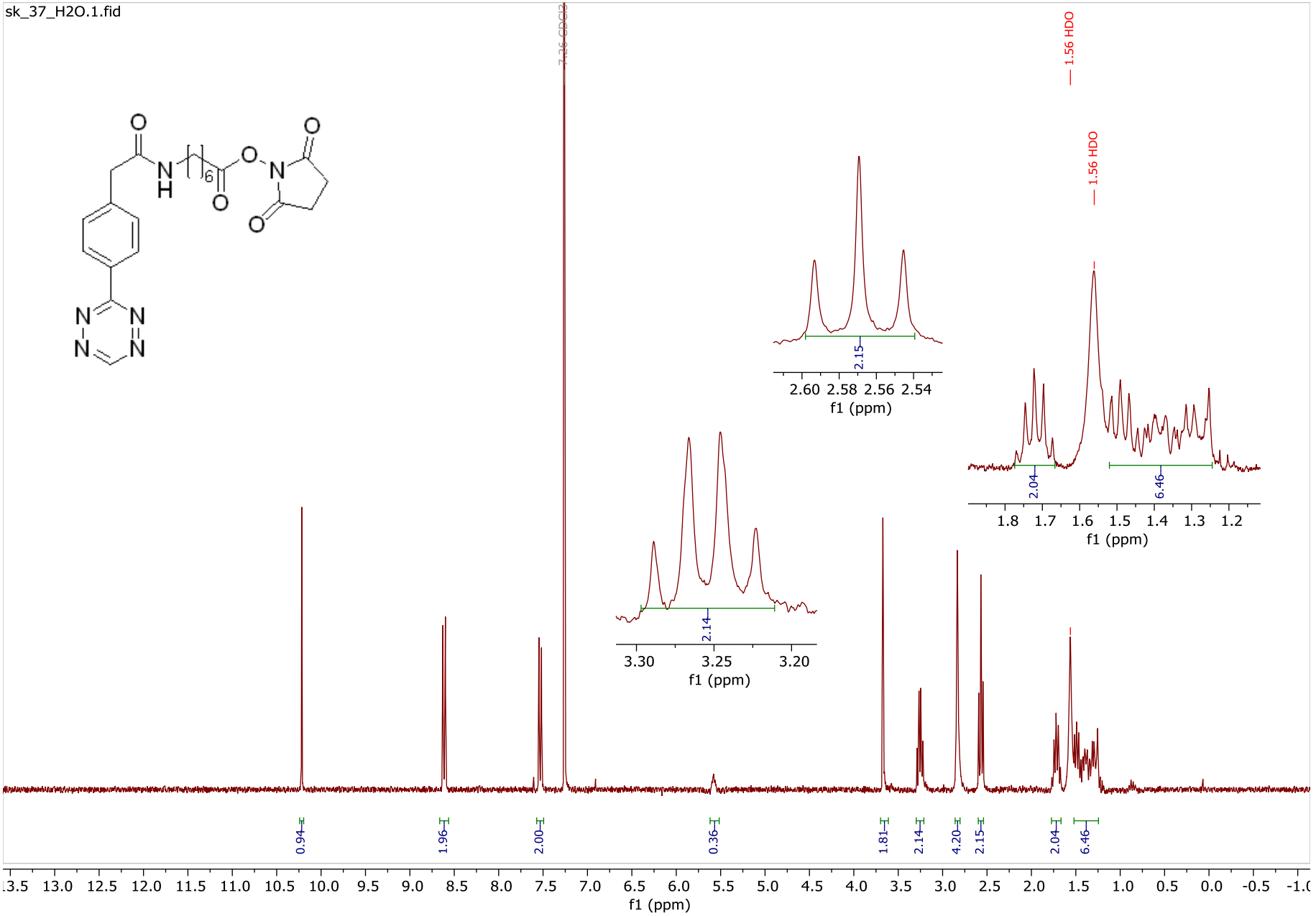
7-(2-(4-(1,2,4,5-Tetrazin-3-yl)phenyl)acetamido)heptanoic acid (5)

[DEPT-135, 75 MHz, DMSO-d_6]



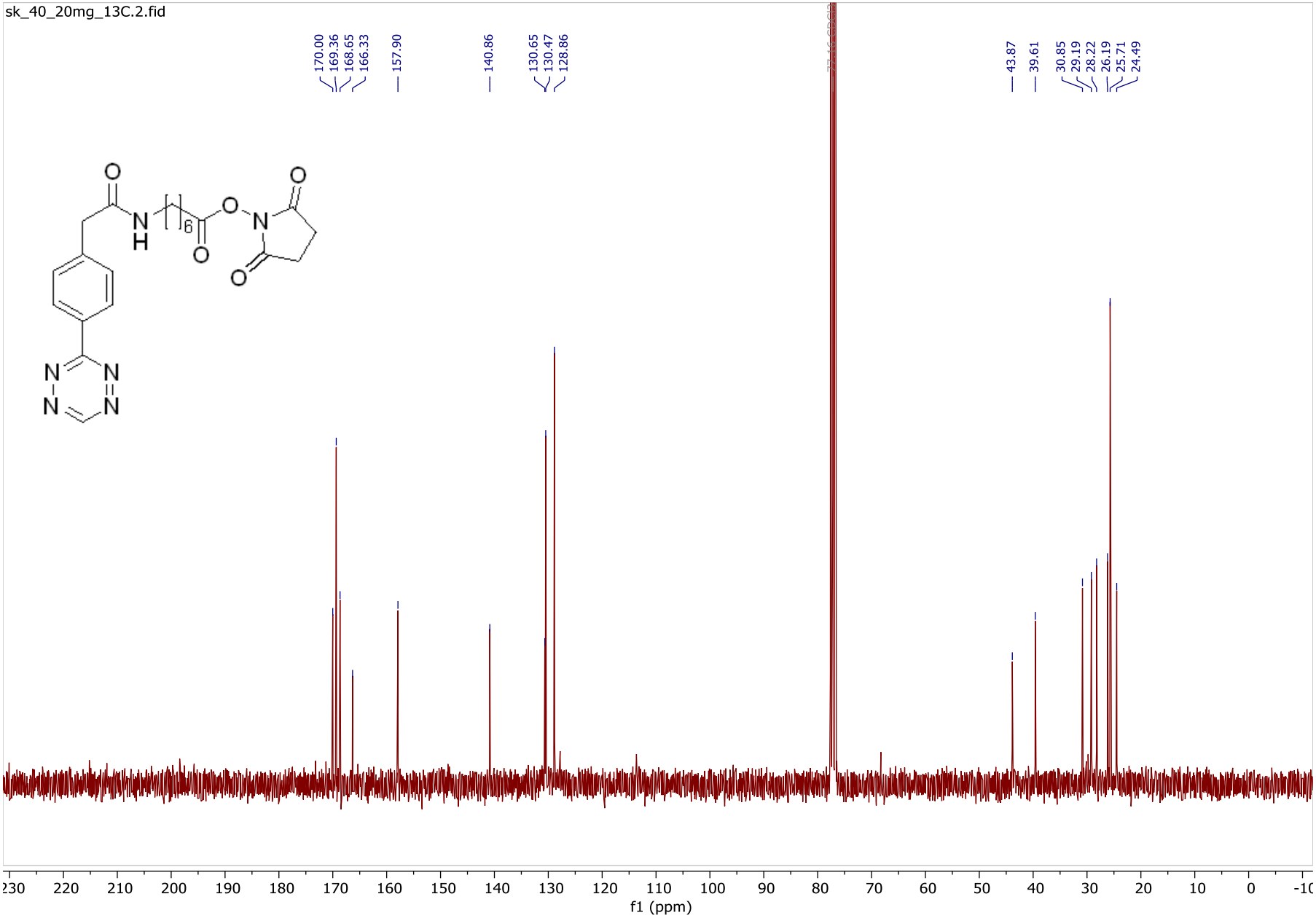
2,5-Dioxopyrrolidin-1-yl 7-(2-(4-(1,2,4,5-tetrazin-3-yl)phenyl)acetamido)heptanoate (6)

[^{1}H-NMR, 300 MHz, D_2O]

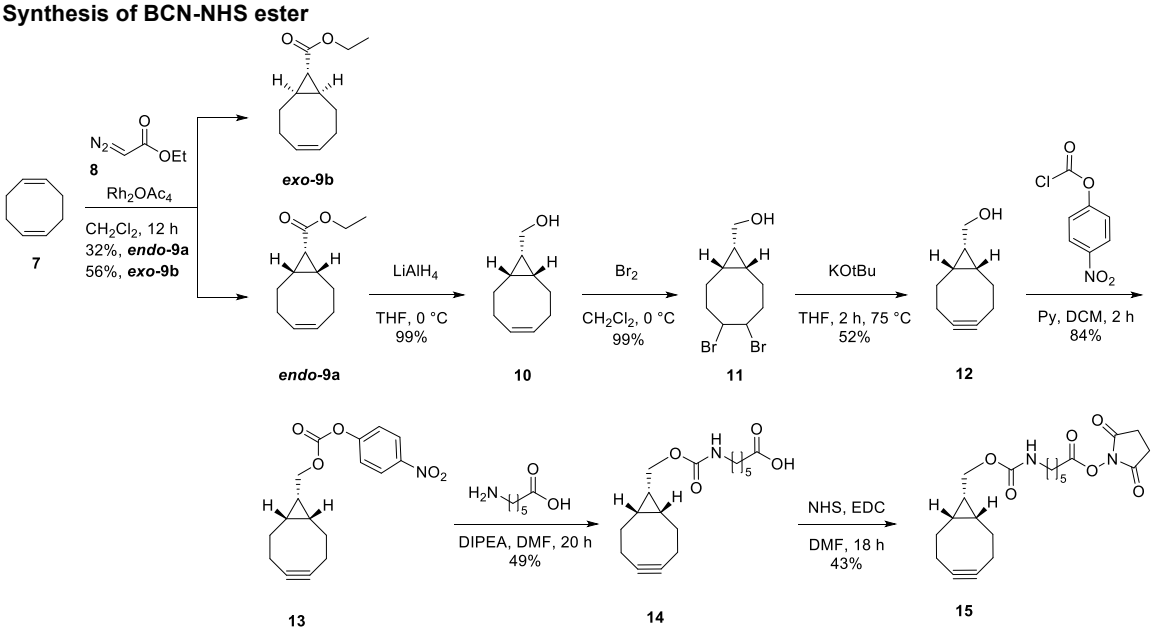


2,5-Dioxopyrrolidin-1-yl 7-(2-(4-(1,2,4,5-tetrazin-3-yl)phenyl)acetamido)heptanoate (6)

[^{13}C-NMR, 75 MHz, $CDCl_3$]

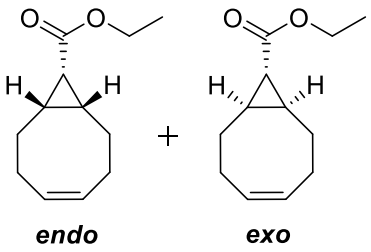


Synthesis of BCN-NHS ester



Scheme S3. **Synthesis of BCN linker.**

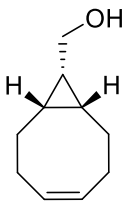
Experimental. Synthesis of BCN linker **15**.[100,101]



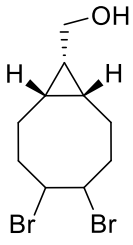
(1*R*,8*S*,9*r*,*Z*)-Ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (*exo*-9b) and (1R,8S,9s,Z)-ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (*endo*-9a). Ethyl diazoacetate (87 wt. % solution in CH_2Cl_2, 2.5 mL, 20.2 mmol, 1 equiv) was added dropwise over 12 hours via syringe pump to a vigorously stirred mixture of 1,5-cyclooctadiene (14.9 mL, 121.24 mmol, 6 equiv) and $Rh_2(OAc)_4$ (45 mg, 0.10 mmol, 0.005 equiv) in anhydrous CH_2Cl_2 (20 mL) under argon atmosphere. Upon complete addition of reagent all volatiles were removed in vacuo and the residual crude mixture of *endo/exo* diastereomers was purified by silica gel flash column chromatography using 2% Et_2O in hexanes as a mobile phase to afford *exo*-9b (2.193 g, 11.29 mmol, 56%) and *endo*-9a (1.264 g, 6.51 mmol, 32%) as colorless oils (combined yield: 3.46 g, 17.79 mmol, 88%).

(1*R*,8*S*,9*s*,*Z*)-Ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (endo-9a). ^{1}H NMR (300 MHz, $CDCl_3$): δ= 5.68 – 5.55 (m, 2H), 4.12 (q, J = 7.1 Hz, 2H), 2.56 – 2.44 (m, 2H), 2.27 – 2.13 (m, 2H), 2.12 – 1.99 (m, 2H), 1.88 – 1.76 (m, 2H), 1.70 (t, J = 8.8 Hz, 1H), 1.45 – 1.33 (m, 2H), 1.26 (t, J = 7.1 Hz, 3H). ^{1}H NMR spectrum was in agreement with that reported in the literature.[101]

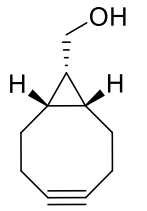
(1*R*,8*S*,9*r*,*Z*)-Ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (exo-9b). ^{1}H NMR (300 MHz, $CDCl_3$): δ= 5.69 – 5.55 (m, 2H), 4.10 (q, J = 7.1 Hz, 2H), 2.38 – 1.98 (m, 6H), 1.61 – 1.41 (m, 4H), 1.25 (t, J = 7.1 Hz, 3H), 1.18 (t, J = 4.6 Hz, 1H). ^{1}H NMR spectrum was in agreement with that reported in the literature.[101]



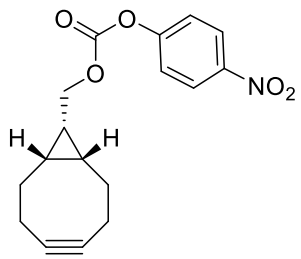
1*R*,8*S*,9*s*,*Z*)-Bicyclo[6.1.0]non-4-en-9-ylmethanol (10). A solution of *endo*–**9a** (1.264 g, 6.51 mmol, 1 equiv) in anhydrous THF (30 mL) under argon atmosphere was cooled to 0 °C (crushed ice bath) and $LiAlH_4$ (4 M solution in THF, 1.71 mL, 6.83 mmol, 1.05 equiv) was added dropwise. Full conversion of the starting material was observed by TLC after stirring for 1 h at room temperature (R_f=0.2, product **10**, R_f=0.7, starting material **9a**; 1:4 EtOAc:petroleum ether; *p*-Anisaldehyde or $KMnO_4$ stain for visualization). Then reaction mixture was cooled down to 0 °C, upon which water (2 mL) was added carefully until evolution of gas has ceased (*caution!*). Anhydrous Na_2SO_4 (5 g) was added to the well-stirred suspension, the resulting solid was filtered off and washed thoroughly with THF. Combined filtrates were concentrated *in vacuo* to afford alcohol **10** as a colorless oil (991 mg, 6.51 mmol, 99%) that was used in the next step without further purification. ^{1}H NMR (300 MHz, $CDCl_3$): δ= 5.70 – 5.56 (m, 2H), 3.71 (dd, *J* = 7.5, 5.1 Hz, 2H), 2.43 – 2.27 (m, 2H), 2.18 – 1.91 (m, 4H), 1.66 – 1.48 (m, 2H), 1.20 – 0.93 (m, 4H).



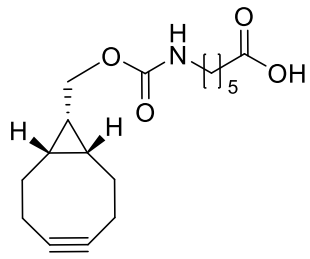
(1*R*,8*S*,9*s*)-4,5-Dibromobicyclo[6.1.0]nonan-9-yl)methanol (11). A solution of Br_2 (370 µL, 7.22 mmol, 1.1 equiv) in CH_2Cl_2 (5 mL) was added dropwise under argon atmosphere to a solution of crude alcohol **10** (991 mg, 6.51 mmol, 1 equiv) in anhydrous CH_2Cl_2 (30 mL) at 0 °C (crushed ice bath) until red color persisted. After warming to room temperature and stirring for 15 min, the dark reaction mixture was quenched with aqueous 10% $Na_2S_2O_3$ solution (12 mL). Layers were separated and the aqueous phase was extracted with CH_2Cl_2 (3×15 mL). Combined organic extracts were washed with brine (15 mL), dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to yield dibromide **11** as a white flakes (2.031 g, 6.51 mmol, 99%), which was used without further purification. ^{1}H NMR (300 MHz, $CDCl_3$): δ= 4.88 – 4.76 (m, 2H), 3.76 (d, *J* = 7.3 Hz, 2H), 2.79 – 2.58 (m, 2H), 2.36 – 2.09 (m, 2H), 2.01 – 1.84 (m, 2H), 1.72 – 1.47 (m, 2H), 1.28 – 1.03 (m, 4H). ^{13}C NMR (75 MHz, $CDCl_3$): δ= 59.8, 56.3, 53.4, 35.1, 22.0, 20.3, 20.1, 19.1, 17.3.



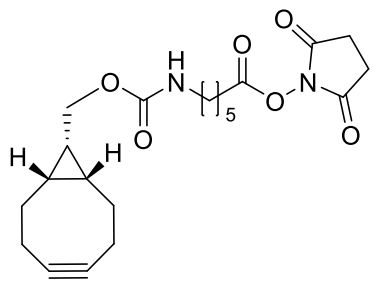
(1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-ylmethanol (12). A solution of KOtBu (1M in THF, 11.22 mL, 1.22 mmol, 3.5 equiv) was added dropwise at 0 °C (crushed ice bath) to a solution of dibromide **11** (1 g, 3.21 mmol, 1 equiv) in anhydrous THF (30 mL) under argon atmosphere, until the suspension turned brown. Then, the reaction mixture was stirred under reflux for 2 h (75 °C), cooled to room temperature, quenched with aqueous sat'd NH_4Cl solution (30 mL) and extracted with CH_2Cl_2 (4×20 mL). Combined organic extracts were dried over Na_2SO_4, filtered and concentrated *in vacuo*. Crude product was purified by silica gel flash column chromatography using gradient elution from 25% Et_2O in hexanes to 50% Et_2O in hexanes, to afford product **12** as a yellowish solid (251 mg, 1.67 mmol, 52%). ^{1}H NMR (300 MHz, $CDCl_3$): δ= 3.73 (d, *J* = 7.9 Hz, 2H), 2.38 – 2.15 (m, 6H), 1.69 – 1.49 (m, 2H), 1.44 – 1.27 (m, 1H), 1.26 – 1.13 (m, 1H), 1.00 – 0.87 (m, 2H). ^{13}C NMR (75 MHz, $CDCl_3$): δ= 99.0, 60.0, 29.1, 21.6, 21.5, 20.1. ^{1}H and ^{13}C NMR spectrum was in agreement with that reported in the literature.[101]



((1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-yl)methyl (4-nitrophenyl) carbonate (13). Compound **13** was synthesized following a modified literature procedure.[102] Accordingly, a solution of 4-nitrophenylchloroformate (361 mg, 1.79 mmol, 1.2 equiv) in anhydrous CH_2Cl_2 (4 mL) was added to a solution of alcohol **12** (224 mg, 1.49 mmol, 1 equiv) in anhydrous CH_2Cl_2 (15 mL) under argon atmosphere. Then, anhydrous pyridine (300 µL, 3.71 mmol, 2.5 equiv) was added dropwise and colorless solution stirred for 2 hours. After full conversion of starting material (observed by TLC; R_f=0.25, product **13**, R_f=0.05, starting material **12**; 1:5 Et_2O:hexanes; $KMnO_4$ stain for visualization) the colorless solution was quenched with aqueous saturated NH_4Cl solution (35 mL). Layers were separated and aqueous layer was extracted with CH_2Cl_2 (4×15 mL). The combined organic extracts were washed with brine (20 mL), dried over Na_2SO_4 and filtered. Solvent was removed *in vacuo* and crude product was purified by silica gel flash column chromatography using gradient elution from 5% Et_2O in hexanes to 20% Et_2O in hexanes to give the title compound **13** as a white solid (393 mg, 1.25 mmol, 84%). ^{1}H NMR (300 MHz, $CDCl_3$): δ= 8.31 – 8.24 (m, 2H), 7.41 – 7.35 (m, 2H), 4.40 (d, *J* = 8.3 Hz, 2H), 2.3 – 2.19 (m, 6H), 1.68 – 1.44 (m, 3H), 1.12 – 0.99 (m, 2H). ^{13}C NMR (75 MHz, $CDCl_3$): δ= 155.7, 152.7, 145.5, 125.4, 121.9, 98.8, 68.1, 29.2, 21.5, 20.6, 17.4. ^{1}H NMR spectrum was in agreement with that reported in the literature.[101]



6-(((((1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbonyl)amino)hexanoic acid (14). Compound 14 was synthesized following a modified literature procedure.[103] Thus, 6-aminocaproic acid (163 mg, 1.25 mmol, 1.25 equiv) was finely ground in a mortar and added to a colorless solution nitrophenyl carbonate 13 (314 mg, 1 mmol, 1 equiv) in anhydrous DMF (13 mL). DIPEA (520 µL, 3.00 mmol, 3 equiv) was added dropwise to the yellow suspension, whereupon the suspension gradually turned to solution. After stirring for 12 h, yellow cloudy solution was diluted with water (100 mL) (*slight exotherm!*) and washed with CH_2Cl_2 (2×25 mL). Clear yellow organic extract was discarded. The aqueous layer was acidified until pH 2-3 by the addition of acetic acid solution in water (5% v/v), whereupon the yellow colour disappeared. The aqueous layer was extracted with CH_2Cl_2 (4×30 mL), combined organic layers were washed with brine (30 mL), dried over $MgSO_4$ and concentrated *in vacuo*. Crude product was purified by silica gel flash column chromatography using gradient elution from 0.5% AcOH in CH_2Cl_2 to 0.5% AcOH in 9:1 CH_2Cl_2:EtOAc. After azeotropic drying with *n*-heptane to remove traces of AcOH and DMF, product 14 was obtained as a colorless viscous oil (149 mg, 0.48 mmol, 49%). ^{1}H NMR (300 MHz, $CDCl_3$, major rotamer signals): δ= 4.97 – 4.59 (br s, 1H), 4.13 (d, *J* = 8.2 Hz, 2H), 3.24 – 3.08 (m, 2H), 2.34 (t, *J* = 7.3 Hz, 2H), 2.30 – 2.12 (m, 6H), 1.72 – 1.46 (m, 6H), 1.45 – 1.28 (m, 3H), 1.01 – 0.86 (m, 2H). ^{13}C NMR (75 MHz, $CDCl_3$): δ= 179.1, 156.9, 98.9, 62.8, 40.9, 34.0, 29.8, 29.2, 26.3, 24.4, 21.5, 20.2, 17.9. ^{1}H and ^{13}C NMR spectrum was in agreement with that reported in the literature.[103]

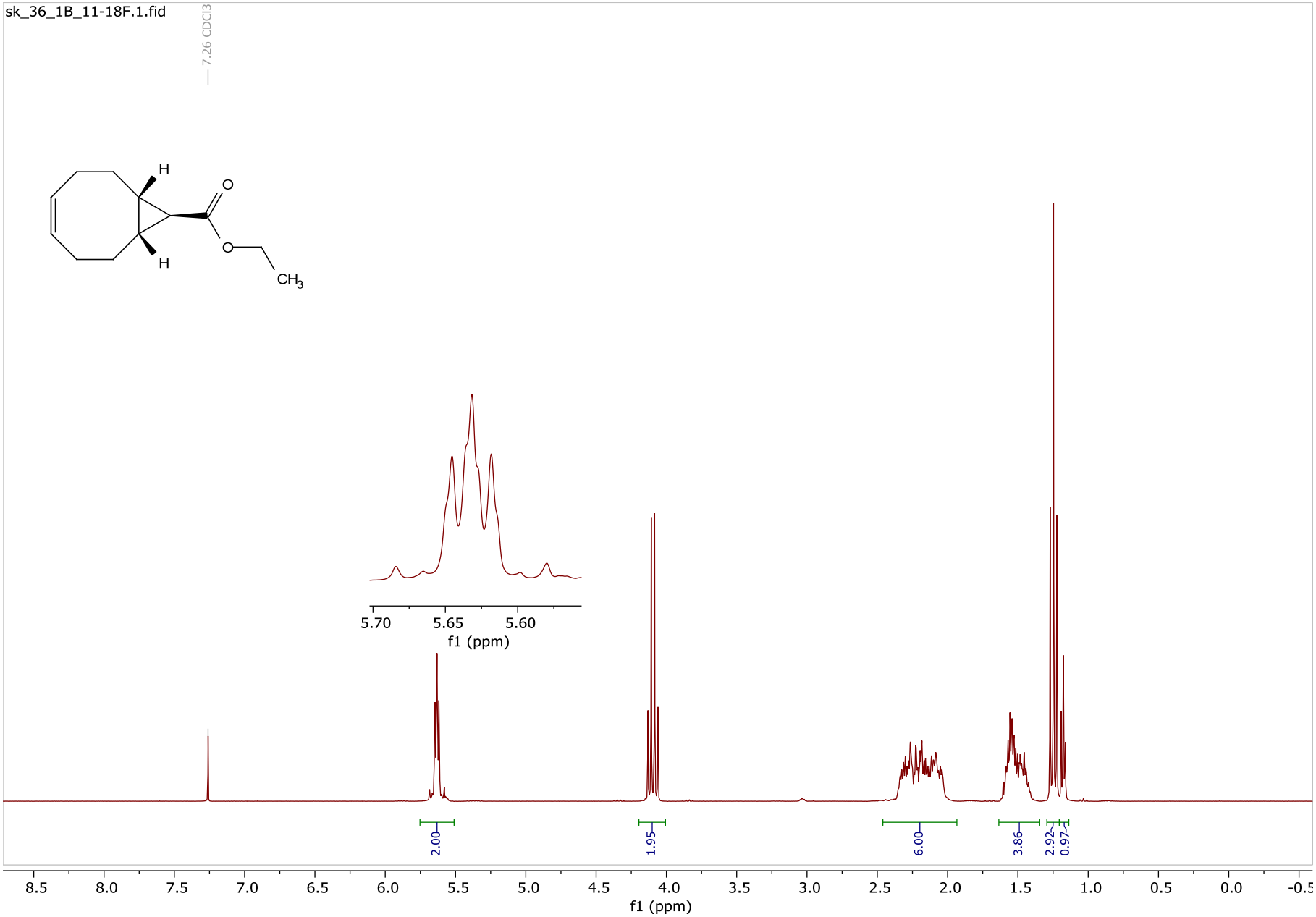


2,5-Dioxopyrrolidin-1-yl 6-(((((1*R*,8*S*,9*s*)-bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbonyl)amino)hexanoate (15)

1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (102 mg, 0.53 mmol, 1.3 equiv) was added to a solution of carboxylic acid **14** (126 mg, 0.41 mmol, 1 equiv) in anhydrous DMF (3.5 mL) under argon atmosphere at 0 °C (crushed ice bath). The resulting suspension was stirred for 15 min at 0 °C, then *N*-hydroxysuccinimide (66 mg, 0.57 mmol, 1.4 equiv) was added. After stirring for 10 min and warming to room temperature, a clear solution was formed, which was stirred for 18 h and concentrated *in vacuo* to remove excess DMF. The residue was dissolved in CH_2Cl_2 (20 mL), washed with cold water (3×10 mL), then with brine (2×10 mL). Organic phase was dried over $MgSO_4$ and concentrated *in vacuo*. Crude product was purified by silica gel flash column chromatography using gradient elution from 5% EtOAc in DCM to 20% EtOAc in DCM, to afford the title compound **15** as a viscous colorless oil (72 mg, 0.18 mmol, 43%). ^{1}H NMR (300 MHz, $CDCl_3$): δ= 4.88 – 4.60 (br s, 1H), 4.10 (d, *J* = 8.0 Hz, 2H), 3.20 – 3.09 (m, 2H), 2.85 – 2.75 (m, 4H), 2.58 (t, *J* = 7.3 Hz, 2H), 2.33 – 2.11 (m, 6H), 1.74 (p, *J* = 7.4 Hz, 2H), 1.65 – 1.25 (m, 7H), 0.98 – 0.83 (m, 2H). ^{13}C NMR (75 MHz, $CDCl_3$): δ= 169.3, 168.5, 156.8, 98.9, 62.6, 40.7, 30.9, 29.5, 29.1, 25.8, 25.7, 24.3, 21.5, 20.2, 17.8. HRMS (ESI/Q-TOF) (m/z) calcd for $C_{21}H_{29}N_2O_6$ $[M + H]^+$ 405.2026, found 405.2027.

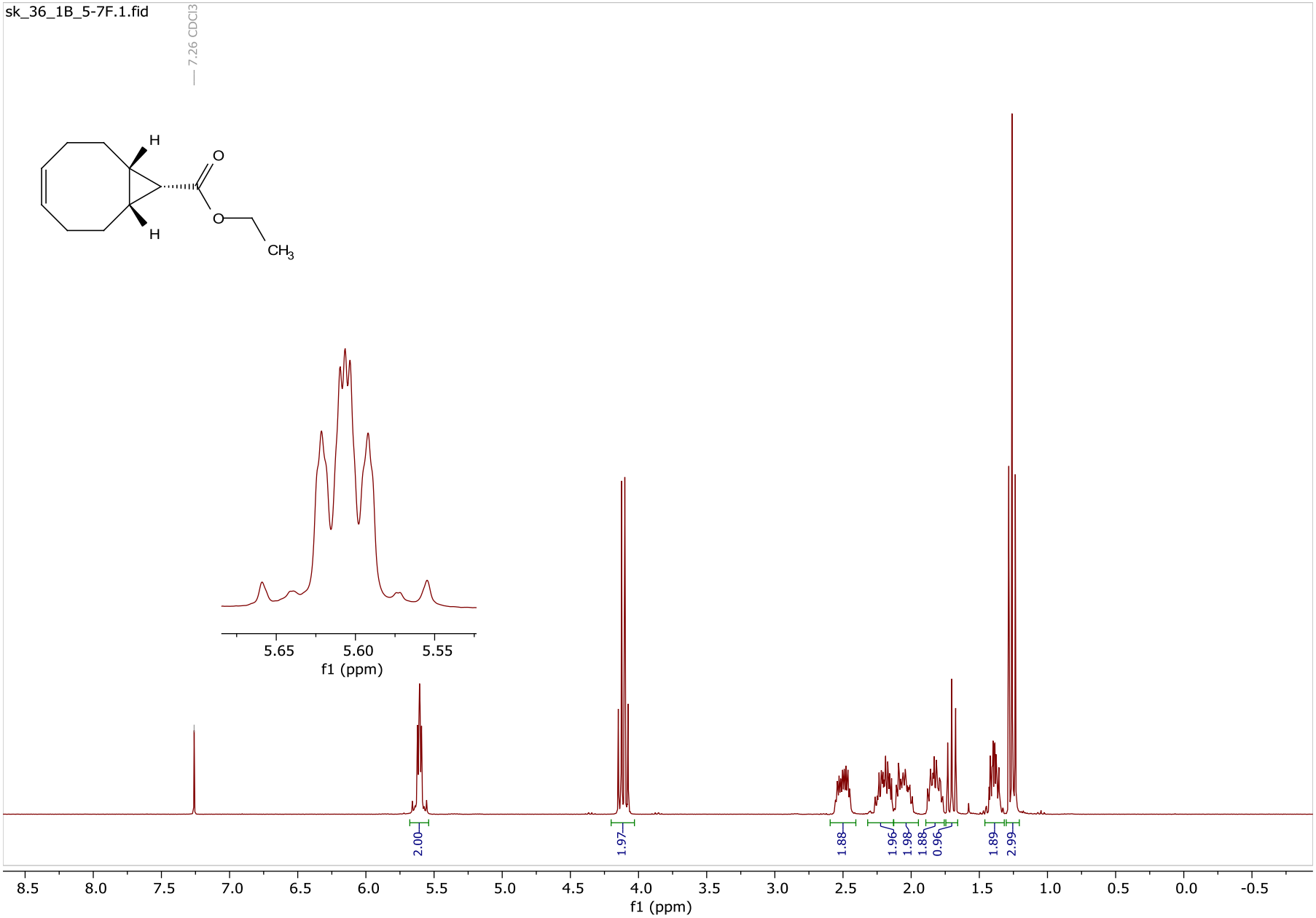
(1*R*,8*S*,9*r*,*Z*)-Ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (*exo*-9b)

[^{1}H-NMR, 300 MHz, $CDCl_3$]



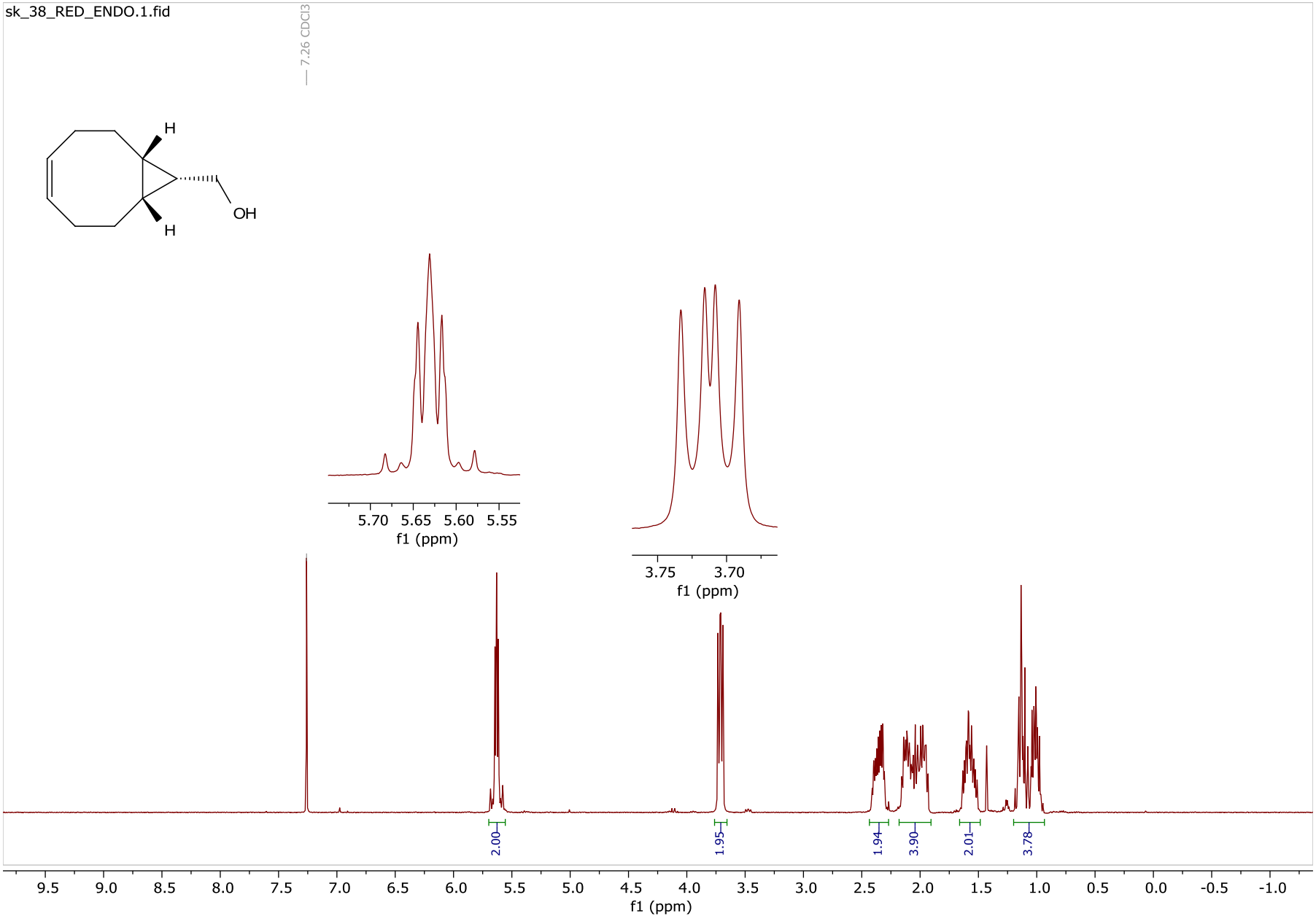
(1*R*,8*S*,9*s*,*Z*)-ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (*endo*-9a)

[^{1}H-NMR, 300 MHz, $CDCl_3$]



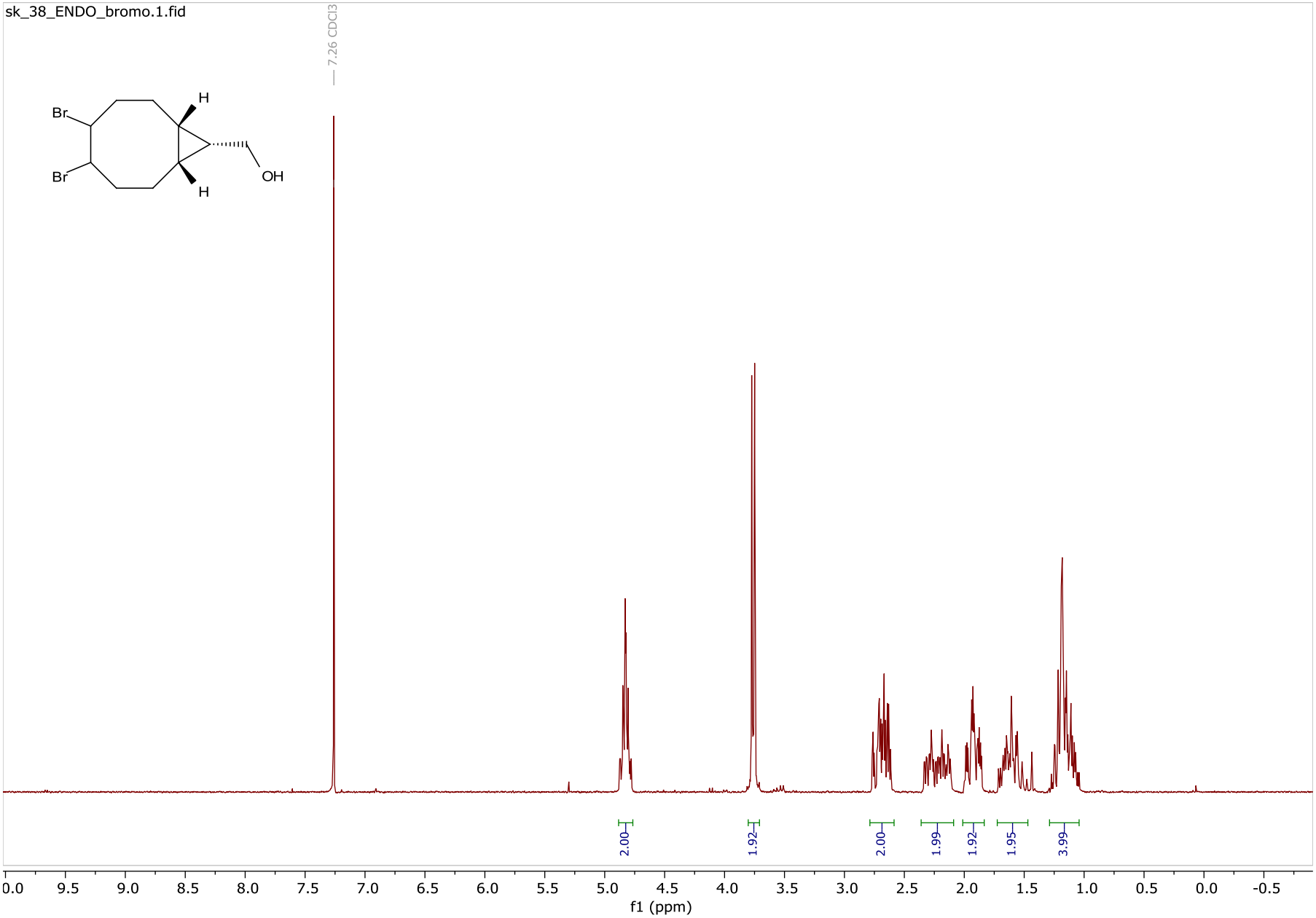
(1*R*,8*S*,9*s*,*Z*)-bicyclo[6.1.0]non-4-en-9-ylmethanol (10)

[^{1}H-NMR, 300 MHz, $CDCl_3$]



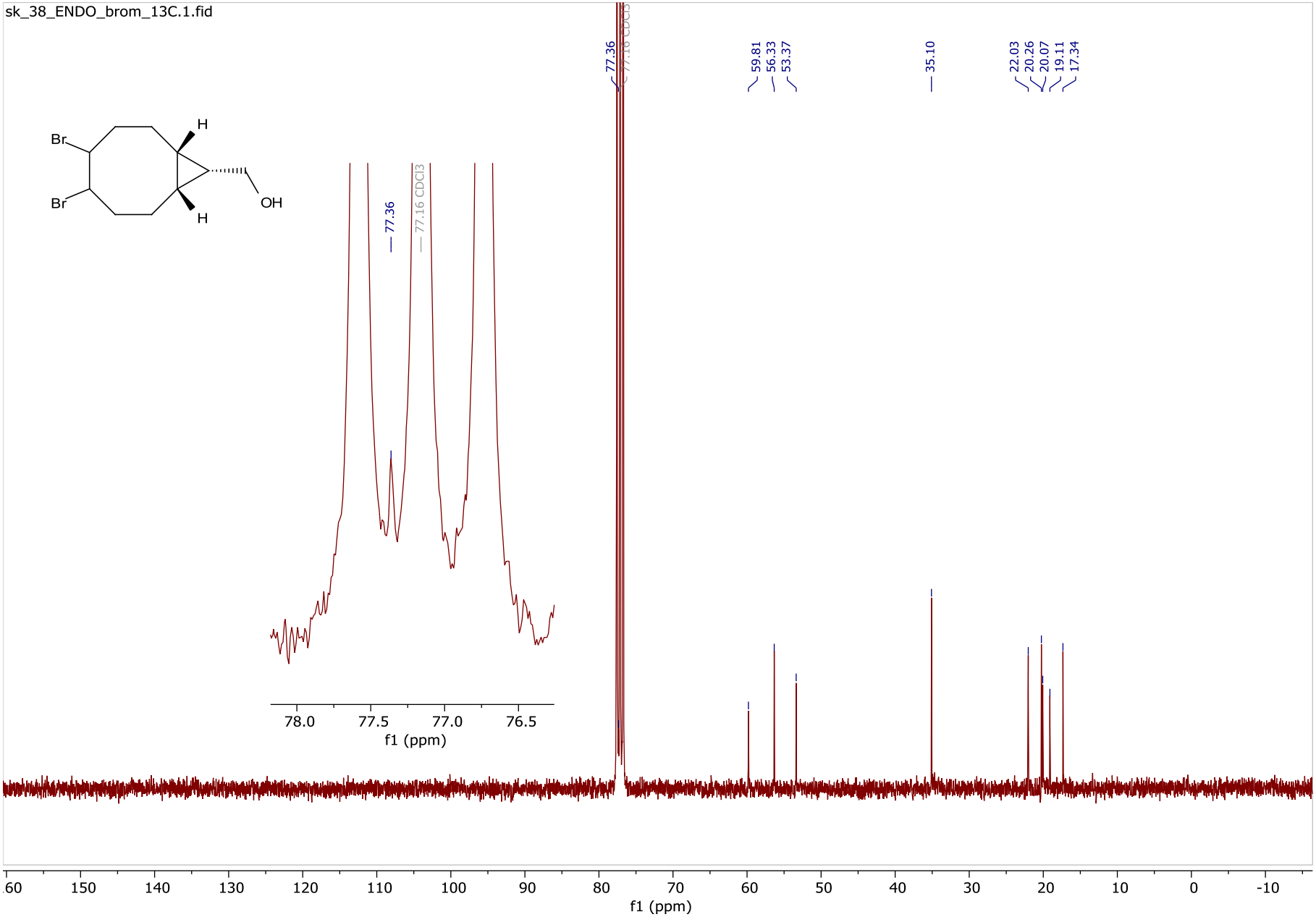
(1*R*,8*S*,9*s*)-4,5-dibromobicyclo[6.1.0]nonan-9-yl)methanol (11)

[^{1}H-NMR, 300 MHz, $CDCl_3$]



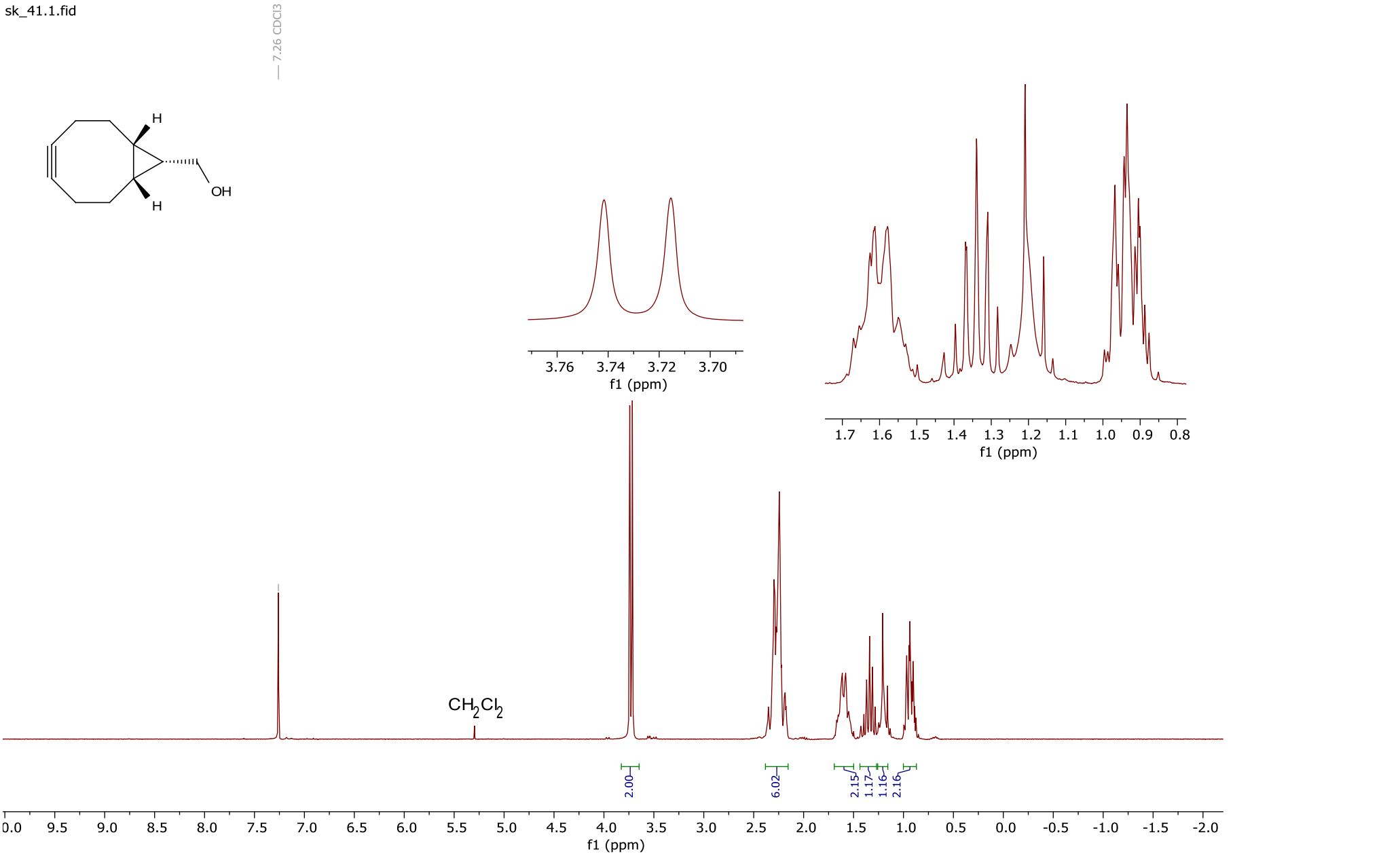
(1*R*,8*S*,9*s*)-4,5-dibromobicyclo[6.1.0]nonan-9-yl)methanol (11)

[^{13}C-NMR, 75 MHz, $CDCl_3$]



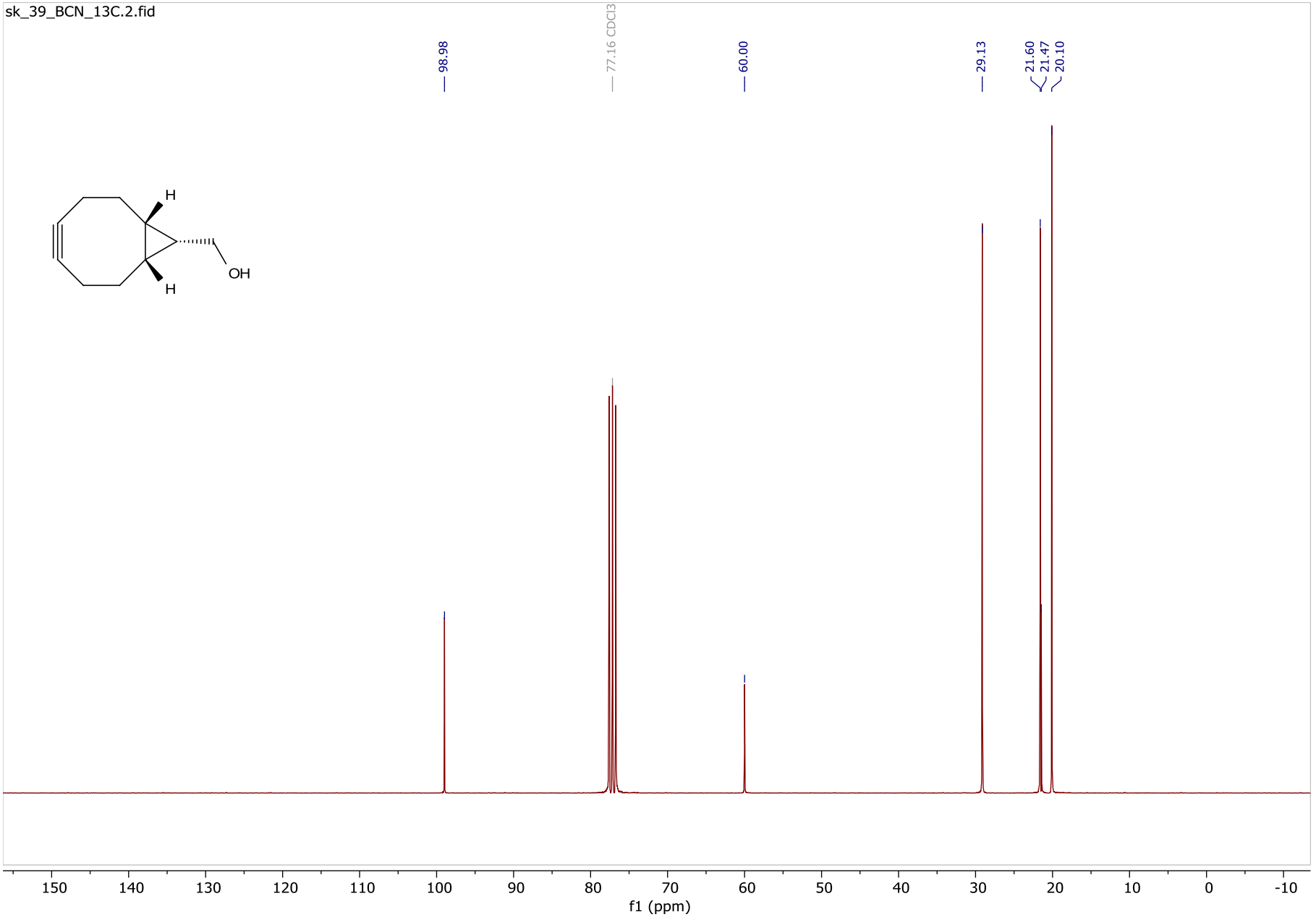
(1R,8S,9s)-bicyclo[6.1.0]non-4-yn-9-ylmethanol (12)

[^{1}H-NMR, 300 MHz, $CDCl_3$]



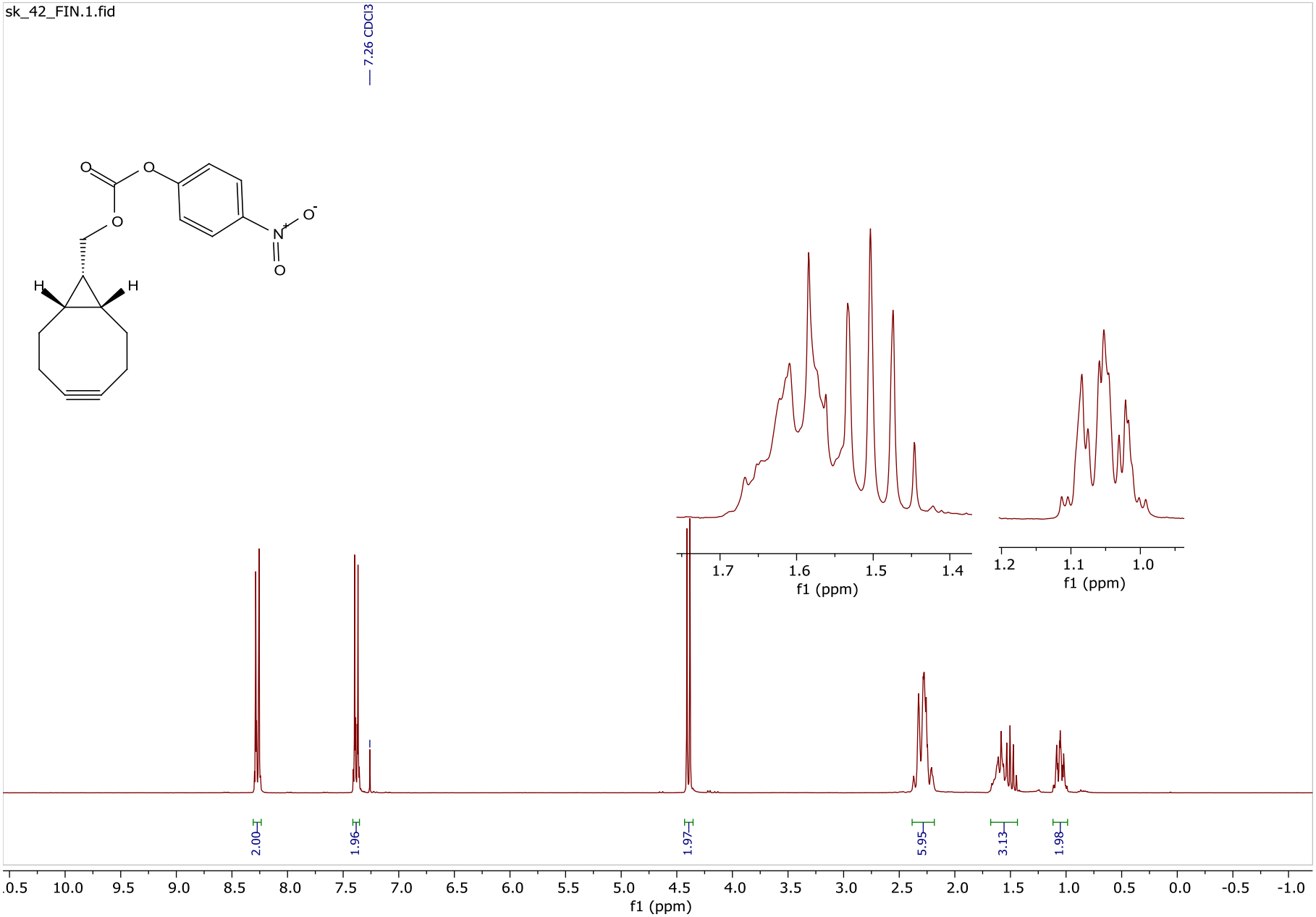
(1R,8S,9s)-bicyclo[6.1.0]non-4-yn-9-ylmethanol (12)

[^{13}C-NMR, 75 MHz, $CDCl_3$]



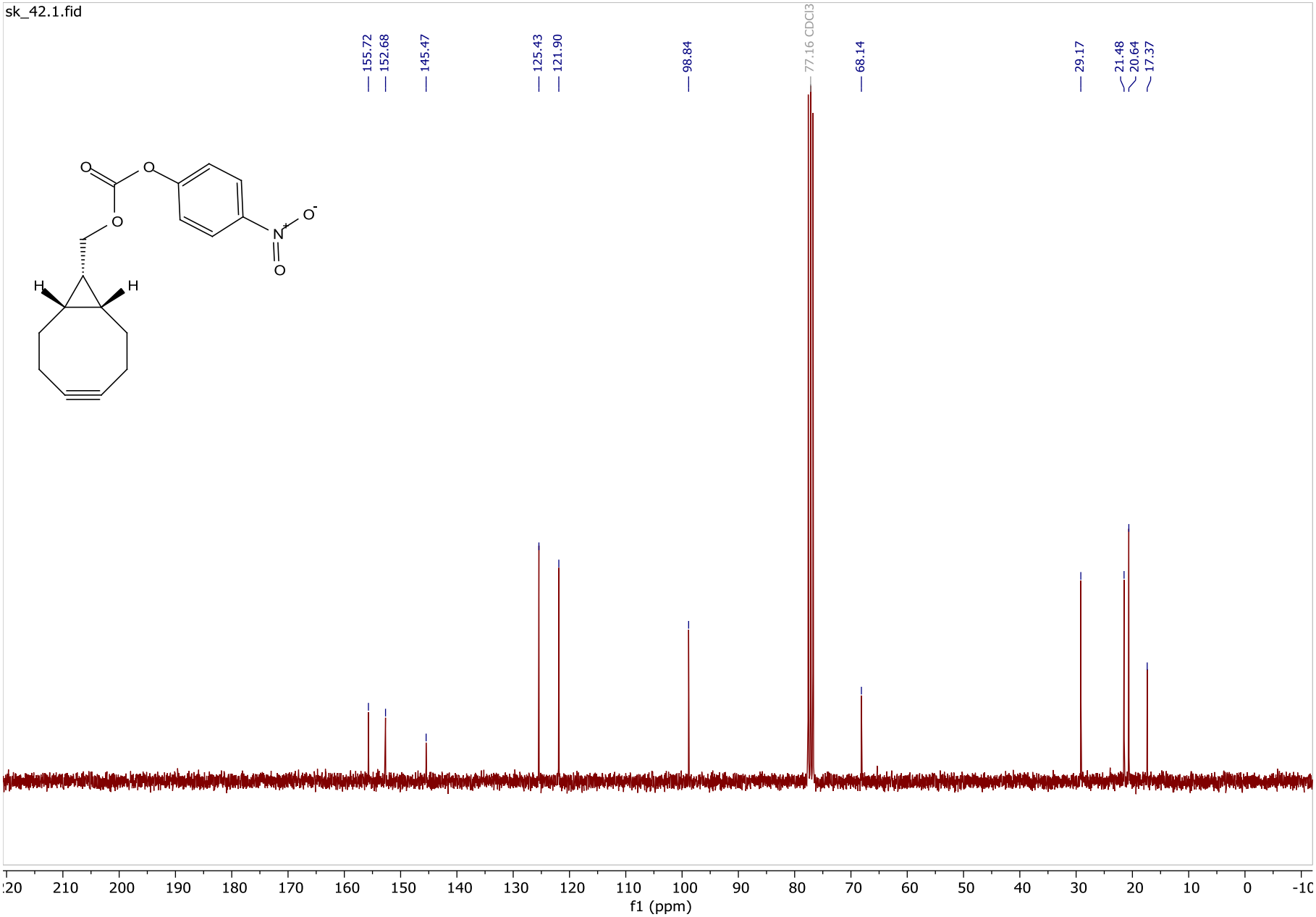
((1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-yl)methyl (4-nitrophenyl) carbonate (13)

[^{1}H-NMR, 300 MHz, $CDCl_3$]



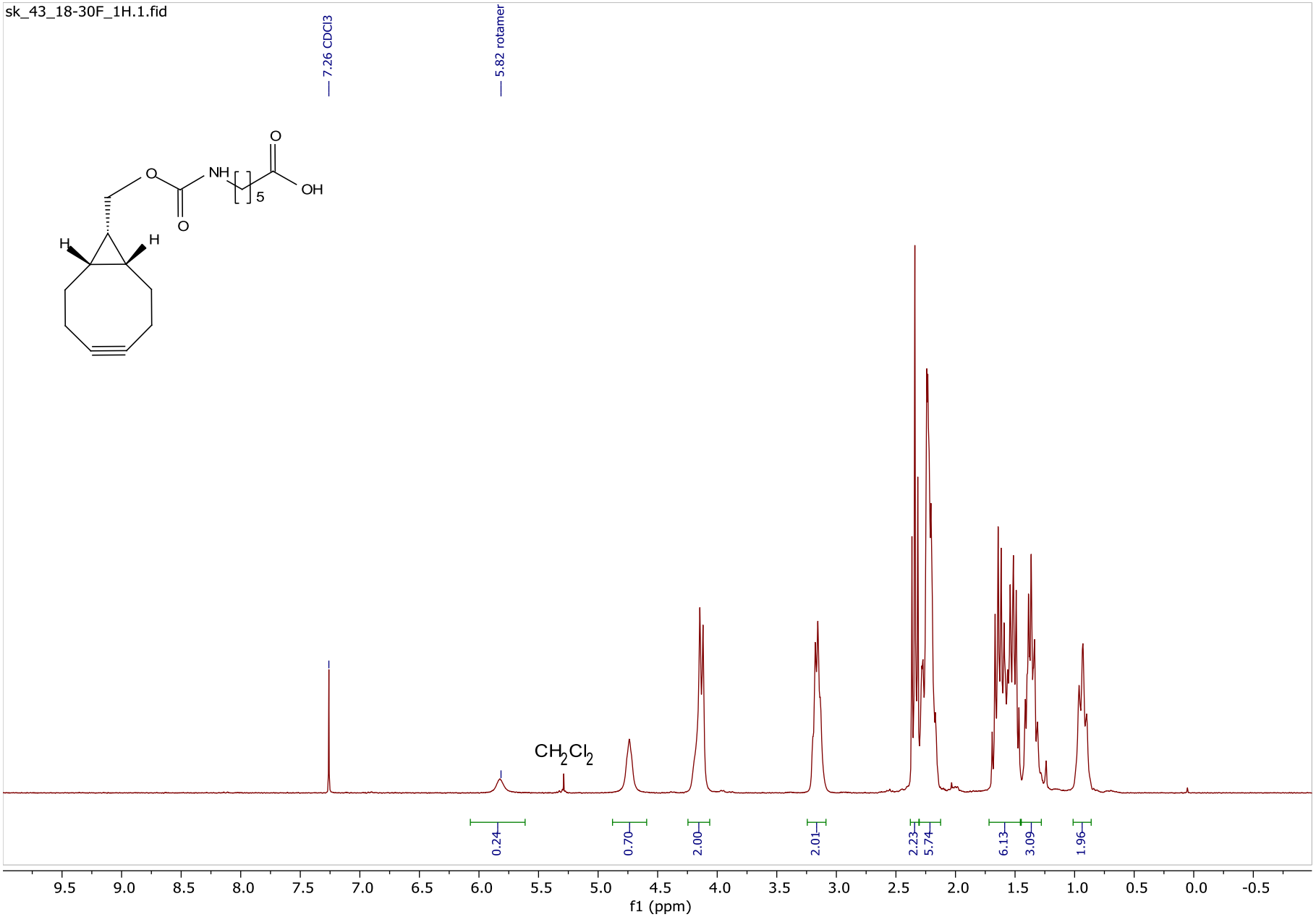
((1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-yl)methyl (4-nitrophenyl) carbonate (13)

[^{13}C-NMR, 75 MHz, $CDCl_3$]



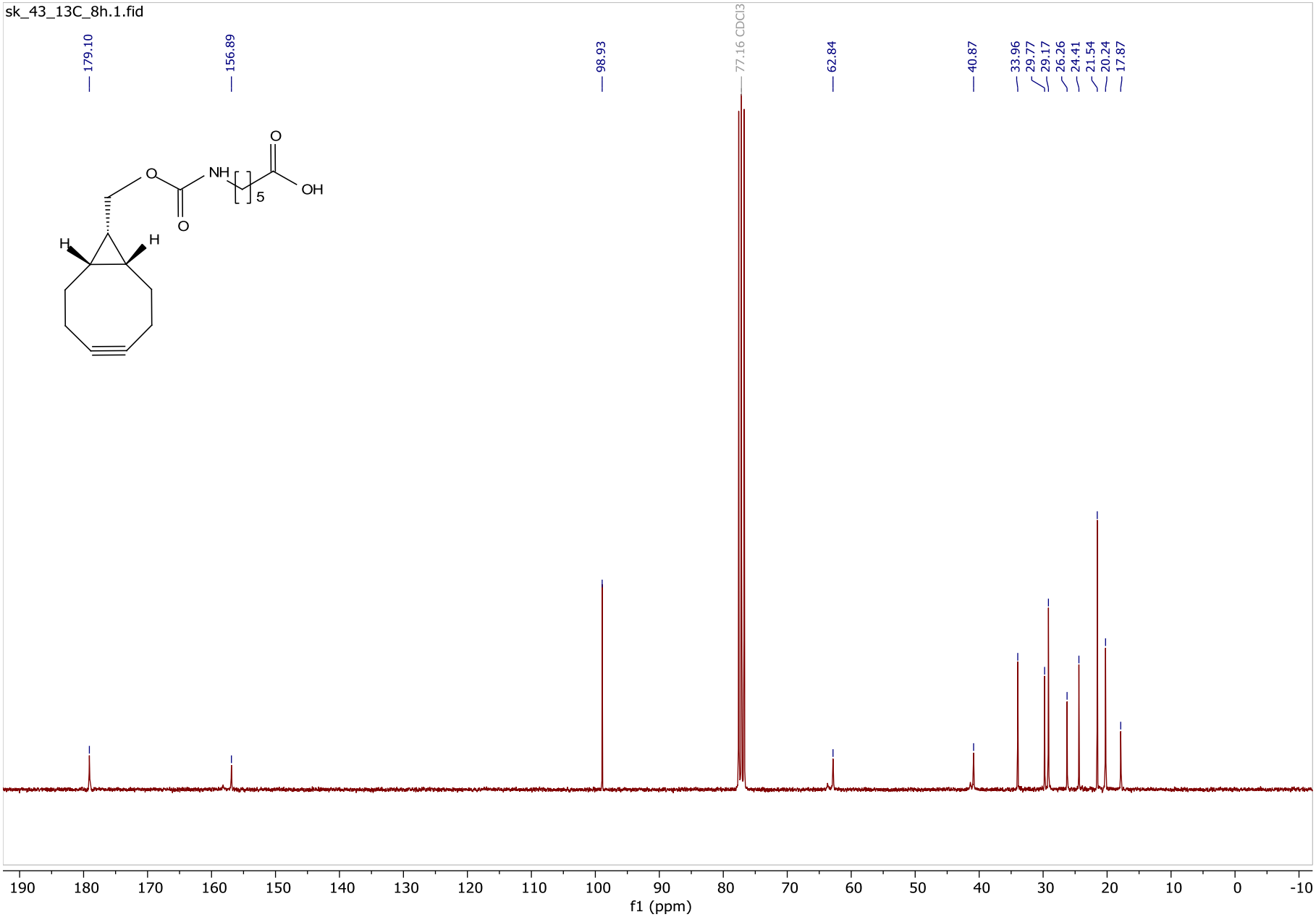
6-(((((1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbonyl)amino)hexanoic acid (14)

[^{1}H-NMR, 300 MHz, $CDCl_3$]



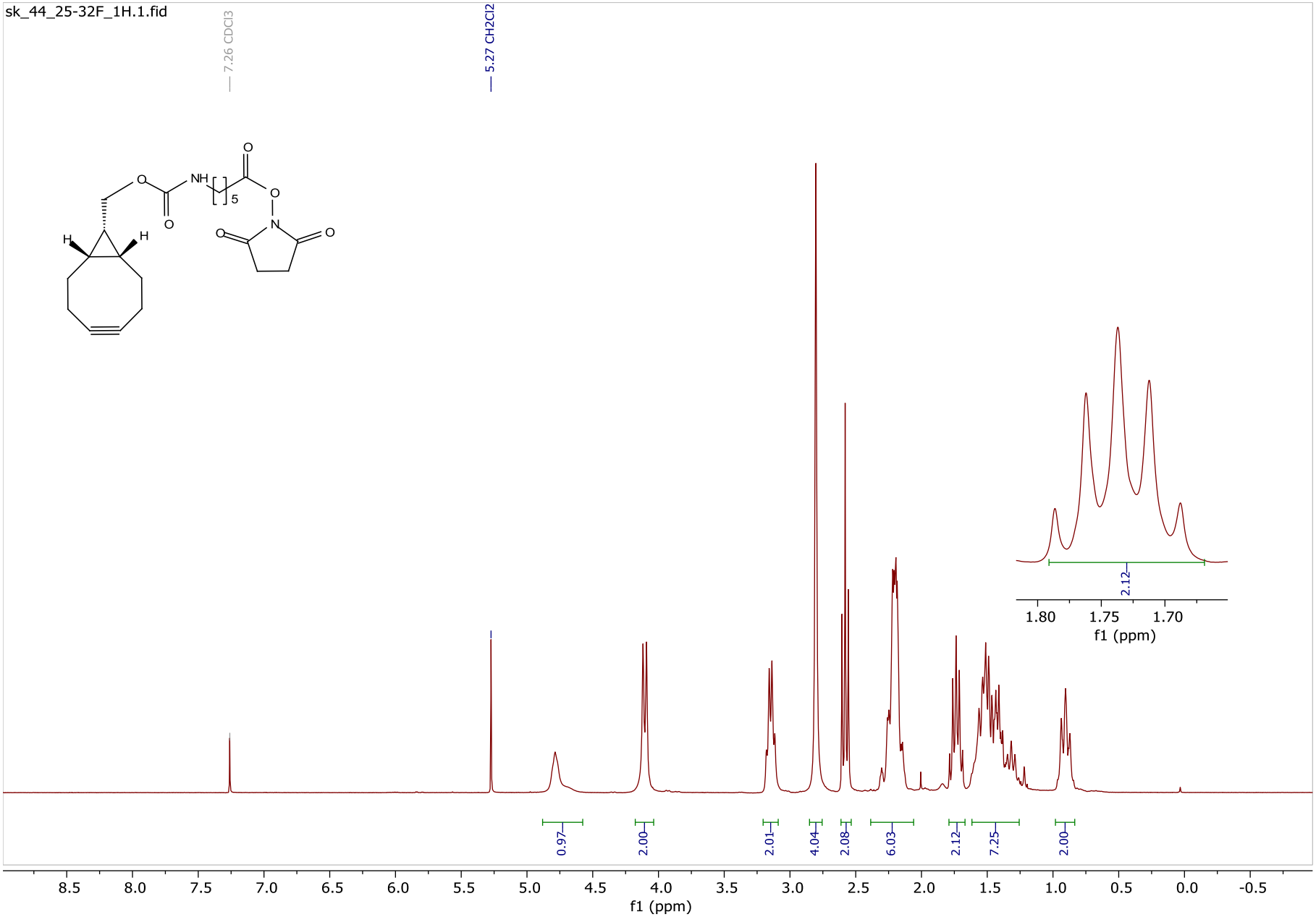
6-(((((1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbonyl)amino)hexanoic acid (14)

[^{13}C-NMR, 75 MHz, $CDCl_3$]



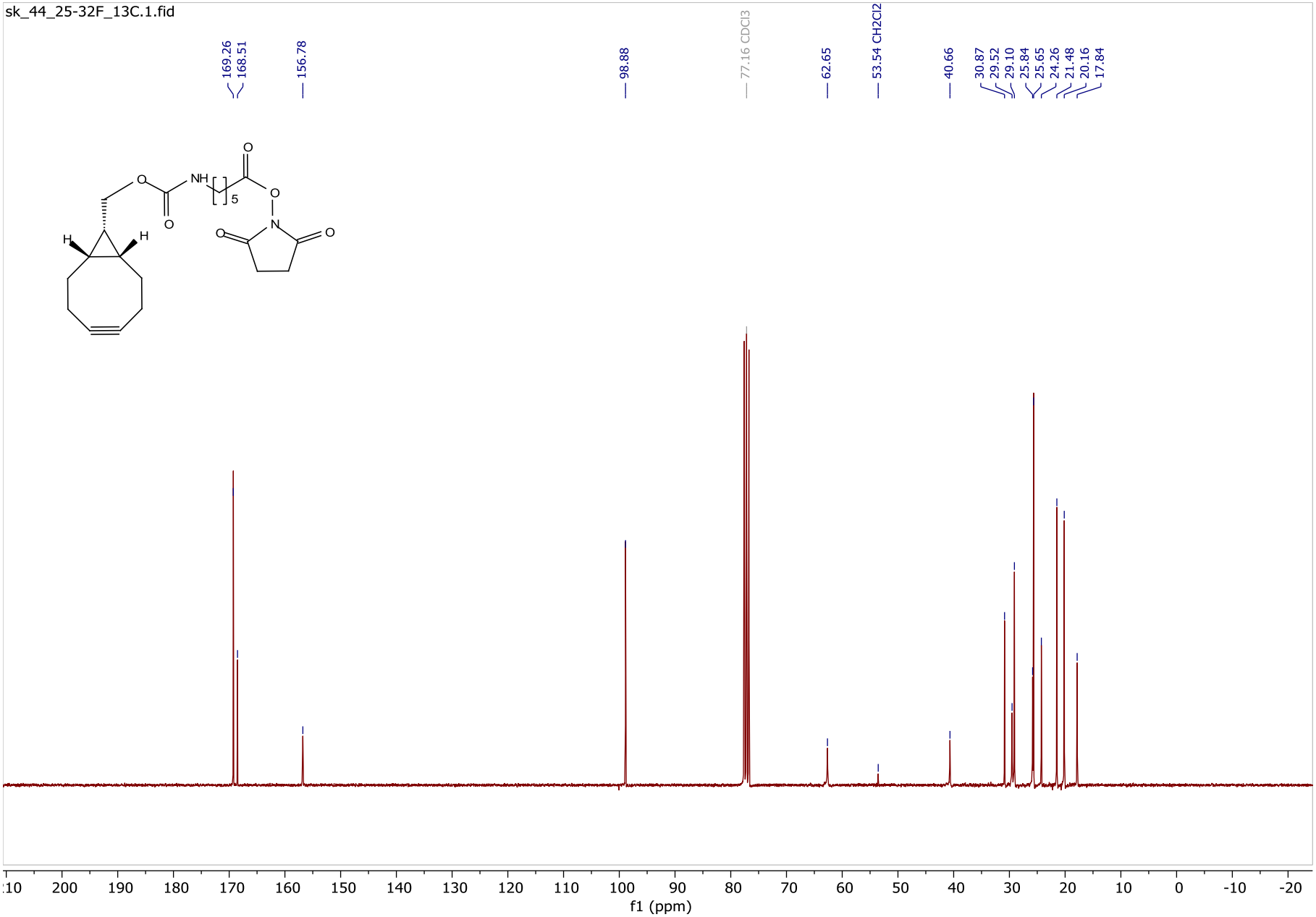
2,5-Dioxopyrrolidin-1-yl 6-(((((1*R*,8*S*,9*s*)-bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbonyl)amino)hexanoate (15)

[^{1}H-NMR, 300 MHz, $CDCl_3$]



2,5-Dioxopyrrolidin-1-yl 6-(((((1*R*,8*S*,9*s*)-bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbonyl)amino)hexanoate (15)

[^{13}C-NMR, 75 MHz, $CDCl_3$]



General Methods and Instrumentation

Cloning and bacterial protein expression methods

Codon-optimized Gene Blocks were ordered from Integrated DNA Technologies and cloned into respective expression vectors via isothermal assembly[104] unless otherwise specified. For *E. coli* expression, standard pET series vectors were used and N-terminal variants were constructed therein. Inserts were verified by Sanger sequencing.

Model protein expression was conducted in *E. coli* BL21(DE3) RIPL according to standard procedures[105], with slight modifications. To suppress gluconoylation, phosphogluconolactonase from a modified pA15 plasmid (maintained with 34 µg/mL chloramphenicol) was co-expressed to degrade (6-phospho)-glucono-1,5-lactone, which otherwise promotes protein gluconoylation in BL21(DE3).[20] To suppress host-derived acetylation, the expression medium was supplemented with 10 mM $MgSO_4$ in some instances. Plasmid maintenance was selected for with 100 µg/mL ampicillin and 50 µg/mL kanamycin, where applicable.

GSS-H_6-SpyCatcher003-mi03 was produced as described in standard BL21(DE3) cells.[40]

Qβ-VLPs were expressed and purified differently (see below).

Protein purifications

Standard E. coli derived proteins

Cells were lysed in 25 mM Tris-HCl pH 7.5-8.0, 300 mM NaCl, 0.1% v/v Triton X-100 (Sigma), 0.1%v/v Tween-20 (Sigma), 10 mM phenylmethylsulfonylfluoride (PMSF) (Sigma), EDTA-free mixed protease inhibitors (Roche), 1 mg/mL chicken lysozyme (Sigma) and 0.1 mg/mL bovine DNAse I (Sigma). Cells were frozen and thawed 3 times (-20 ºC / RT). Next $MgSO_4$ was added to 5 mM to aid DNA digestion, followed by incubation with for 15 min at 4 ºC. Lysates were centrifuged at 21,000 g at 4 °C for 30 min. The supernatant was spun once more, and was then bound to equilibrated free-flowing Nickel NTA agarose (Cube Biotech), followed by separation on a polypropylene gravity column. The resin was washed with 20-30 CV 25 mM Tris-HCl pH 7.5, 300 mM NaCl, 40 mM imidazole, and then eluted with 300-500 mM imidazole / HCl pH 7.5. Fractions containing target protein were identified by SDS-PAGE, pooled and dialysed 4 times against half-strength PBS (Sigma). Regenerated cellulose spin filters (Amicon, Sigma) were used to concentrate monomeric proteins where applicable. Monomeric proteins were stored at 20 ºC.

Note: G-H_6-Cys-ΔN1SpyCatcher (pKDB_368) was purified in the presence of beta-mercaptoethanol (Sigma), to partially reduce the installed Cys residue in order to prevent full disulfide dimerization. All other proteins were purified in non-reducing condtions.

GH_3-VLPs purification

GH_3-VLPs were prepared as above but with 500 mM NaCl in the lysis buffer. GH_3-VLPs were eluted with 1M imidazole and were buffer-exchanged on 100 kDa Amicon, or 300 kDa MWCO PES membrane spin filter (Vivaspin, Sartorius) into PBS, or PBS buffer supplemented with 400 mM NaCl to counteract precipitation for pKDB_164. Protein concentrations were determined by UV absorbance and/or Bradford reagent assay. GH_3-VLPs were stored at 4 ºC until use.

SDS-PAGE

SDS-PAGE was performed on 4-12 gradient, or 12, 16, and 18% fixed gels (Life Technologies) using an XCell SureLock system (Life Technologies) with MES running buffer. Samples were mixed with 4× LDS loading buffer to a final concentration of 1× and heated for 10 min at 70 °C. Gels were run at 200 V for 35-40 minutes. For reduced gels 50 mM DTT was included in the SDS loading buffer. Gels were stained with Colloidal Coomassie Blue stain (Kang formulation) and imaged with a Canon Canon CanoScan 9950F, or by imaging with a Xiaomi Mi A2 smartphone camera.

mP-AGE

14% polyacrylamide mP-AGE was prepared by supplementing Tris-Glycine SDS-PAGE resolving gel mixture with 3-methacrylamidophenylboronic acid (MBPA) to 0.09-0.1% w/v. MBPA was synthesized according literature.[18] To run a gel, 6x SDS-PAGE loading buffer [0.23 M Tris·HCl, pH 6.8, 120 µM bromophenol blue, 0.23 M SDS]) was added to the samples to 1x, the mixture was heated to 75ºC for 3 minutes, samples were applied to the gel and electrophoresed (180 V for 1 h 20 min) in Tris-Glycine (25 mM Tris-HCl, 192 mM glycine, and 0.1% SDS, pH 8.2) buffer followed by staining with Coomassie. Notably, a polymerizable phenylboronic acid derivative has recently become commercially available (Sigma Cat. No. 771465-1G).

MALDI-TOF MS/MS

4-Chloro-α-cyanocinnamic acid (ClCCA, Sigma Cat. 94141) matrix was prepared monthly at 4 mg/mL in 90% ACN, 0.1% TFA (v/v) and stored at 4 ºC until use. Protein and peptide samples were mixed 1:1 on a polished steel target and allowed to air dry.

2,5-Dihydroxybenzoic acid (Sigma) was used for acquisition of permethylated glycans.

MALDI TOF and MALDI TOF/TOF spectra were recorded in positive mode on a 5800 TOF/TOF (AB SCIEX) instrument. Calibration was performed against Cal Mix 2 (AB SCIEX). The extent of labelling was estimated as the ratio of the TIC peak values for labelled protein (XIC) divided by the sum of all IC value for both labelled and unlabelled protein, i.e. species that could participate in the reaction. MS/MS studies were performed using argon or air as a collision gas at 2 kV collision energy.

UHPLC-MS Q-TOF

Peptide mapping and the studies on smaller were resolved on a C18 Reversed Phase CSH column (1.0x 50 mm) (Waters) on an Acquity UPLC (Waters) coupled to a Xevo G2 Q-TOF MS (Waters). The following solvent system was used at a flow rate of 0.35 mL min^{-1}: solvent A, water containing 0.1% formic acid (v/v); solvent B, acetonitrile containing 0.1% formic acid (v/v). The column was eluted using a linear gradient from 0 to 100% of solvent B over a period of 40 minutes. Larger proteins were resolved on a BEH300 C4 Reversed Phase column (1.0x 50mm, 17 µm) (Waters) linked to the Q-TOF, eluted using a linear gradient from 0 to 90% of solvent B over 20 minutes. The extent of labelling was estimated as the ratio of the specific XIC peak values for labelled protein

divided by the sum of all IC values for both labelled and unlabelled protein that could participate in the reaction, from deconvoluted MS data.

Some samples (VLPs and mAb) were treated with either dithiothreitol (DTT) (100 mM, 37 ºC for at least 10 minutes), or TCEP (50 mM, 60 ºC, for 10 minutes), followed by quickly cooling and mixing to 10% (v/v) acetic acid, from a 50% (v/v) acetic acid solution. Some VLP samples were resolved as above but using 0.05% (v/v) TFA as the mobile phase acid modifier.

Sample digestion for glycoproteomics

Yeast and HEK-derived RBD proteins were buffer exchanged into 0.1 M Tris/3 M guanidine HCl (pH 8.5) with Amicon ULTRA-0.5 centrifugal filter units (3 kDa nMWCO) and incubated with 25 mM DTT at 37°C for 30 min. The reduced samples were carboxymethylated by addition of 0.1 M Iodoacetic acid and reacted at room temperature in the dark for 30 min. The reduced and carboxymethylated samples were buffer exchanged into 50 mM ammonium bicarbonate, pH 8.4 with Amicon® ULTRA-0.5 centrifugal filter units (3 kDa MWCO). Samples were digested with Trypsin (Trypsin-ultra™ - 1:50 w/w enzyme:protein ratio) at 37°C for 3 hours.

Glycoproteomics Analysis of digested samples

LC ES-MS (MSe) was performed on the products of digestion using an Acquity UPLC coupled with a Xevo G2- XS Q-TOF Mass Spectrometer (Waters, Milford, MA). Separations were achieved on the Acquity UPLC using a 100 mm x 2.1mm C18 reversed phase column eluted with a linear gradient of 0.1% formic acid in water (Buffer A) and 0.1% formic acid in acetonitrile (Buffer B) at a flow rate of 0.4 mL/min. The gradient used was: 0 min, 0% B; 2 min, 0% B; 62 min, 26% B; 68 min, 31% B; 74 min, 45% B; 74.5 min, 85% B; 79.5 min, 85% B. Interpretations of glycopeptide ESMS and ESMS/MS (MSe) data were performed manually by previously reported methods based on amino acid and sugar masses together with the known fragmentation mechanisms of biopolymers.[106,107] The theoretical masses of identified peptides and glycopeptides were calculated using MassLynx.

Production of RBD in Pichia pastoris yeast

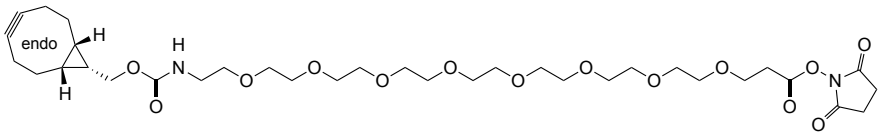
A PCR fragment encoding SARS-CoV-2 RBD sequence with an N-terminal 6x-His tag was cloned into pPICZα vector (Invitrogen) behind the α-factor secretion signal using XhoI and NotI restriction sites in a way to restore the Kex2 signal cleavage site. The RBD DNA sequence of the RBD gene and the predicted protein sequence are provided in the sequence section. After sequencing the plasmid was linearized at MssI site and transformed by electroporation in *Pichia pastoris* (reassigned as *Komagataella phaffii*) strain X-33. Mut+ transformants were obtained on YPD agar plates containing 400 µg/mL zeocin. The selected clone was cultivated 24 h in BMGY medium at 24 °C with aeration (250 rpm) followed by daily addition of 1% methanol and cultivation was continued for two more days. The cell pellet was removed by low-speed centrifugation. To supernatant, 50 mM Tris HCl and 300 mM NaCl were added and pH was adjusted to 8.0 with NaOH. 1 L of supernatant was passed through the Ni-affinity column (bed volume 20 mL) and bound material was eluted with linear 0-300 mM imidazole gradient. Peak protein fractions were concentrated until ~2 mL with Amicon filter device (MWCO 10 kDa) and subjected to size-exclusion chromatography on Superdex 200 column (bed volume 120 mL) in PBS. Target protein fractions were selected by Coomassie-stained SDS-PAGE and subsequently used for chemical coupling.

Chemical coupling of RBD to bacteriophage Qβ VLPs for immunogen preparation.

6AGDL:BCN protein chemistry

Purified RBD was concentrated to 0.5-3 mg/ml and transferred to 0.2 M HEPES (pH 7.5) solution with 1 ml Zeba Spin Desalting Columns (Thermo Fisher Scientific). Dry 6AGDL was added to the RBD solution to the final concentration of 100 mM, then the reaction mixture was incubated at room temperature for 2 hours. Unreacted 6AGDL reagent was removed and the buffer was changed to a 50 mM borate (pH 8.2) with a desalting column. Analytical gel filtration on a Superdex S200 (Akta) column revealed some partial aggregation (Figure S62).

Qβ VLPs were purified as described previously[62], concentrated to 3-5 mg/ml and transferred to a PBS buffer with desalting column. Qβ VLPs were labelled with BCN-endo-PEG8-NHS ester (SiChem, SC-8108) by adding it to the protein solution to the final concentration of 2 mM from a 10x stock solution in DMSO. The reaction mixture was then incubated at room temperature for 2 hours, and the unreacted BCN-PEG8-NHS ester was removed, and the buffer changed to a 50 mM borate (pH 8.2) with a desalting column.



Structure of BCN-endo-PEG8-NHS ester according to manufacturer.

Labelled RBD-6AGDL and Qβ-BCN-PEG8-NHS were mixed at mass ratios of 1:1, EDTA added to the final concentration of 1 mM and incubated for 1-2 days at +4° C. The excess free RBD was separated from the Qβ-RBD conjugate by gel filtration on a 16/600 Superdex 200 PG (GE Healthcare), or Superdex 200 10/300 GL (GE Healthcare) column in 50 mM borate buffer (pH 8.2). Endotoxin levels were not measured. Fractions containing conjugates were pooled and concentrated to 0.7-1.2 mg ml, flash frozen in liquid nitrogen and stored at -20 °C until immunization.

Tetrazine:BCN protein chemistry

RBD was labelled with tetrazine-NHS ester (6) by adding it to the protein solution in PBS from a stock solution in DMSO at a 0.9:1 molar ratio in respect to RBD. The mixture was incubated at room temperature for 2 hours. The unreacted tetrazine-NHS label was removed, and the buffer was changed to a 50 mM phosphate solution (pH 7.2) with a desalting column.

Qβ VLPs were labelled with BCN-NHS ester (15) by adding it to the protein solution to the final concentration of 2 mM from a 50x stock solution in DMSO and afterwards the unreacted BCN-NHS ester was removed, and the buffer was changed to a 50 mM phosphate (pH 7.2) with a desalting column.

Labelled RBD-tetrazine-NHS and Qβ-BCN-NHS reaction at mass ratios of 1:1 yielded visible precipitate, and required reaction ratio optimization. The most appropriate ratio for coupling was 1 mg of Qβ-BCN-NHS per 0.5 mg RBD. The solution was further incubated for 6 hours at +4° C, and then the excess free RBD immediately separated from the Qβ-RBD conjugate by gel filtration on a 16/600 Superdex 200 PG (GE Healthcare) or Superdex 200 10/300 GL (GE Healthcare) column in 50 mM phosphate solution (pH 7.2). Fractions containing conjugates were pooled and concentrated to 0.7-1.2 mg ml, flash frozen in liquid nitrogen and stored at -20 °C until immunization. Endotoxin levels were not measured.

Transmission electron microscopy (TEM)

Negative staining with uranyl acetate was used for the imaging of purified VLP conjugates. 5 µL of sample were placed on Formvar coated Cu 200 grid and incubated for 3 minutes. The grids were dried, briefly washed with 1 mM EDTA solution, and stained with 1% uranyl acetate for 1 minute. Grids were dried and recorded on a Jeol JEM-1230 TEM electron microscope at 100 kV.

Native agarose gel electrophoresis (NAGE)

Electrophoretic mobilities of the conjugates, purified via size-exclusion, were analysed with NAGE in standard 1% (w/v) agarose and 1x TAE buffer, followed by staining with EtBr, and Coomassie. The Coomassie image in Figure 5H was brightened for improved visibility in the main text. The unmodified image is provided as Figure S61.

Immunizations

Seven weeks old BALB/CR female mice purchased from the University of Tartu Laboratory Animal Centre (Tartu, Estonia) were immunized subcutaneously with three doses of each vaccine antigen at a dose of 25 µg of protein per mouse, two weeks apart. The proteins were diluted to 0.5 mg/mL in PBS and 50 µL of antigen were combined with 50 µL of AddaVax adjuvant (Invivogen SAS, France). The control group received sterile PBS with adjuvant. Before each immunization mice were bled via tail vein and two weeks after the last immunization mice were bled via retro-orbital sinus to obtain serum samples. Animal experiments were approved by the Latvian Veterinary and Food Department - Animal procedure permission 89.

Endpoint ELISA

For ELISA, EIA/RIA 96-Well microtitration plates (Corning Life Sciences, Tewksbury, MA, USA) were coated with 100 µL of the recombinant proteins at a concentration of 10 µg/mL overnight at +4 °C. Plates were washed with 0.05% Tween 20 in PBS (PBST) and blocked with 100 µL of 2% bovine serum albumin (BSA) in PBS for 30 min at +37 °C. Three-fold dilutions of sera were prepared starting from 1:50 and added to the coated wells (150 µL/well). The plates were incubated for 1 h at +37 °C and washed four times with PBST. Bound IgG antibodies were detected with 100 µL of horseradish peroxidase (HRP)-conjugated polyclonal rabbit anti-mouse IgG (Sigma–Aldrich, USA) diluted 1:1000 in blocking buffer and incubated for 1 h at +37 °C. Color reactions were developed for 20 min by adding the o-phenylenediamine substrate (1-Step Ultra TMB–ELISA Substrate Solution, Thermo Scientific, Rockford, IL, USA). The HRP enzymatic reaction was terminated by the addition of 2.0 M H2SO4, and the optical density was measured at 492 nm

using an ELISA plate reader (BDSL Immunoskan MS, Finland). The endpoint serum IgG antibody titers were determined as the last serum dilution with OD492 value exceeding at least triple the mean OD values of the control wells (all the components except mouse sera).

SARS-CoV2 pseudotyped murine leukemia virus neutralization assay

The CMV-Gag-Pol murine leukemia virus (MLV) packaging construct, encoding the MLV gag and pol genes (pTG5349 -Transgene) and the MLV-GFP plasmid, encoding an MLV-based transfer vector containing a CMV-GFP internal transcriptional unit were used as described.[108] For SARS CoV2 pseudotyping the plasmid pcDNA3.1-SARS2-Spike[109] encoding full length SARS-CoV2 Spike protein gene (GenBank number QHD43416.1) was used.

To generate SARS CoV2 pseudoparticles, HEK 293-T/17 cells (ATCC, CRL-11268) grown in 10 cm tissue dishes at 40% confluency were transfected with packaging (8 ug), GFP reporter (8 ug) and SARS CoV2 S protein plasmids mix by calcium phosphate precipitation method (CalPhos™ Mammalian Transfection Kit, Takara, Japan). The medium, containing 20% bovine fetal serum (7,5 ml/dish) was replaced 16 hours after transfection. Supernatants containing the pseudoparticles, were harvested 24 hours later, filtered through 0.45-um pore size membranes, aliquoted to 0.3- 0,9 ml portions, flash-frozed in liquid nitrogen and placed into -70°C for storage.

Target HEK 293-T/17 were seeded in 24 well plate in 0,5 ml medium, 4x104 cells per well. Next day the cells were transfected with human ACE2 expressing plasmid – pcDNA3.1-hACE2[109] (Addgene plasmid # 145033) by calcium phosphate precipitation method (Takara, Japan) 500 ng plasmid per well. After approximately 20 hours, the infection with pseudotyped particles and virus neutralization assay was performed. Pooled from each experimental group vaccinated mice sera were analysed at 100, 400 and 1600 dilutions. For several groups all mice sera were analysed individually at serum dilution 400.

First, the dilutions in DMEM of sera samples or SARS CoV2 S specific antibody (Sino Biological, Cat: 40592-T62) were prepared. Next, the pseudoparticle aliquotes were thawed at room temperature, 5 ul of each serum/antibody sample was mixed with 0.5 ml of pseudoparticle preparation and incubated 20 minutes at room temperature. After incubation the medium from human ACE2 transfected HEK 293-T/17 cells was replaced with 240 ul of pseudoparticle preparation per well, every sample in duplicate. The cells were incubated at 37 0C for 5 hours, then 0,5 ml of complete cell medium was added per well and cells were incubated for further 67 hours. Transduction efficiency was determined by the percentage of GFP-positive cells measured by flow cytometry (FACSAria, BD) and analysed by FlowJo_V10 software. The normalized % of neutralization was calculated as follows: N(y) = 100 – 100*y/k, N(y) – neutralization (%); y - % of GFP positive cells with serum; k - % of GFP positive cells without serum. Neutralization titers 50, 70 and 90 (NT50, NT70, NT90) were calculated by nonlinear regression using GraphPad Prism 7: dose-response inhibition, log-inhibitor vs. response (three parameter).

Passaging of VeroE6 cells.

5 ml of warmed trypsin-EDTA solution was added to one 10 cm Petri plate (>70% confluency) and incubated for 5 min at 37°C. The cells were harvested and diluted in 50 ml of DMEM + 10% heat inactivated FBS (fetal bovine serum). 1 ml of obtained suspension was added per well of 24well plate. Cells were further incubated for 12-24h until 70% confluency.

SARS-CoV-2 Neutralization Assay.

27 µL of mice serum samples were mixed with 3 µL of SARS-CoV-2 at multiple of infection (MOI) = 0.1 and incubated for 30 min. 1 mL of DMEM+FBS was added, and the obtained mixture used to infect one well of previously prepared 24 well plate with VeroE6 cells. Plate was further incubated for 1 h at 37 °C, supernatant removed and 2 mL of DMEM with 10% heat inactivated FBS added. The plate was further incubated for 48 h at 37 °C. Supernatant was centrifuged for 1 min at 10,000g, and 150 µL of superantant was transferred to a new tube, containing 150 µL of RNA Shield reagent. The mix was vortexed and kept at +4 °C until RNA extraction.

RNA extraction

RNA extraction was performed by Quick-RNA viral kit (Zymo Research) following manufacturer's instructions.

qPCR.

The qPCR mix (qScript XLT 1-Step Tough Mix, QuantaBio) was prepared following manufacturer's recommendations, using the primers: FW 5'-gaccccaaaatcagcgaaat-3'; RV 5'-tctggttactgccagttgaatctg and probe: 5'-FAM-accccgcattacgtttggtggacc-BHQ1-3'. PCR conditions were: Hold stage: 50 °C 15 min + 95°C 1 min (Ramp = 1,6 °C / sec, PCR: 95 °C 10 sec + 60 °C 1 min X 40 cycles. (Ramp = 1,6 °C / sec).

Statistic analysis of VNT results

GraphPad Prism was used for statistical analysis. An Anderson-Darling test was used to query the normality of C_q values. Several groups were not normally distributed. Hence, VNT data was analysed with Kruskal–Wallis test followed by Dunn's multiple comparison test. Plotted VNT data for Figure 5 are provided in Table S11, and the C_q values used to calculate them are provided in Table S12.

Software

MALDI TOF MS and MS/MS data was analysed with Data Explorer 4.9, build 115 (Ab Sciex).
Q-TOF MS spectra were analyzed using Masslynx 3.2 (Waters).
MaxEnt3 Electrospray plugin (Waters) was used for deconvolution (until convergence, or 15 cycles).
Curve integrations and density measurements were performed with ImageJ[110], FIJI[111] Version 2.1.0/1.53c.
GraphPad Prism 7/9 (GraphPad Software, Inc., San Diego, CA, USA) was used for statistical analyses.
Chemical structures were drawn with ChemDraw (18.2.0.37).
Scheme and figure composition was performed with Adobe Illustrator CS6 (version 16.0.0).
The manuscript was written in Microsoft Word 2011, with Zotero reference manager (Version 5.0.96.1).

Further Information

Further information and request for resources and reagents should be directed to and will be fulfilled by the corresponding authors Karl D. Brune (6AGDL, subject to MTA) and Kaspars Tars (Qβ VLP materials, yeast protein).

Figure specific methods

Method for Figure 1**.**

20 µM GSS-H_6-AffiEGFR was reacted with 100 mM 6AGDL in 0.5 M HEPES buffer (NaOH), pH 7.5 (total volume 50 µL), RT. After 1h 4 µL of concentrated acetic acid were added and 10 µl reaction mixture aliquot was desalted and prepared for MALDI analysis using 0.1% TFA and final elution into 10 µL of 70%ACN/0.1%TFA C18 ZipTip method. 0.5 µL eluate were mixed 1:1 with ClCCA matrix [4 mg/mL in 90% ACN, 0.1% TFA (v/v)] and spotted onto a MALDI steel target prior to detection in reflector positive mode MALDI-TOF MS. The extent of labelling was calculated for non-acetylated species.

LC-MS Q-TOF

20 µM GSS-H_6-AffiEGFR was reacted with 100 mM 6AGDL in the presence of 0.5 M HEPES (NaOH), pH 7.5, at RT (total reaction volume 102.5 µL). After 1h the reaction was mixed with 3 µL of concentrated acetic acid. A 10 µL aliquot of the TFA-treated reaction mixture was desalted over a C18 ZipTip by repeatedly washing with 0.1% TFA (v/v), followed by elution with 10 µL of 70% ACN (v/v), 0.1% TFA (v/v). 8 µL of the eluent was then diluted with 0.1 % (v/v) TFA to a final volume of 16 µL. 10 µL of this mixture were subjected to UHPLC-MS on an Acquity UPLC (Waters) coupled to a Xevo G2 Q-TOF MS (Waters) on a C18 Reversed Phase CSH column (1.0x 50 mm) (Waters). The following solvent system was used at a flow rate of 0.35 mL min^{-1}: solvent A, water containing 0.1% formic acid (v/v); solvent B, acetonitrile containing 0.1% formic acid (v/v). The column was eluted using a linear gradient from 0 to 100% of solvent B over a period of 40 minutes. The extent of labelling was estimated as the ratio of the specific XIC peak values for labelled protein divided by the sum of all IC value for both labelled and unlabelled protein, that could participate in the reaction from deconvoluted MS data.

Method for Figure S39 **and** Figure S40**.**

100 µM peptide [GGKWSKR-Beltide1 (PP7) or GGTYSDH-Beltide1 (PP10)] was labelled with 100 mM 6AGDL (adding pre-mixed reaction mixture components into an Eppendorf tube with dry 1.5 mg of 6AGDL powder) in 1 M HEPES (NaOH), pH 7.5 at 4 ºC in a reaction volume of 75 µL. After 3 h another 100 mM 6AGDL label was added (reaction mixture was transferred into a new Eppendorf tube with dry 1.5 mg 6AGDL) and the reaction mixture was kept for another 3 h at the same temperature. Upon reaction completion acetonitrile was added to 10% (v/v) final, to solubilize any precipitated Beltide-1. To determine the extent of labelling, 10 µL of the reaction mixture were acidified with acetic acid, purified using ZipTip C18 chromatography and analysed by MALDI-TOF.

Method for Table S3 **and** Figure S41

4-methoxyphenyl 2-azidoacetate was synthesized according to literature.[19] Stock solution was prepared by diluting light yellow, oily substance (10.14 mg) to a final volume of 49 µL with ACN to give a 1 M stock solution. This stock solution was diluted 1:40 with ACN [195 µL ACN to 5 µL of 1M stock 4MPAA] to give 25 mM reaction stock and was used to supply 4MPAA to the reactions. 6AGDL was prepared by dissolving 3.21 mg of white powder in 16 µL H_2O to give a 1 M stock 6AGDL solution. This solution was used to supply to the reaction. Peptides were reacted at 1 mM in a volume of 25 µL in 200 mM HEPES buffer at pH 7.5, 10% (v/v) ACN (final concentration) at 4 ºC or 22 ºC (RT) for the times indicated. ACN was supplied to 4MPAA samples as the solvent for the reagent and the ACN concentration in the final sample was also 10% (v/v). 2.5 µL ACN was supplied to 6AGDL samples to help dissolve peptide PP1 which was visibly precipitating [final concentration ACN 10% (v/v)]. 5 µL reaction mixture were withdrawn after 1, 18 and 24 h, mixed with 45 µL H_2O to give a 100 µM peptide solution, which was acidified with 25 µL 20% (v/v) acetic acid [final concentration of ~1.18 M acetic acid]. 10 µL of this preparation was used a substrate for standard C18 Zip Tip purification. MALDI MS and quantification were performed as before.

Method for Figure S45**.**

PP6 peptide (GSS-H6-Beltide-1), previously labelled with 6AGDL, purified by C18 reverse phase chromatography, lyophilifzed and stored at -20ºC, was retrieved and resuspended to 1 mM in 10% (v/v) ACN aqueous solution. 10 µL reactions of 100 µM peptide in 100 mM potassium phosphate buffer (pH 8) were prepared and optionally supplemented to 50 mM boric acid from a 500 mM stock solution titrated to pH 8 with NaOH. For CuAAC reaction, additives were supplied premixed to yield as final concentration 0.1 mM $CuSO_4$, 0.5 mM Tris(3-hydroxypropyltriazolyl-methyl)amine, and 5 mM aminoguanidine.[46] Next 2-fold molar excess of propargylamine (200 µM) was supplied at 2-fold molar excess from a 4 mM aqeuos stock solution. CuAAC reactions were mix-welled prior to addition of sodium ascorbate as the final reactant, mixed once more and spun-down and left to react at 37 ºC for 15-30 min. For SPAAC reaction, DBCO-acid was supplied to a 5-fold molar excess from a 10 mM aqueous stock solution, prepared from a 1:10 dilution of 100 mM DBCO-acid in DMF. Samples were analysed by acidification, followed by C18 ZipTip clean-up and MALDI TOF MS analysis.

Method for Figure S46**.**

6AGDL-labelled GSS-H6-AffiEGFR was prepared as before and excess 6AGDL-label was removed by spin filtration into 50 mM potassium phosphate buffer pH 7.0. 20 kDa PEG-alkyne (α-Methoxy-ω-alkyne, Cat. No. 1220000-70, Rapp Polymere, Germany) was dissolved in 50% aq. DMSO to prepare a 10 mM stock solution. Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA, Sigma) was prepared as an aqueous 50 mM stock, to give 2.5 mM final reaction concentration. 20 mM $CuSO_4$ stock was prepared and was used at a final reaction concentration of 0.5 mM. Presolski et al. 2011 was followed with slight modification. Fresh sodium ascorbate (100 mM stock) was prepared by quenching ascorbic acid with sodium bicarbonate. After cessation of effervescence, sodium ascorbate was quickly added to the pre-mixed THPTA and $CuSO_4$, master mix solution to give a final reaction concentration of sodium ascorbate of 3.75 mM. Azidated protein (30 µM) was mixed with 20 kDa-PEG-alkyne at 30x equivalent excess (900 µM). The activated ligand solution was then added to achieve specified reaction concentrations in a volume of 45 µL. The reaction was mixed well, capped, and incubated at 37 ºC for 1 ½ h. EDTA was added to 5 mM concentration from 0.5 M stock in order to stop the reaction. RNAseB was included as an internal, non-reactive control. The reaction was analyses by SDS-PAGE, and staining with Coomassie.

Method for Figure S48**.**
GSS-H_6-AffiEGFR was labelled at 75 µM with 100 mM 6AGDL in 500 mM HEPES (NaOH), pH 7.5 (NaOH) at room temperature (23.8 ºC) for 1h by adding buffered protein solution (72 µL) to dry 6AGDL powder (1.57 mg). The reaction was stopped by addition of acetic acid to 0.8 M. The reaction was split over three 3 kDa MWCO spin filter devices, which were washed 5 times with a 9-fold excess of the respective storage buffer to remove any remaining free label and other buffering components. The final volume of each retentate was adjusted to ~90 µL to give ~20 µM solutions with different pH values. Buffer solutions employed were: pH 4.5 [100 mM acetic acid (NH_3)], pH 7.5 [50 mM HEPES (NaOH)], and pH 8.8 [100 mM NH_3 (acetic acid)]. The reactions were stored at room temperature and 10 µL aliquots were taken at each time and desalted using ZipTip C18 resin with elution in 10 µL of 70% (v/v) ACN, 0.1 % (v/v) TFA. Eluate was mixed 1:1 with ClCCA matrix mixture and spotted onto a MALDI steel target prior to detection in linear positive mode on the MALDI-TOF MS and labelling extent was determined by calculating the ratio of labelled (m_1/z) and unlabelled (m_2/z) protein at the secondary charge state with the following formula ($[m_1/z/(m_1/z+m_2/z)]*100\%$, @*z* +2).

Method for Figure S49**.**
575 µL of 55 µM GSS-H_6-AffiEGFR was labelled with 300 mM GDL in the presence of 650 mM HEPES (NaOH), pH 7.5 at room temperature (22 ºC) for 1 h. Excess GDL label was removed via repeated spin filtering and washing into ~575 µL of 50 mM potassium phosphate buffer (KOH), pH 7.5 through a 3 kDa MWCO Amicon cellulose filter unit. A 10 µL aliquot was taken and acidified with 2 µL acetic acid to assess the degree of labelling via ZipTip C18 desalting and MALDI-TOF MS analysis. The spin-filtered reaction volume was evenly split across 6 standard polypropylene tubes (50 µL per tube) to give a concentration of ~105 µM. One sample was lyophilised and then stored at room temperature (time at -80 ºC for freezing ~30 minutes); the other samples were stored in solution either at -80 ºC (frozen), -11 ºC (liquid), RT or 37 ºC. After 1 week the lyophilised sample was reconstituted with the initial amount of ultrapure H_2O (50 µL) and all samples were equilibrated to room temperature prior to cleaning up 10 µL of analyte (acidified with 2 µL of concentrated acetic acid) via the C18 ZipTip procedure, followed by MALDI-TOF MS analysis in positive linear mode. Three independent experiments were performed.

Method for Figure S55
20 µL of aqueous 200 µM GDL-conjugated peptide solution (buffered with 20 mM AMBIC, pH 8.44, 10% v/v ACN) was mixed with 2.5 µL of 20 mM 2-formylphenylboronic acid solution. Then 2.5 µL of 20 mM N-t-butylhydroxylamine hydrochloride was added to the reaction mixture. After 2h of reaction at room temperature, 10 µL of reaction mixture was taken, mixed with 2 µL of 50% (v/v) acetic acid, cleaned using C18 ZipTip procedure and analysed by MALDI-MS.

Peptide sequences

Peptide sequences.

ID	Description	Sequence
PP1	GHHHHHH-Beltide-1	GHHHHHHDWLKAFYDKVAEKLKEAF
PP2	GHHH-Beltide-1	GHHHDWLKAFYDKVAEKLKEAF
PP3	GHHHH-Beltide-1	GHHHHDWLKAFYDKVAEKLKEAF
PP4	GGHHHHHH-Beltide-1	GGHHHHHHDWLKAFYDKVAEKLKEAF
PP5	HHHHHH-Beltide-1	HHHHHHDWLKAFYDKVAEKLKEAF
PP6	GSSHHHHHH-Beltide-1	GSSHHHHHHDWLKAFYDKVAEKLKEAF
PP7	Nef(7)-Beltide-1	GGKWSKRDWLKAFYDKVAEKLKEAF
PP8	Tc(7)-Beltide-1	GASGSKGDWLKAFYDKVAEKLKEAF
PP9	GG-Beltide-1	GGDWLKAFYDKVAEKLKEAF
PP10	EMP1varC6D(7)-Beltide-1	GGTYSDHDWLKAFYDKVAEKLKEAF
PP11	Beltide-1-DELYKHHHHHH	DWLKAFYDKVAEKLKEAFDELYKHHHHHH
PP12	Beltide-1	DWLKAFYDKVAEKLKEAF
EMP1	EMP1	GGTYSCHFGPLTWVCKPQGG
GRGDSPC	GRGDSPC	GRGDSPC
PP13	GSSH6-EMP1varC6A,varC15A	GSSHHHHHHGGTYSAHFGPLTWVAKPQGG
PP14	EMP1varC5A,varC15A	GGTYSAHFGPLTWVAKPQGG

Protein and DNA sequences

Gene insert are reported for new expression vectors. Expected amino acid sequences are reported from DNA sequences with uncleaved initiator formylmethionine. The name of each protein is given as >protein name, followed by expression plasmid identifier in parentheses.

>GSS-H_6-SUMO(pKDB_013)
ATGGGTAGCAGCCATCACCATCATCATCACTCGGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAGTCAAG
CCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTTTAAGAAGGCT
GATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCTTGTACGACGGTATTAGAATTCAAGCTGATCAG
ACCCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCACAGAGAACAGATTGGTGGTTAA

M**GSSHHHHHH**SDSEVNQEA**K**PEV**K**PEV**K**PETHINL**K**VSDGSSEIFF**K**I**KK**TTPLRRLMEAFA**K**RQG**K**EMDSLRFLYDGIRIQADQTPE
DLDMEDNDIIEAHREQIGG*

---

pET28-SUMO-SpyTag[112] was a gift from Mark Howarth, and the resulting protein is here referred to as GSS-H_6-SUMO-SpyTag.
>GSS-H_6-SUMO(pET28-SUMO-SpyTag)

M**GSSHHHHHH**GSDSEVNQEA**K**PEV**K**PEV**K**PETHINL**K**VSDGSSEIFF**K**I**KK**TTPLRRLMEAFA**K**RQG**K**EMDSLRFLYDGIRIQADQTP
EDLDMEDNDIIEAHREQIGGAHIVMVDAY**K**PT**K**GY*
Cyan: SpyTag

---

pET28a-SnoopTag-AffiEGFR-SpyTag[27] (GenBank accession no. KU296973), was a gift from Mark Howarth, and the resulting protein is here referred to as GSS-H_6-AffiEGFR.
>GSS-H_6-AffiEGFR(pET28a-SnoopTag-AffiEGFR-SpyTag)
M**GSSHHHHHH**SSGLVPRGSHMG**K**LGDIEFI**K**VN**K**GSGESGSGASMTGGQQMGRDPGVDN**K**FN**K**EMWAAWEEIRNLPNLNGWQM
TAFIASLVDDPSQSANLLAEA**KK**LNDAQAP**K**GLEGSGEGSGAHIVMVDAY**K**PT**K***
Green: SnoopTag. Cyan: SpyTag

---

pET28a-SpyTagMBP[43] (Addgene plasmid # 35050) was a gift from Mark Howarth, and the resulting protein is here referred to as GSS-H_6-MBP.
>GSS-H_6-MBP(pET28a-SpyTagMBP)
M**GSSHHHHHH**SSGLVPRGSHMGAHIVMVDAY**K**PT**K**GSGESG**K**IEEG**K**LVIWINGD**K**GYNGLAEVG**KK**FE**K**DTGI**K**VTVEHPD**K**LEE
KFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPD**K**AFQD**K**LYPFTWDAVRYNG**K**LIAYPIAVEALSLIYN**K**DLLPNPP**K**TWEEIPAL
D**K**EL**K**A**K**G**K**SALMFNLQEPYFTWPLIAADGGYAF**K**YENG**K**YDI**K**DVGVDNAGA**K**AGLTFLVDLI**K**N**K**HMNADTDYSIAEAAFN**K**GETA
MTINGPWAWSNIDTS**K**VNYGVTVLPTF**K**GQPS**K**PFVGVLSAGINAASPN**K**ELA**K**EFLENYLLTDEGLEAVN**K**D**K**PLGAVAL**K**SYEEEL
A**K**DPRIAATMENAQ**K**GEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEAL**K**DAQTNSSS*
Cyan: SpyTag

The DNA template for FliC was a gift from Sung-il Yoon. A construct with N-terminal G-hexahistidine tag fused via TEV-protease cleavage site to a-type flagellin from *Pseudomonas aeruginosa* with internal SpyTag was constructed.
>GSS-H_6-aFliC(pKB_259)
ATGGGCCACCATCACCATCACCATGAGAACCTGTATTTTCAGGGCGGAAGCCAGGTCAACGGCCTGAACGTGGCTACCAAGAAC
GCCAACGATGGTATCTCCCTGGCGCAGACCGCTGAAGGCGCCCTGCAGCAGTCGACCAACATCCTGCAGCGTATGCGTGACCT
GTCCCTGCAGTCGGCCAACGGCTCCAACAGCGACTCCGAGCGTACCGCTCTGAATGGCGAAGTGAAGCAACTGCAGAAAGAAC
TGGATCGTATCAGCAACACCACCACCTTCGGTGGCCGCAAGCTGCTCGACGGTTCCTTCGGCGTCGCCAGCTTCCAGGTGGGT
TCGGCCGCCAACGAAATCATCAGCGTCGGCATCGACGAGATGAGCGCAGAGTCGCTGAACGGCACCTACTTCAAGGCTGACGG
CGGCGGCGCGGTCACTGCTGCAACCGCTTCGGGCACCGTCGACATCGCGATCGACATCACCGGCGGCAGCGCCGTGAACGTC
AAGGTCGACATGAAGGGCAACGAAACCGCCGAGCAGGCGGCTGCCAAGATCGCCGCTGCGGTCAACGACGCCAACGTCGGCA
TCGGTGCCTTCACCGACGGCGCGCAGATCAGCTACGTGTCCAAGGCTAGTGCGGATGGTACTACCAGTGCGGTTTCCGGCGTA
GCCATCACCGACACCGGCAGCGGTGGTTCAGGTGCGCATATTGTAATGGTCGATGCTTATAAACCTACCAAAGGTTCAGGTGGA
GCAGGTACCGCGGCAGGCACCACGACGTTCACCGAGGCTAACGACACCGTCGCCAAGATCGACATCTCCACCGCGAAGGGCG
CTCAGTCCGCCGTGCTGGTGATCGACGAGGCGATCAAGCAGATCGACGCCCAGCGTGCCGACCTCGGTGCGGTGCAGAACCG
CTTCGACAACACCATAAACAACCTGAAGAACATCGGTGAGAACGTATCGGCTGCTTAA

M**GHHHHHH**ENLYFQGGSQVNGLNVAT**K**NANDGISLAQTAEGALQQSTNILQRMRDLSLQSANGSNSDSERTALNGEV**K**QLQ**K**ELDR
ISNTTTFGGR**K**LLDGSFGVASFQVGSAANEIISVGIDEMSAESLNGTYF**K**ADGGGAVTAATASGTVDIAIDITGGSAVNV**K**VDM**K**GNET
AEQAAA**K**IAAAVNDANVGIGAFTDGAQISYVS**K**ASADGTTSAVSGVAITDTGSGGSGAHIVMVDAY**K**PT**K**GSGGAGTAAGTTTFTEAN
DTVA**K**IDISTA**K**GAQSAVLVIDEAI**K**QIDAQRADLGAVQNRFDNTINNL**K**NIGENVSAA*
Cyan: SpyTag

---

pET28a-SnoopCatcher[27] (Addgene plasmid # 72322) was a gift from Mark Howarth, and the resulting protein is here referred to as GSS-H_6-SnoopCatcher.

> GSS-H_6-SnoopCatcher(pET28a-SnoopCatcher)

M**GSSHHHHHH**SSGLVPRGSHM**K**PLRGAVFSLQ**K**QHPDYPDIYGAIDQNGTYQNVRTGEDG**K**LTF**K**NLSDG**K**YRLFENSEPAGY**K**PV QN**K**PIVAFQIVNGEVRDVTSIVPQDIPATYEFTNG**K**HYITNEPIPP**K***

G-H_6-Cys-ΔN1SpyCatcher expression plasmid pKDB_368 was build after ΔN1SpyCatcher[113] from DNA template pDEST14-SpyCatcher[43], a gift from Mark Howarth (Addgene plasmid # 35044). The protein has an N-terminus proximal Cys residue for covalent dimerization.

>G-H_6-Cys-ΔN1SpyCatcher(pKDB_368)

ATGGGCCACCATCACCATCACCATGATTGCGACATCCCAACGACCGAAAACCTGTATTTTCAGGGCTCCGGTGATAGTGCTACC
CATATTAAATTCTCAAAACGTGATGAGGACGGCAAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTA
TTAGTACATGGATTTCAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGAC
GGTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACG
CTCATATTTAA

M**GHHHHHH**D**C**DIPTTENLYFQGSGDSATHI**K**FS**K**RDEDG**K**ELAGATMELRDSSG**K**TISTWISDGQV**K**DFYLYPG**K**YTFVETAAPDGY
EVATAITFTVNEQGQVTVNG**K**AT**K**GDAHI*

>GH_3-MS2(pKDB_156)

ATGGGTCATCATCATCCAGATCATCGTGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGTGATGTAACAG
TAGCTCCAAGCAACTTCGCTAACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCG
TTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTGGTGTAG
AGCTTCCTGTAGCCGCATGGCGTTCGTACTTAAATATGGAATTAACCATTCCAATTTTCGCTACGAATTCCGACTGCGAGCTTATT
GTTAAGGCAATGCAAGGTCTCTTAAAAGACGGAAACCCGATTCCATCAGCAATCGCAGCAAACTCCGGCATCTACTAA

M**GHHH***PDHR*ASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGVEL
PVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY*

Note: The PDHR-linker is derived from an assembling HiPerX MS2-variant.[35]

>GH_3- PQ-465 (pKDB_164)

ATGGGTCATCATCATGCAAGTGGCGGTAGCGGCGGTAGCGGCGGTAGCGCCGCACAGCATAATATGCGTCTTCAACTCACGTC
TGGCACAAGCCTAACATGGGTTGACCCTAACGACTTTCGTTCAACTTTCCGCATTAACTTGAATGTTAATCAGAAAGTTGCCGGC
GCCGTGTCTGTTTATAACGCTCGTAGTGAGGTTATTACAAACCGCGCACCTTTAGTCGTTATCGAAGGTTGCACTGATGCCTGTA
GTGTTAATCGTGAGAATATCTCGATCCGTACCACGATTAGTGGATCTGTCGAAAATAAGGCTGCCGTTCTGGCTGCCCTTCTCGA
CCATCTTCACAATCTTGGTTTGGCTCGCGATGATTTGGTCGCTGGTCTACTACCAACCACCATTCAACCTGTTGTAGAATACACA
GGTTCTTAA

M**GHHH***ASGGSGGSGGSA*AQHNMRLQLTSGTSLTWVDPNDFRSTFRINLNVNQKVAGAVSVYNARSEVITNRAPLVVIEGCTDACSV
NRENISIRTTISGSVENKAAVLAALLDHLHNLGLARDDLVAGLLPTTIQPVVEYTGS*

Linker in italics.

>GH_3-Shahe-1(pKDB_178)

ATGGGTCATCATCATGCAAGTGGCGGTAGCGGCGGTAGCGCCACTACTACTCGTTTATTCTCAAGTGACCAAGGCGGTCATAAA
TATACCGTCCCTGGCGCTACCGACGGCTACATGGTAACTTTCAAAACATCAAAATCCACTAAGGATGCAGGTGGTTTCAAAGTTC
CGAATCACAAGCTCGAGGTCTTTGTTAATACTCAGGCCCCGGTCTCCGTGACTTCATCAACATGTACACCGAAGTGCGGTCAAG
AAAATCTGAGCGTCCGGATTATAGTATCGGGCTCGCCGTTGAATAAGGACGAACTAGATCTTCTAGTTACCGGCTTATGCGGCCA
GATGAAGACTTGGTTTGAGAGCGAATACGTCTTTGACGGATACGCTCCCACTACGTTACCATTGTACTATACGGGCACCGTCGA
GTAA

M**GHHH***ASGGSGGSA*TTTRLFSSDQGGHKYTVPGATDGYMVTFKTSKSTKDAGGFKVPNHKLEVFVNTQAPVSVTSSTCTPKCGQE
NLSVRIIVSGSPLNKDELDLLVTGLCGQMKTWFESEYVFDGYAPTTLPLYYTGTVE*

Linker in italics.

>GH_3-AP205d(pKDB_124)
ATGGGTCATCATCATGCAAGTGAAAGTGGTGCAGCGAACAAACCGATGCAGCCGATCACCTCTACCGCGAACAAAATCGTTTGG
TCTGACCCGACCCGTCTGTCTACCACCTTCTCTGCGTCTCTGCTGCGTCAGCGTGTTAAAGTTGGTATCGCGGAACTGAACAAC
GTTTCTGGTCAGTACGTTTCTGTTTACAAACGTCCGGCGCCGAAACCGGAAGGTTGCGCGGACGCGTGCGTTATCATGCCGAAC
GAAAACCAGTCTATCCGTACCGTTATCTCTGGTTCTGCGGAAAACCTGGCGACCCTGAAAGCGGAATGGGAAACCCACAAACGT
AACGTTGACACCCTGTTCGCGTCTGGTAACGCGGGTCTGGGTTTCCTGGACCCGACCGCGGCGATCGTTTCTTCTGACACCACC
GCGGGTTCTGCAAATAAGCCAATGCAACCGATCACATCTACAGCAAATAAAATTGTGTGGAGTGATCCAACTCGTTTATCAACTA
CATTTTCAGCAAGTCTGTTACGCCAACGTGTTAAAGTTGGTATTGCCGAACTGAATAATGTTTCAGGTCAATATGTATCTGTTTATA
AGCGTCCTGCACCTAAACCGGAAGGTTGTGCAGATGCCTGTGTCATTATGCCGAATGAAAACCAATCCATTCGCACAGTGATTTC
AGGGTCAGCCGAAAACTTGGCTACCTTAAAAGCAGAATGGGAAACTCACAAACGTAACGTTGACACACTCTTCGCGAGCGGCAA
CGCCGGTTTGGGTTTCCTTGACCCTACTGCGGCTATCGTATCGTCTGATACTACTGCTTAA

M**GHHH***ASESGA*ANKPMQPITSTANKIVWSDPTRLSTTFSASLLRQRVKVGIAELNNVSGQYVSVYKRPAPKPEGCADACVIMPNENQ
SIRTVISGSAENLATLKAEWETHKRNVDTLFASGNAGLGFLDPTAAIVSSDTTA*GS*ANKPMQPITSTANKIVWSDPTRLSTTFSASLLRQ
RVKVGIAELNNVSGQYVSVYKRPAPKPEGCADACVIMPNENQSIRTVISGSAENLATLKAEWETHKRNVDTLFASGNAGLGFLDPTAAI
VSSDTTA* *Linker* in italics.

---

pET28a-SpyCatcher003-mi3 encoding GSS-H_6-SpyCatcher-mi3 has been previously described.[40]
>GSS-H_6-SpyCatcher-mi3(pET28a-SpyCatcher003-mi3)
ATGGGTTCAAGTCACCACCATCATCATCATGGCAGCGGCGTAACCACCTTATCAGGTTTATCAGGTGA
ACAAGGTCCGTCCGGTGATATGACAACTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTG
ATGAGGACGGCCGTGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGT
ACATGGATTTCAGATGGACATGTGAAGGATTTCTACCTGTATCCAGGAAAATATACATTTGTCGAAAC
CGCAGCACCAGACGGTTATGAGGTAGCAACTCCAATTGAATTTACAGTTAATGAGGACGGTCAGGTTA
CTGTAGATGGTGAAGCAACTGAAGGTGACGCTCATACTGGGGGCTCAGGGGGATCTGGCGGCTCTGG
CGGTTCAATGAAAATGGAAGAGCTTTTCAAGAAACATAAGATTGTCGCAGTATTGCGCGCAAACAGC
GTGGAGGAAGCGAAGAAAAAGGCTTTAGCAGTGTTCTTGGGAGGGGTGCATCTTATAGAGATCACGT
TCACAGTACCTGATGCGGATACTGTTATTAAGGAATTGTCCTTCTTAAAGGAAATGGGGGCCATTATT
GGTGCAGGGACAGTCACTTCCGTTGAGCAAGCCAGAAAAGCTGTGGAAAGCGGAGCTGAATTTATAG
TCTCCCCGCACTTGGACGAGGAGATCAGTCAATTTGCCAAAGAAAAAGGGGTATTTTATATGCCAGGT
GTAATGACACCTACCGAGTTAGTGAAAGCAATGAAGCTGGGTCACACTATATTAAAGTTGTTTCCGGG
GGAAGTTGTAGGGCCACAATTCGTTAAGGCTATGAAAGGGCCATTCCCAAATGTTAAATTCGTTCCTA
CTGGTGGTGTAAATCTTGACAATGTCTGTGAGTGGTTTAAAGCCGGTGTCCTTGCCGTAGGAGTCGGA
TCGGCACTGGTTAAGGGAACGCCCGTGGAGGTAGCGGAGAAGGCTAAGGCTTTCGTGGAAAAAATTC
GTGGCTGTACAGAATAATAA

M**GSSHHHHHH**GSGVTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATMELRDSSGKTIST
WISDGHVKDFYLYPGKYTFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGGSGGSGGSGGS
MKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSV
EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKA
MKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTE*

>GH_3-Qbeta(pKDB_307)
ATGGGTCATCATCATCCAGATCATCGTGCAAAATTAGAGACTGTTACTTTAGGTAACATCGGGAAAGATGGAAAACAAACTCTGG
TCCTCAATCCGCGTGGGGTAAATCCCACTAACGGCGTTGCCTCGCTTTCACAAGCGGGTGCAGTTCCTGCGCTGGAGAAGCGT
GTTACCGTTTCGGTATCTCAGCCTTCTCGCAATCGTAAGAACTACAAGGTCCAGGTTAAGATCCAGAACCCGACCGCTTGCACTG
CAAACGGTTCTTGTGACCCATCCGTTACTCGCCAGGCATATGCTGACGTGACCTTTTCGTTCACGCAGTATAGTACCGATGAGGA
ACGAGCTTTTGTTCGTACAGAGCTTGCTGCTCTGCTCGCTAGTCCTCTGCTGATCGATGCTATTGATCAGCTGAACCCAGCGTAT
TAA

M**GHHH***PDHR*AKLETVTLGNIGKDGKQTLVLNPRGVNPTNGVASLSQAGAVPALEKRVTVSVSQPSRNRKNYKVQVKIQNPTACTANG
SCDPSVTRQAYADVTFSFTQYSTDEERAFVRTELAALLASPLLIDAIDQLNPAY*
Note: GHHH-Qbeta expressed well (as judged by reducing SDS-PAGE), but did not assemble into VLPs (as determined by NAGE). The PDHR-linker is derived from an assembling HiPerX MS2-variant.[35]

>GH_3-2PP7(pKDB_158)
ATGGGTCATCATCATGGTGGTCCAAGTGAAAGTGGTTCCAAAACCATCGTTCTGTCGGTCGGCGAGGCTACTCGCACTCTGACT
GAGATCCAGTCCACCGCAGACCGTCAAATCTTTGAAGAGAAGGTCGGGCCTCTGGTGGGTCGTCTGCGCCTGACGGCTTCGCT
GCGTCAAAACGGTGCCAAGACCGCGTATCGCGTCAACCTGAAACTGGATCAGGCGGACGTCGTTGATTGCTCCACCAGCGTCT
GCGGCGAGCTGCCGAAAGTGCGCTACACTCAGGTATGGTCGCACGACGTGACAATCGTTGCGAATAGCACCGAGGCCTCGCG
CAAATCGCTGTACGATCTGACCAAGTCCCTGGTCGCGACCTCGCAGGTCGAAGATCTGGTCGTCAACCTGGTGCCGCTGGGCC
GTGCATATGGTGGTTCCAAAACGATTGTGCTGAGTGTAGGTGAAGCAACCCGTACCCTGACGGAAATTCAAAGCACTGCGGATC
GCCAAATTTTTGAGGAAAAAGTGGGTCCGCTGGTTGGGCGCCTGCGTCTGACAGCGAGCCTGCGCCAGAATGGCGCTAAAACT
GCCTACCGCGTGAACCTGAAGCTGGACCAAGCCGATGTAGTGGACTGCAGCACGAGTGTGTGTGGTGAACTGCCTAAGGTTCG
TTATACACAAGTTTGGAGCCATGATGTTACCATTGTGGCTAATTCAACGGAAGCGAGTCGTAAGTCCCTGTATGACCTGACCAAA
AGCCTGGTGGCAACTAGCCAAGTGGAGGATCTGGTCGTAAATCTGGTTCCGCTGGGCCGTTAA

M**GHHH***GGPSESG*SKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRLRLTASLRQNGAKTAYRVNLKLDQADVVDCSTSVCGELP
KVRYTQVWSHDVTIVANSTEASRKSLYDLTKSLVATSQVEDLVVNLVPLGRAYGGSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLV
GRLRLTASLRQNGAKTAYRVNLKLDQADVVDCSTSVCGELPKVRYTQVWSHDVTIVANSTEASRKSLYDLTKSLVATSQVEDLVVNLV
PLGR*

Note: GH_3-2PP7[42] expressed well (SDS-PAGE), likely assembled into VLPs (as determined by NAGE post IMAC), but was prone to aggregation/precipitation in 40 mM Tris-HCl, 300 mM NaCl, 0.5 M imidazole (HCl), pH 7.5; when stored at 4 ºC within days. It is possible that different linkers, buffer screening, and/or removal of intermediate VLP-assembly products by size-exclusion/ultra-centrifugation could solve the precipitation issues.

>hRBD (obtained from Absolute Antibodies).

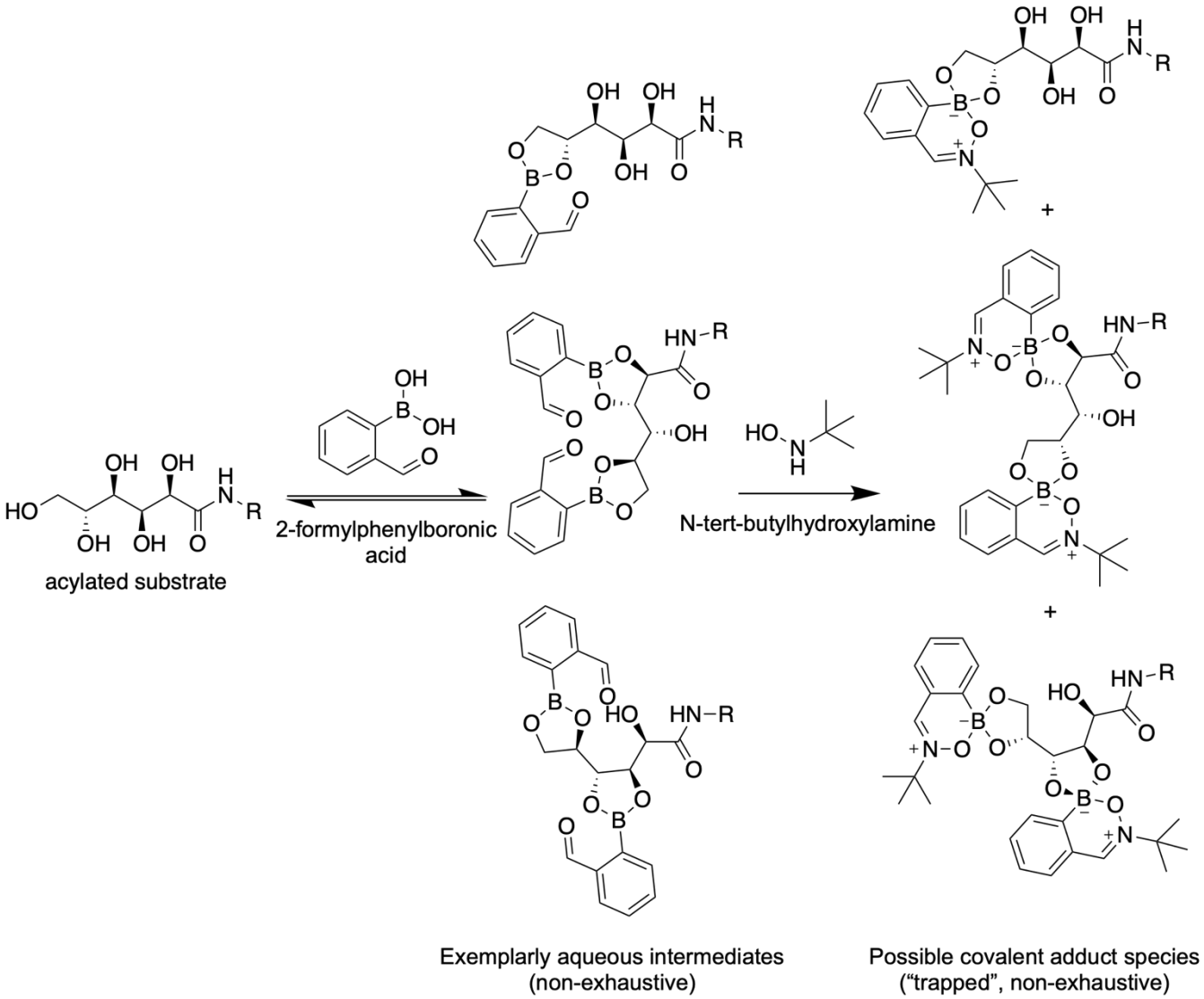
GHHHHHHTNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQ
TGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNG
VGYQPYRVVVLSFELLHAPATVCG*

>yRBD [cloned into pPICα (Thermo Fisher Scientific, US)]
ATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGA
AACGGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAAC
AGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAG
GCGGTCACCATCACCATCACCATACCAACCTGTGCCCGTTTGGCGAAGTGTTTAACGCGACCCGCTTTGCGAGCGTGTATGCGT
GGAACCGCAAACGCATTAGCAACTGCGTGGCGGATTATAGCGTGCTGTATAACAGCGCGAGCTTTAGCACCTTTAAATGCTATG
GCGTGAGCCCGACCAAACTGAACGATCTGTGCTTTACCAACGTGTATGCGGATAGCTTTGTGATTCGCGGCGATGAAGTGCGCC
AGATTGCGCCGGGCCAGACCGGCAAAATTGCGGATTATAACTATAAACTGCCGGATGATTTTACCGGCTGCGTGATTGCGTGGA
ACAGCAACAACCTGGATAGCAAAGTGGGCGGCAACTATAACTATCTGTATCGCCTGTTTCGCAAAAGCAACCTGAAACCGTTTGA
ACGCGATATTAGCACCGAAATTTATCAGGCGGGCAGCACCCCGTGCAACGGCGTGGAAGGCTTTAACTGCTATTTTCCGCTGCA
GAGCTATGGCTTTCAGCCGACCAACGGCGTGGGCTATCAGCCGTATCGCGTGGTGGTGCTGAGCTTTGAACTGCTGCATGCGC
CGGCGACCGTGTGCGGCTAA

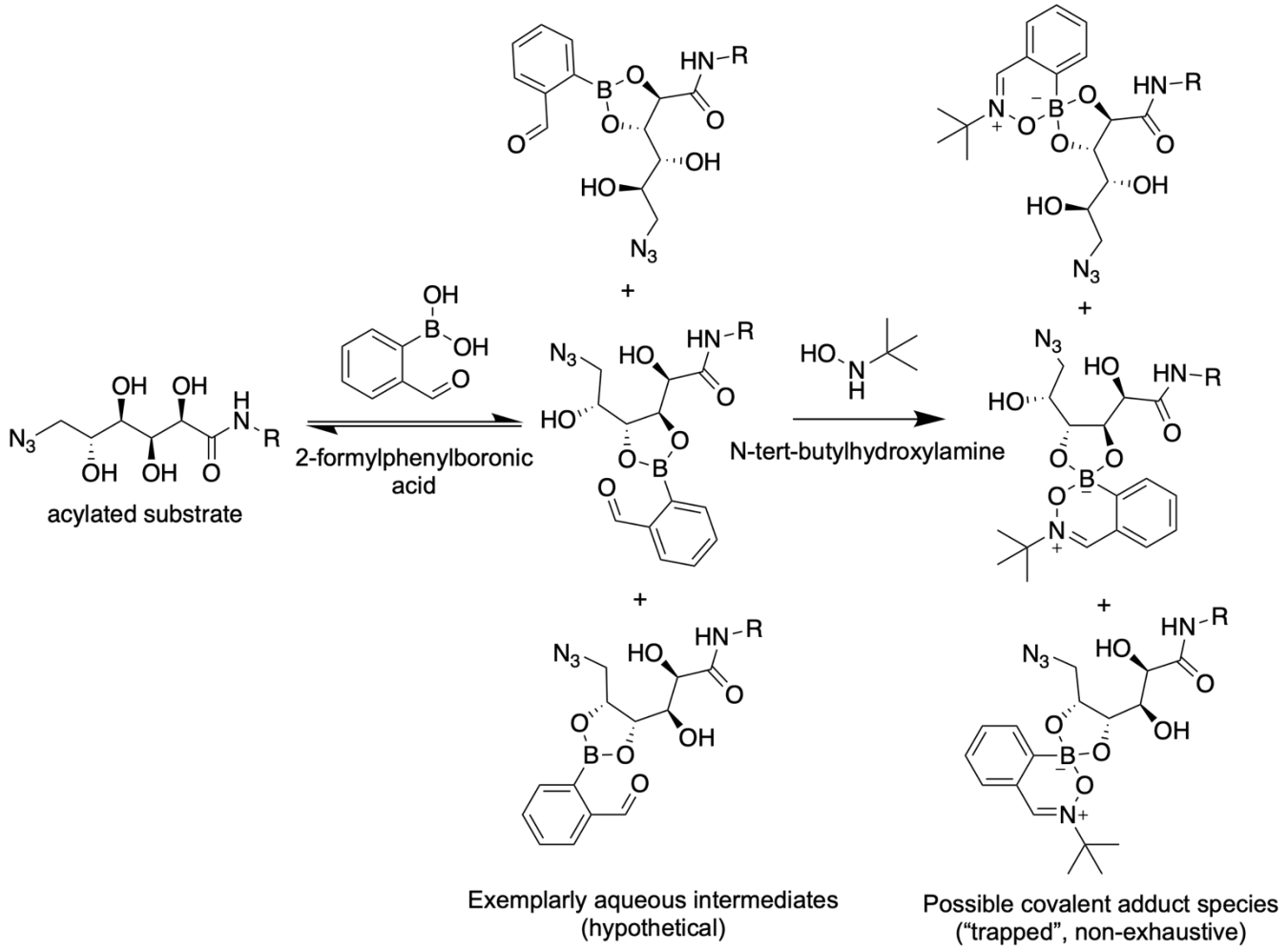
MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGSLEKR^**GGHHHH**
HHTNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQP
YRVVVLSFELLHAPATVCG*
Yellow: α-factor signal sequence (cleaved off during secretion), ^indicates cleavage site; N-terminal peptide of secretion product found to be GGHH… by MS/MS sequencing; RBD sequence of Wuhan isolate include S protein residues 333 to 526.

Atezolizumab
>Atezolizumab_heavy_chain
GHHHHHHEVQLVESGGGLVQPGGSLRLSCAASGFTFSDSWIHWVRQAPGKGLEWVAWISPYGGSTYYADSVKGRFTISADTSKNTA
YLQMNSLRAEDTAVYYCARRHWPGGFDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALT
SGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMI
SRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYASTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAK
GQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCS
VMHEALHNHYTQKSLSLSPGK
>Atezolizumab_light_chain
DIQMTQSPSSLSASVGDRVTITCRASQDVSTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQ
QYLYHPATFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSS
TLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Supporting Schemes



Scheme S4. **Possible formation of 2-formylpheylboronic acid : gluconoyl diol esters and trapping with *N*-tert-buthylhydroxylamine.**



Scheme S5. **Possible formation of 2-formylpheylboronic acid : azidogluconoyl diol esters and trapping with *N*-tert-buthylhydroxylamine.**

Supporting Figures

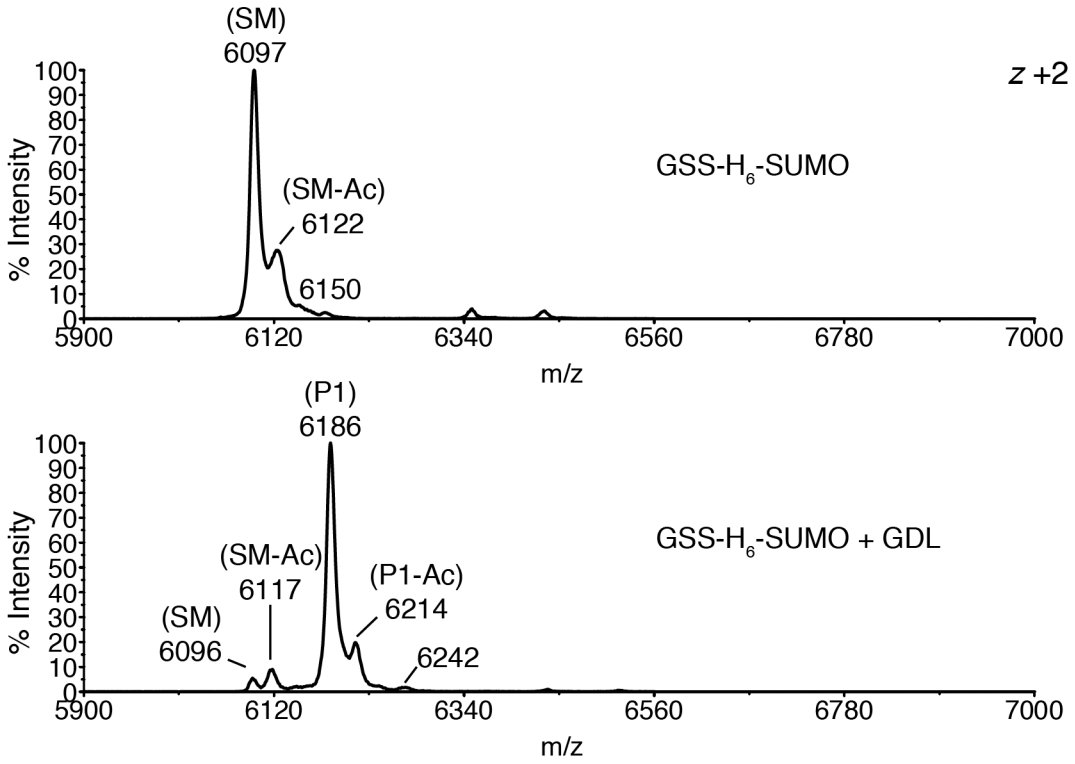


Figure S1. **GSS-H_6-SUMO.** 20 µM protein was labelled with 100 m GDL at pH 7.5 (HEPES/NaOH buffer) for 1 h at RT. Spectra were obtained after C18ZipTipping.

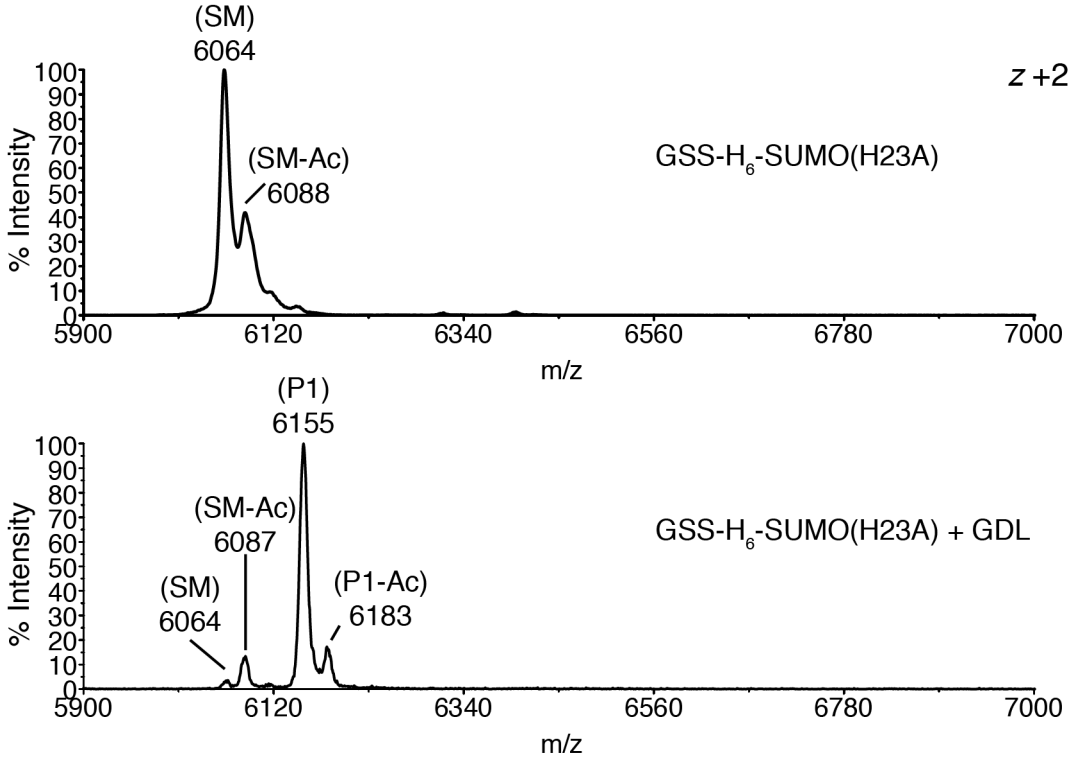


Figure S2. **GSS-H_6-SUMO(H23A).** 20 µM protein was labelled with 100 m GDL at pH 7.5 (HEPES/NaOH buffer) for 1 h at RT. Spectra were obtained after C18ZipTipping.

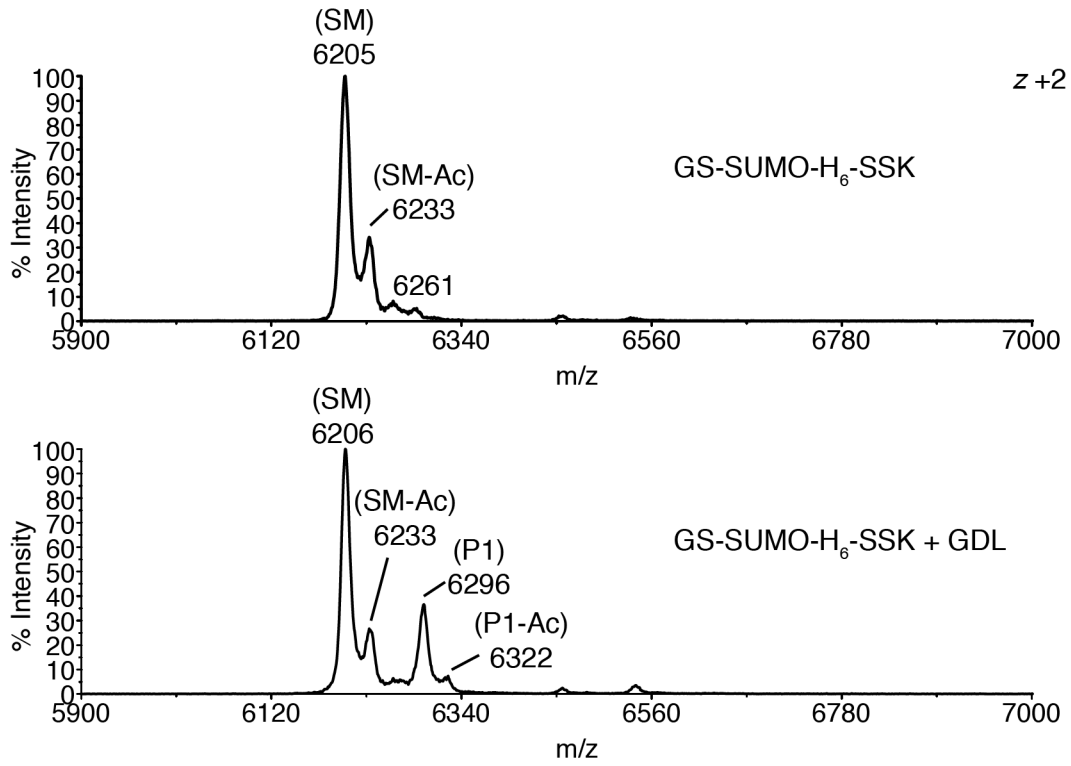


Figure S3. **GS-SUMO-H_6-SSK.** 20 µM protein was labelled with 100 m GDL at pH 7.5 (HEPES/NaOH buffer) for 1 h at RT. Spectra were obtained after C18ZipTipping.

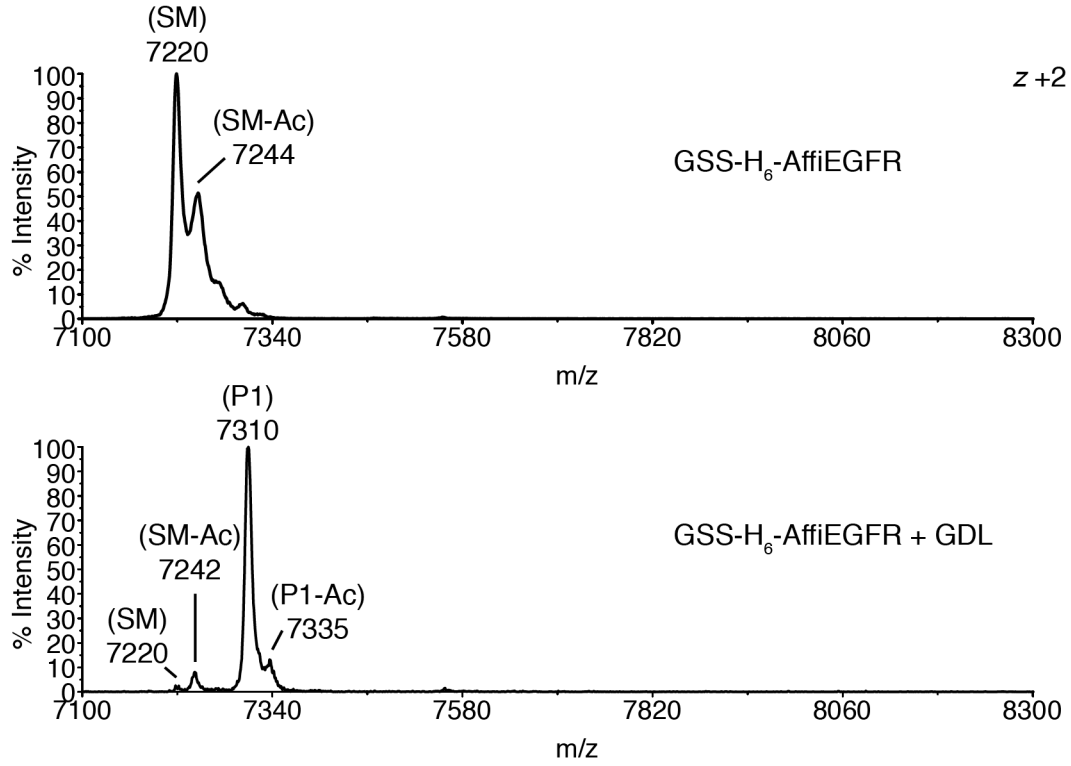


Figure S4. **GSS-H_6-AffiEGFR.** 20 µM protein was labelled with 100 m GDL at pH 7.5 (HEPES/NaOH buffer) for 1 h at RT. Spectra were obtained after C18ZipTipping.

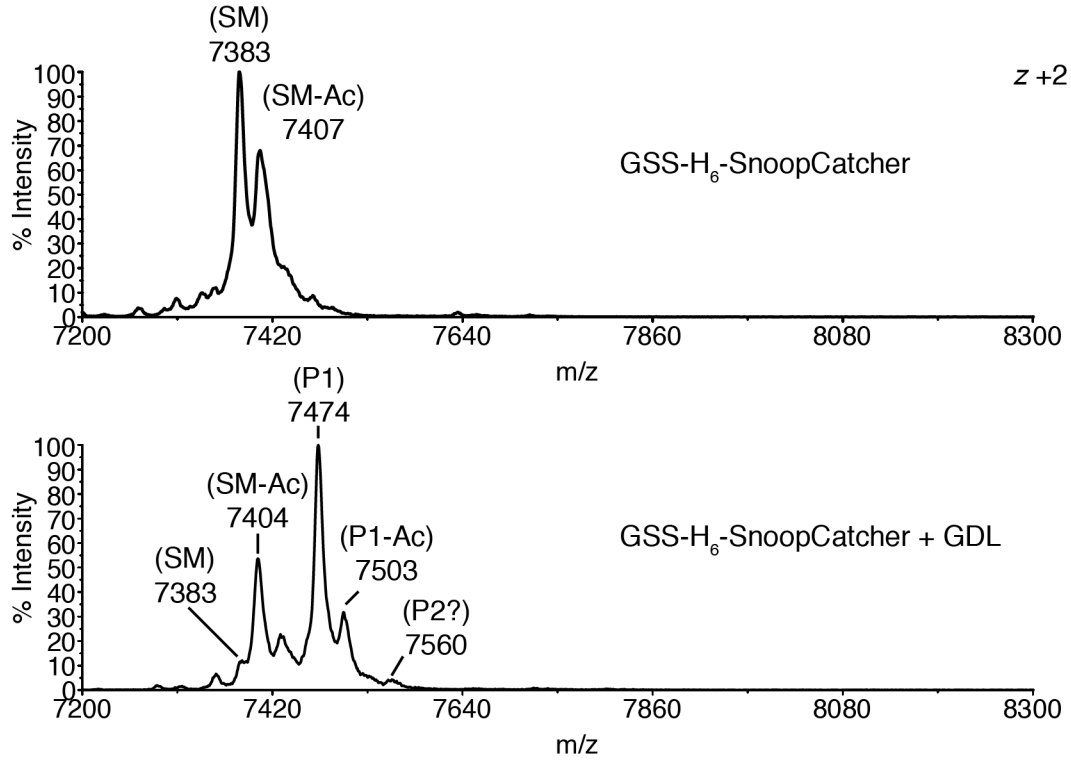


Figure S5. **GSS-H_6-SnoopCatcher.** 20 µM protein was labelled with 100 m GDL at pH 7.5 (HEPES/NaOH buffer) for 1 h at RT. Spectra were obtained after C18ZipTipping.

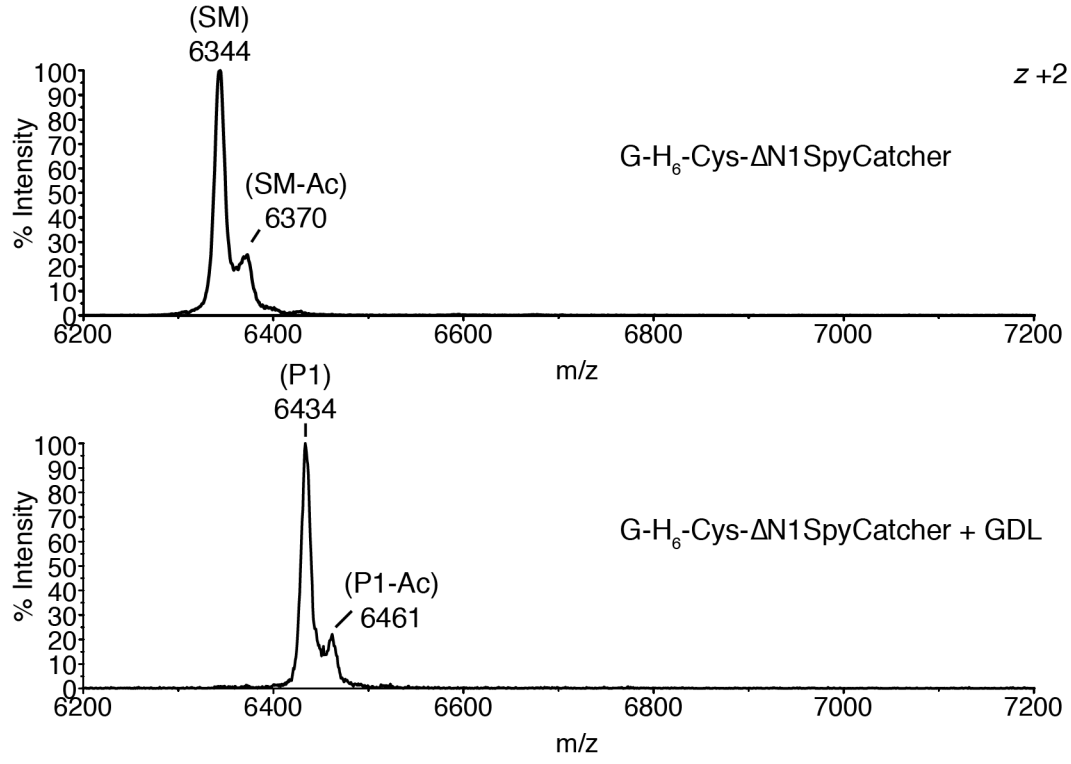


Figure S6. **G-H_6-C-ΔN1SpyCatcher.** 20 µM protein was labelled with 100 m GDL at pH 7.5 (HEPES/NaOH buffer) for 1 h at RT. Spectra were obtained after C18ZipTipping.

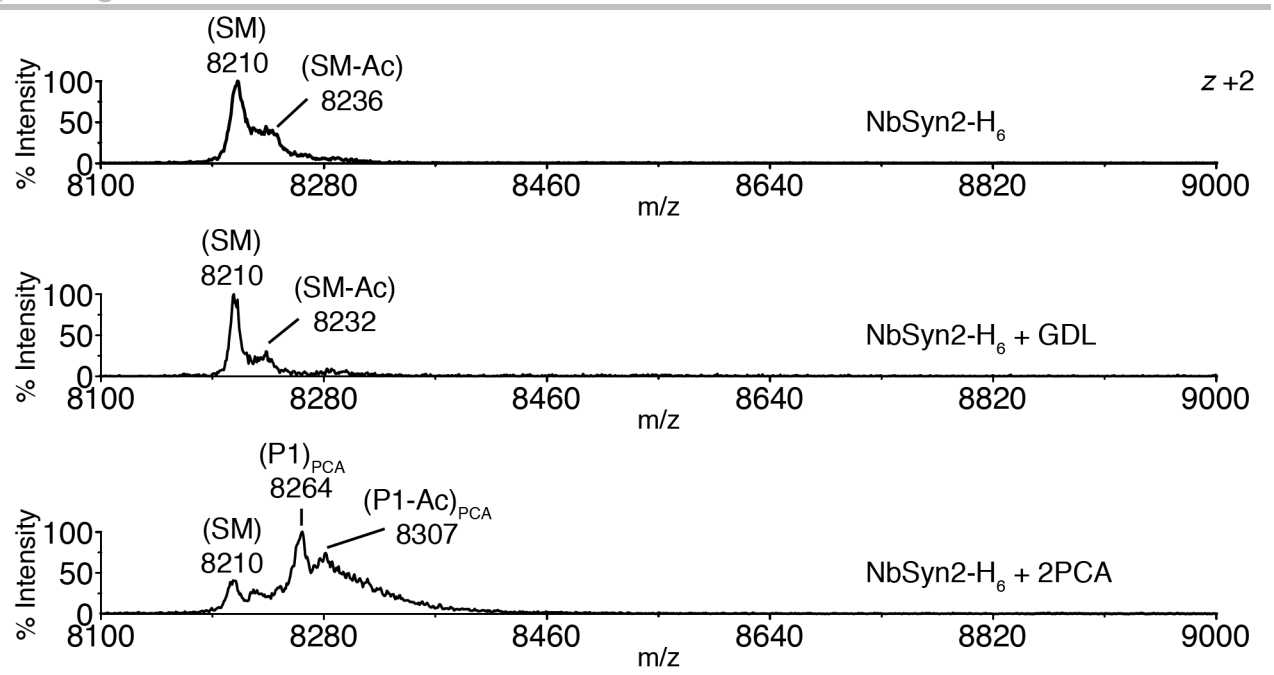


Figure S7. **NbSyn2-H_6 is a poor substrate for GDL labelling.** 20 µM protein was labelled with 100 m GDL at pH 7.5 (HEPES/NaOH buffer) for 1 h at RT. Spectra were obtained after C18ZipTipping. N-terminal amino group availability is demonstrated by labelling with 2PCA in a separate experiment.

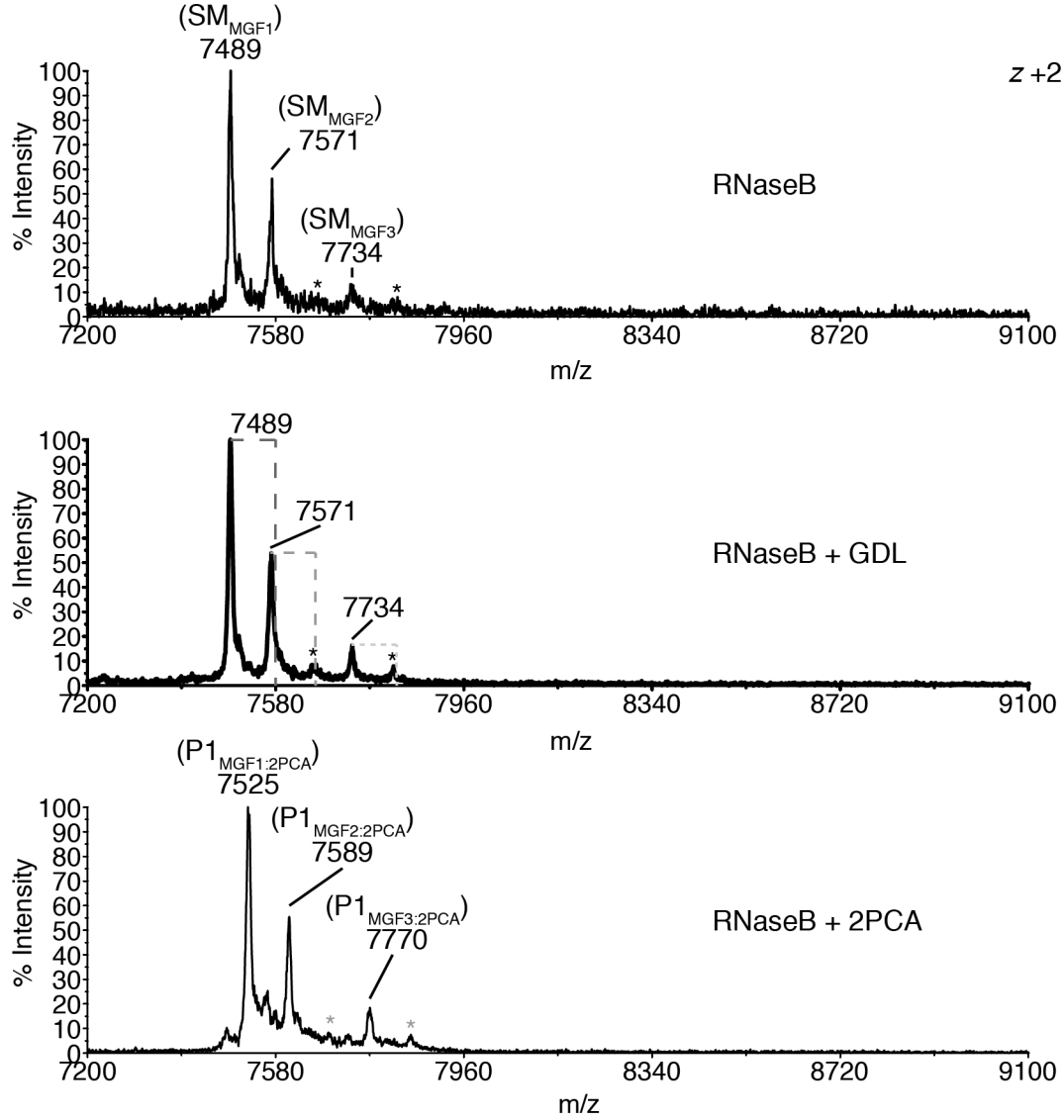


Figure S8. **RNAseB is a poor substrate for GDL labelling.** 20 µM protein was labelled with 100 m GDL at pH 7.5 (HEPES/NaOH buffer) for 1 h at RT. Spectra were obtained after C18ZipTipping. N-terminal amino group availability is demonstrated by labelling with 2PCA in a separate experiment.

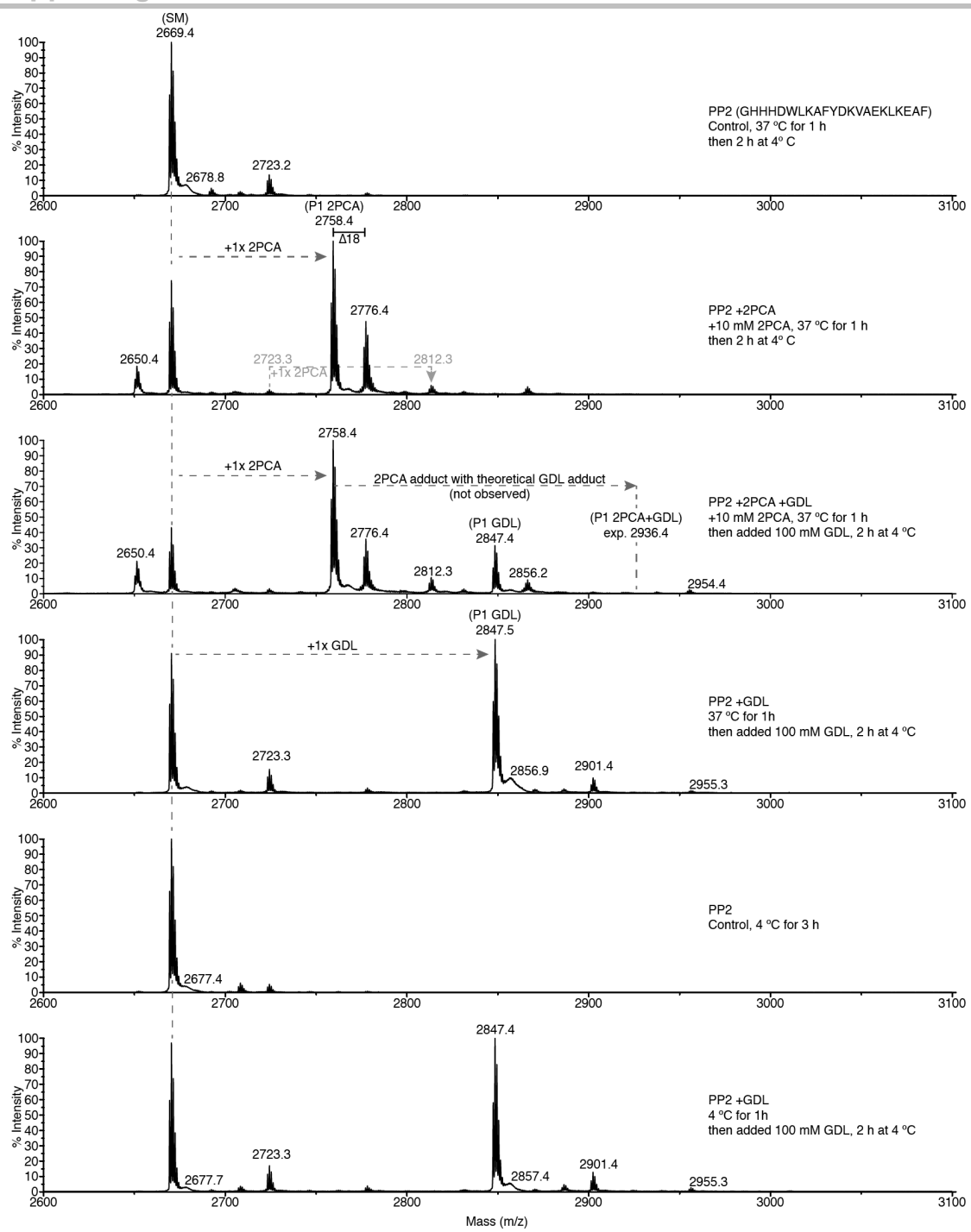


Figure S9**. Pre-reaction with 2PCA precludes subsequent labeling with GDL.** 100 µM peptide PP2 (GHHHDWLKAFYDKVAEKLKEAF) was incubated with 10 mM 2PCA in 200 mM potassium phosphate buffer (KOH) pH 7.5 for 1 h at 37 ºC. GDL was then added to the 2PCA-treated reaction at 100 mM, and the reaction was allowed to proceed for 2 h at 4 ºC. In parallel, the same peptide without prior 2PCA treatment was also incubated with 100 mM GDL. A loss of 19 amu is observed in 2PCA treated material that we cannot explain (*m/z* 2650.4, no precursor with *m/z* 2561.4 is observed in the control either) . A control reaction at the same temperature (37 ºC) does not show this mass loss.

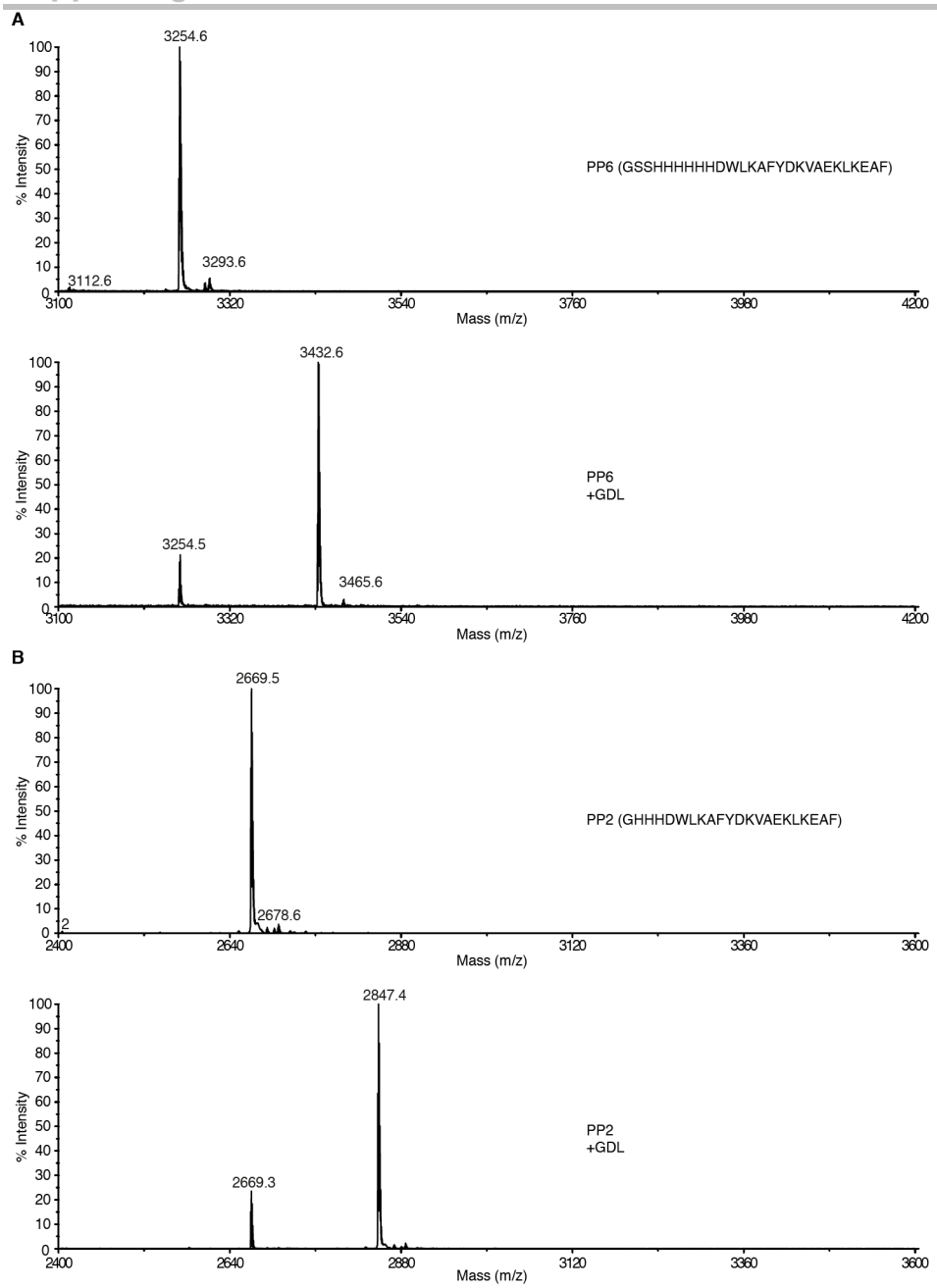


Figure S10 **Suboptimal labelling to discern reactivity differences of various peptides. (A)** PP6 and **(B)** PP2. Top: native, and bottom: the peptides were reacted at 50 μM concentration with 50 mM GDL for 1h in potassium phosphate buffer (pH 7.5) at room temperature (22 ºC).

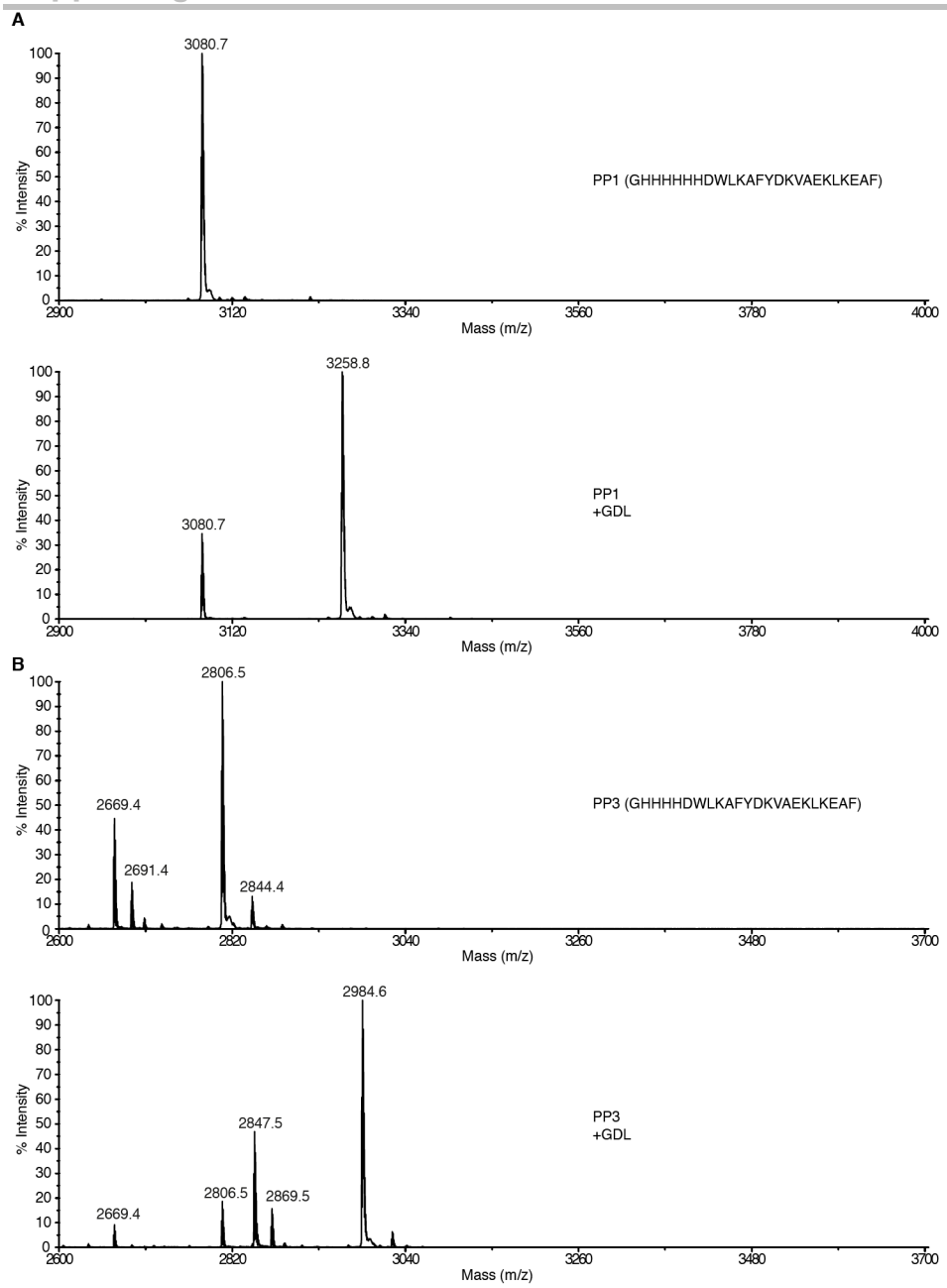


Figure S11 **Suboptimal labelling to discern reactivity differences of various peptides. (A)** PP1 and **(B)** PP3. Top: native, and bottom: the peptides were reacted at 50 μM concentration with 50 mM GDL for 1h in potassium phosphate buffer (pH 7.5) at room temperature (22 ºC).

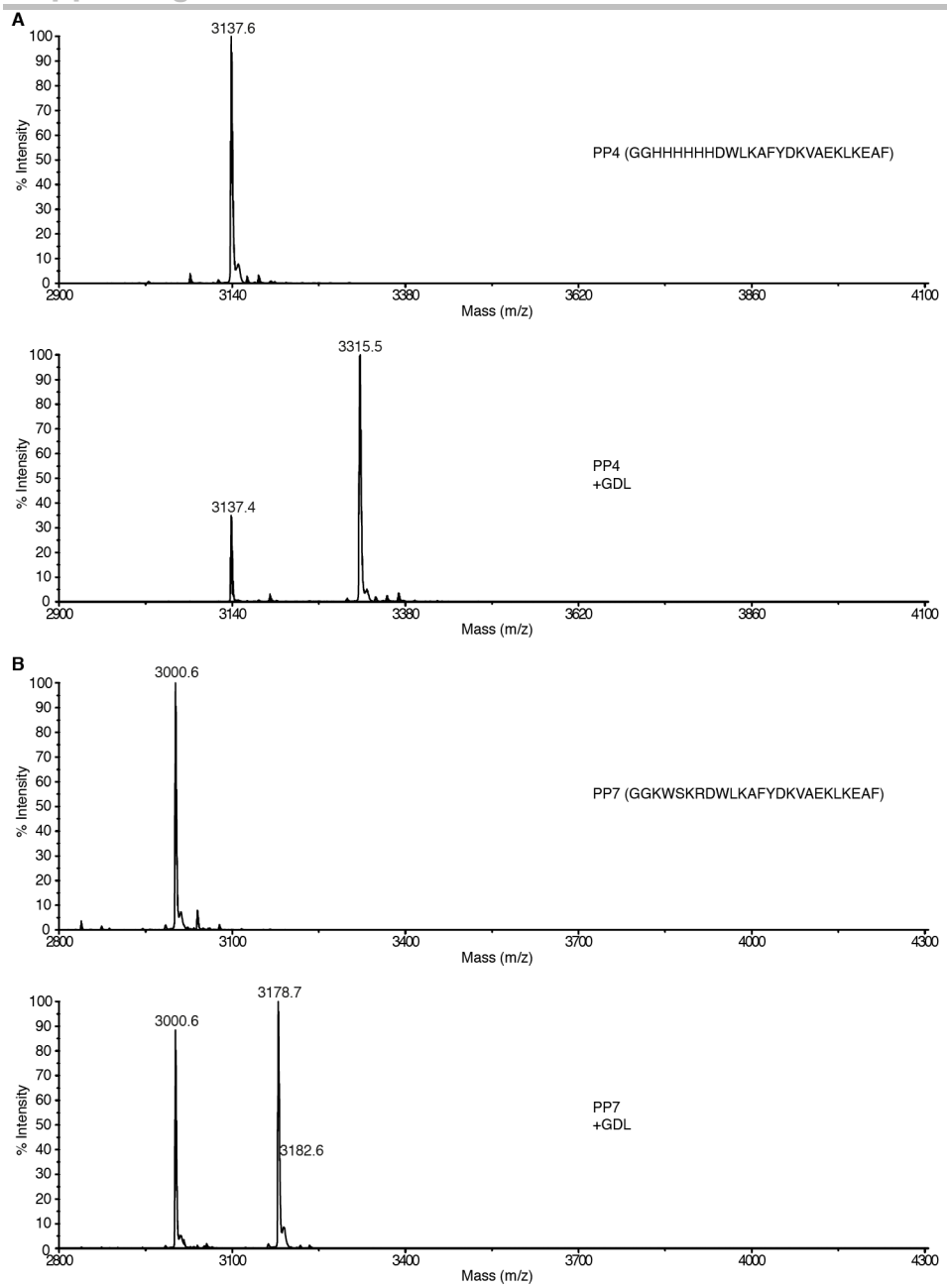


Figure S12 **Suboptimal labelling to discern reactivity differences of various peptides. (A)** PP4 and **(B)** PP7. Top: native, and bottom: the peptides were reacted at 50 μM concentration with 50 mM GDL for 1h in potassium phosphate buffer (pH 7.5) at room temperature (22 ºC).

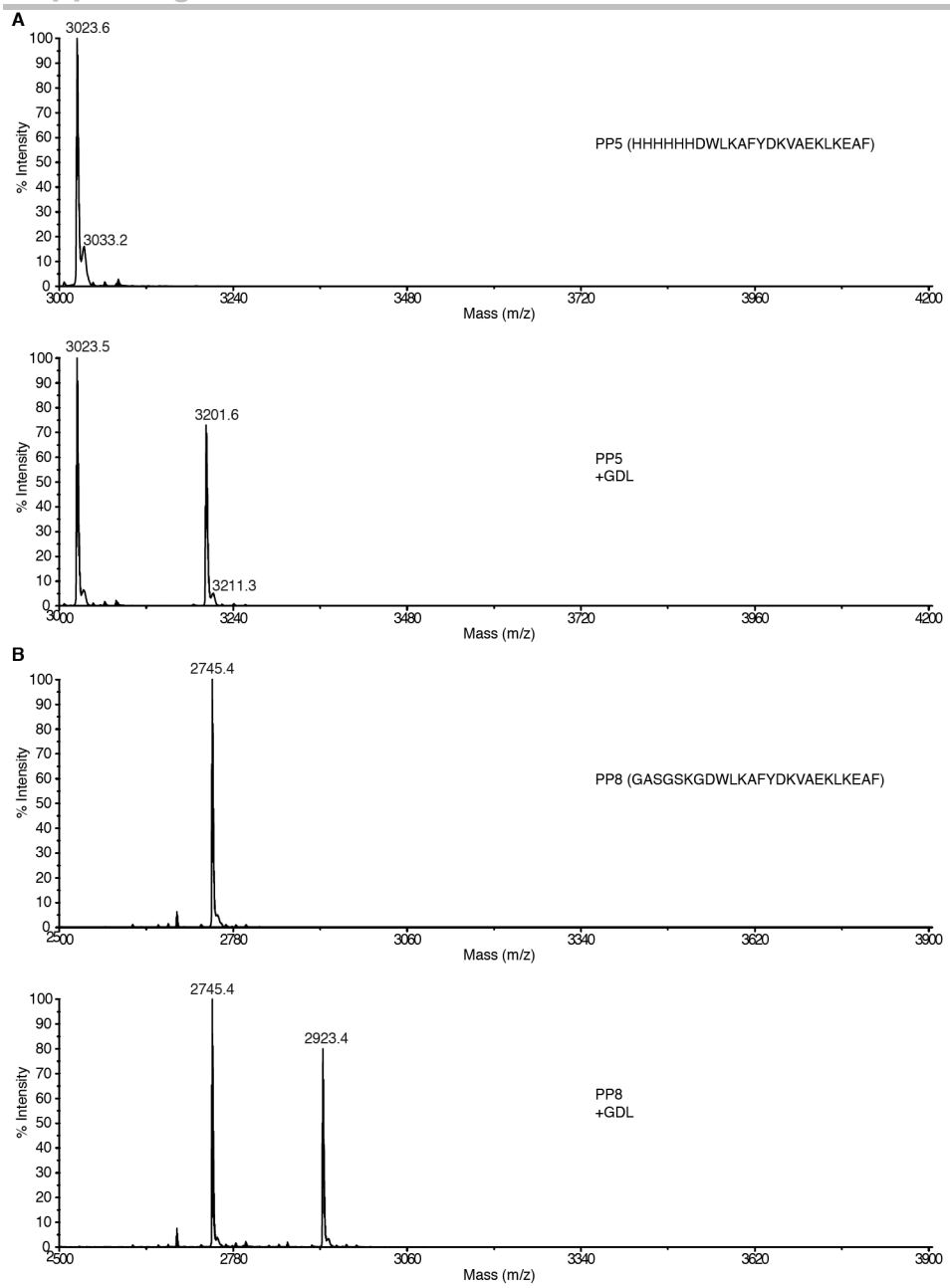


Figure S13 **Suboptimal labelling to discern reactivity differences of various peptides. (A)** PP5 and **(B)** PP8. Top: native, and bottom: reacted with 50 mM GDL for 1h at pH 7.5.

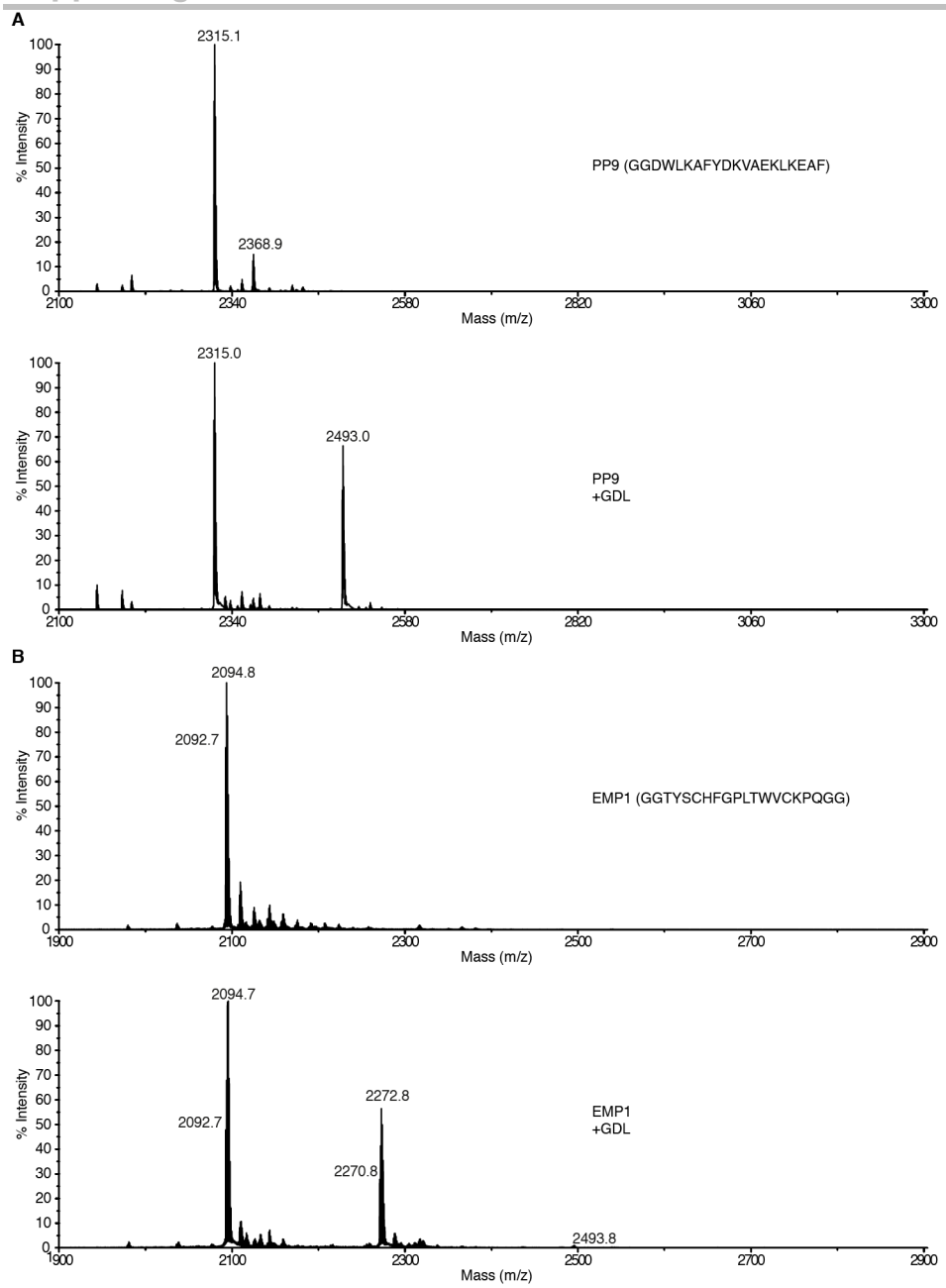


Figure S14 **Suboptimal labelling to discern reactivity differences of various peptides. (A)** PP9 and **(B)** EMP1. Top: native, and bottom: the peptides were reacted at 50 µM concentration with 50 mM GDL for 1h in potassium phosphate buffer (pH 7.5) at room temperature (22 ºC).

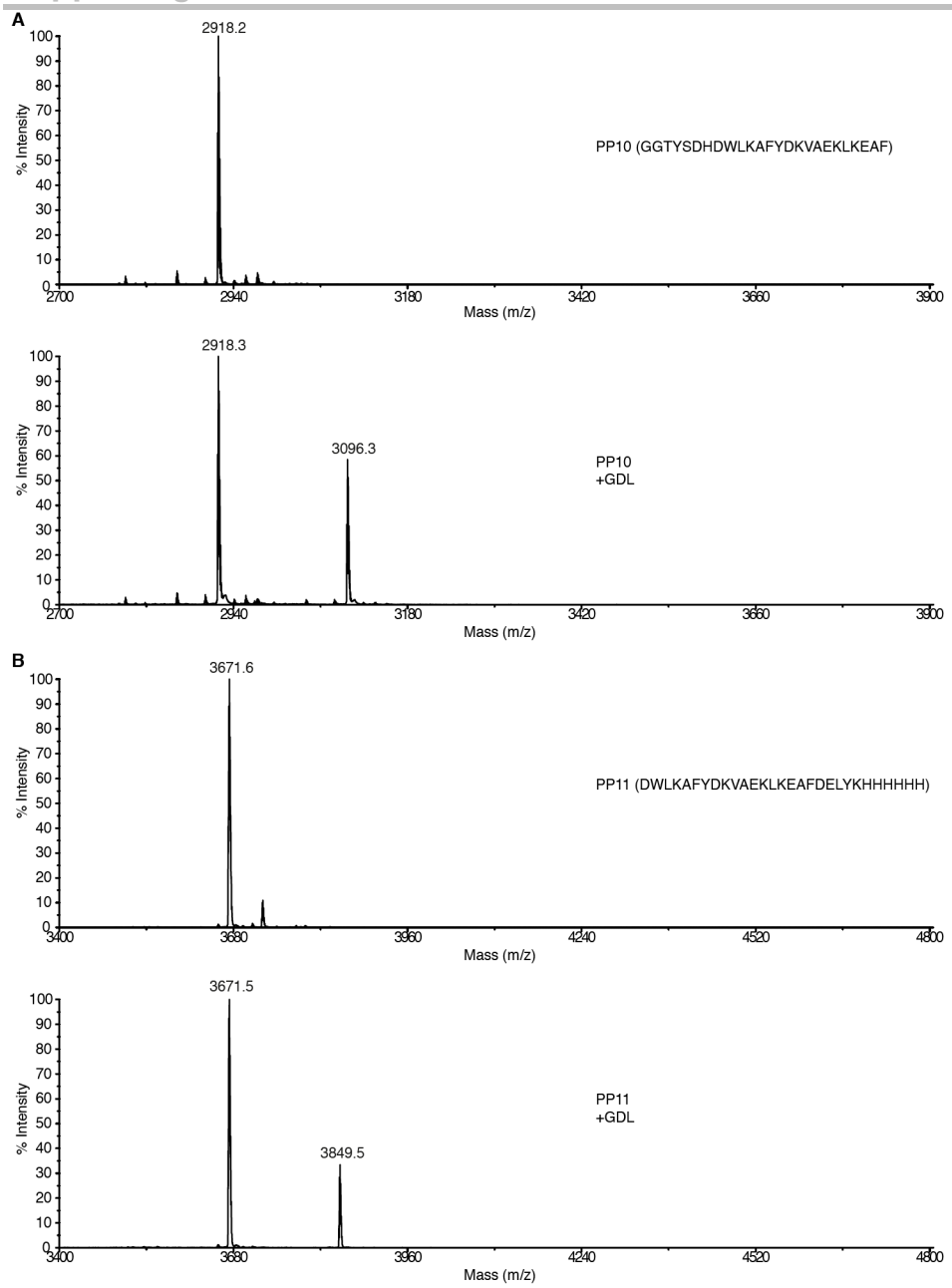


Figure S15 **Suboptimal labelling to discern reactivity differences of various peptides. (A)** PP10 and **(B)** PP11. Top: native, and bottom: the peptides were reacted at 50 μM concentration with 50 mM GDL for 1h in potassium phosphate buffer (pH 7.5) at room temperature (22 ºC).

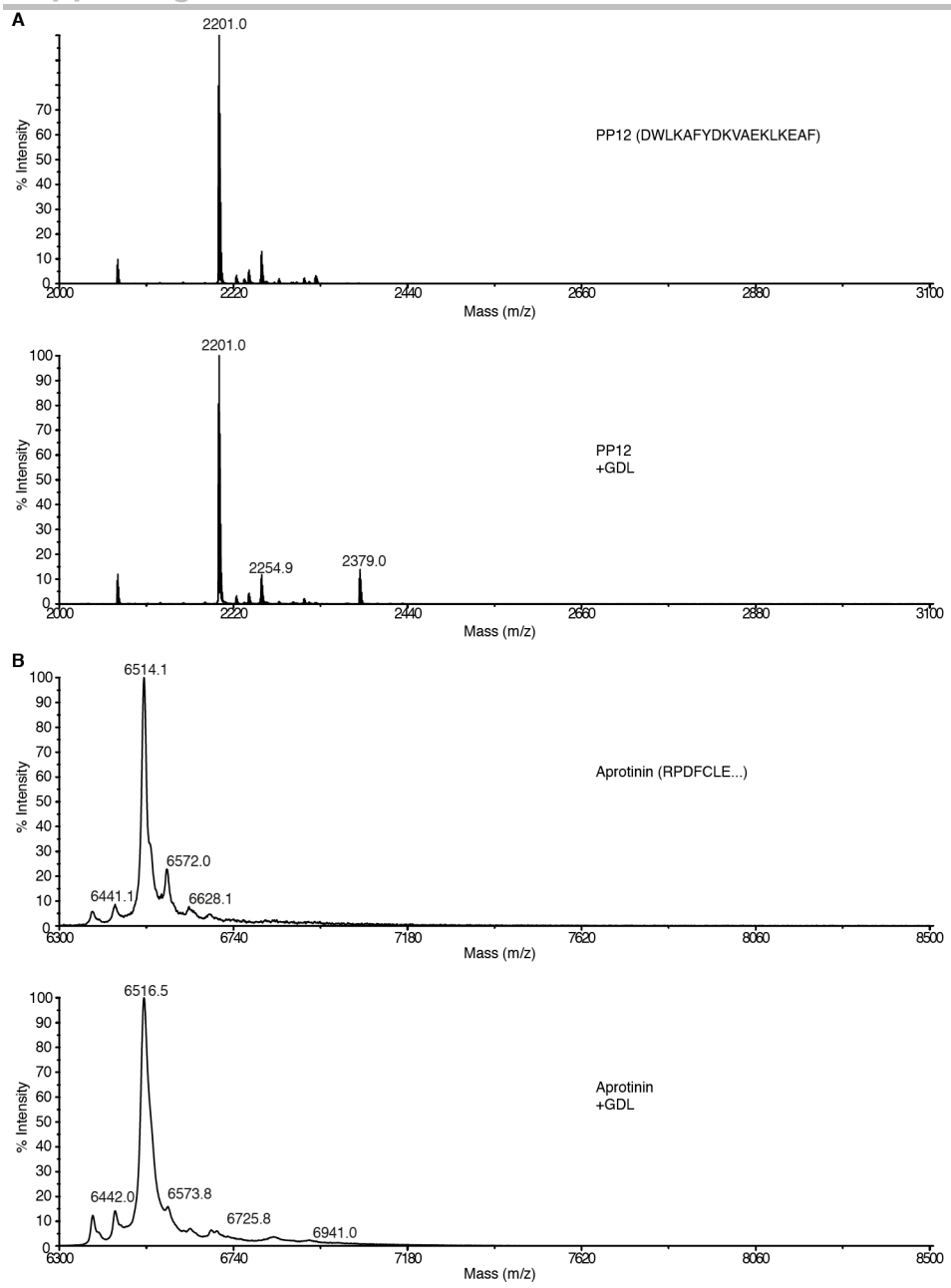


Figure S16 **Suboptimal labelling to discern reactivity differences of various peptides. (A)** PP12 and **(B)** Aprotinin, which is poorly reactive. Top: native, and bottom: the peptides were reacted at 50 µM concentration with 50 mM GDL for 1h in potassium phosphate buffer (pH 7.5) at room temperature (22 ºC).

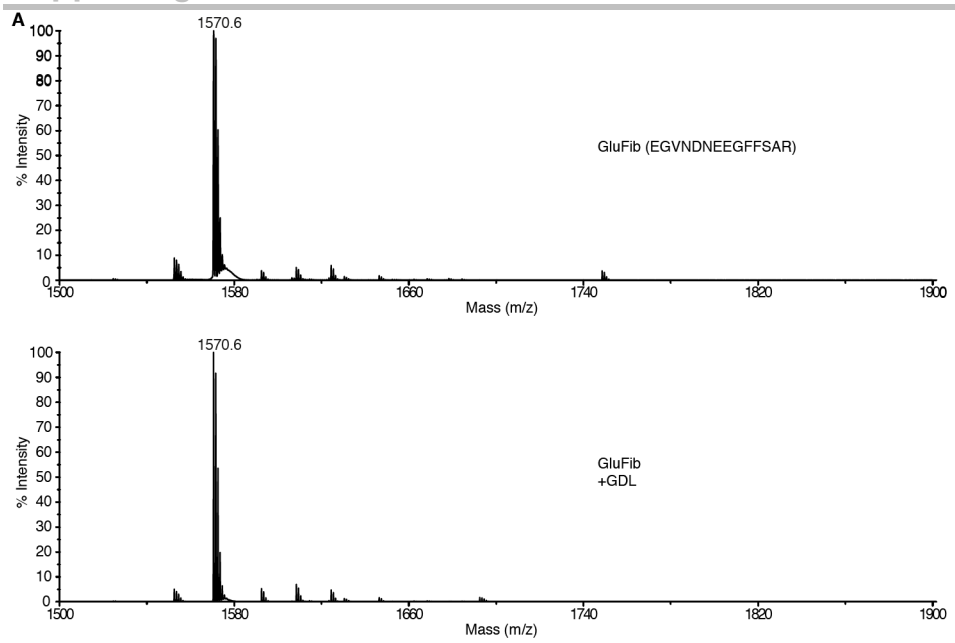


Figure S17 **Suboptimal labelling to discern reactivity differences of various peptides.** GluFib is a poor substrate for gluconoylation. Top: native, and bottom: the peptides were reacted at 50 µM concentration with 50 mM GDL for 1h in potassium phosphate buffer (pH 7.5) at room temperature (22 ºC).

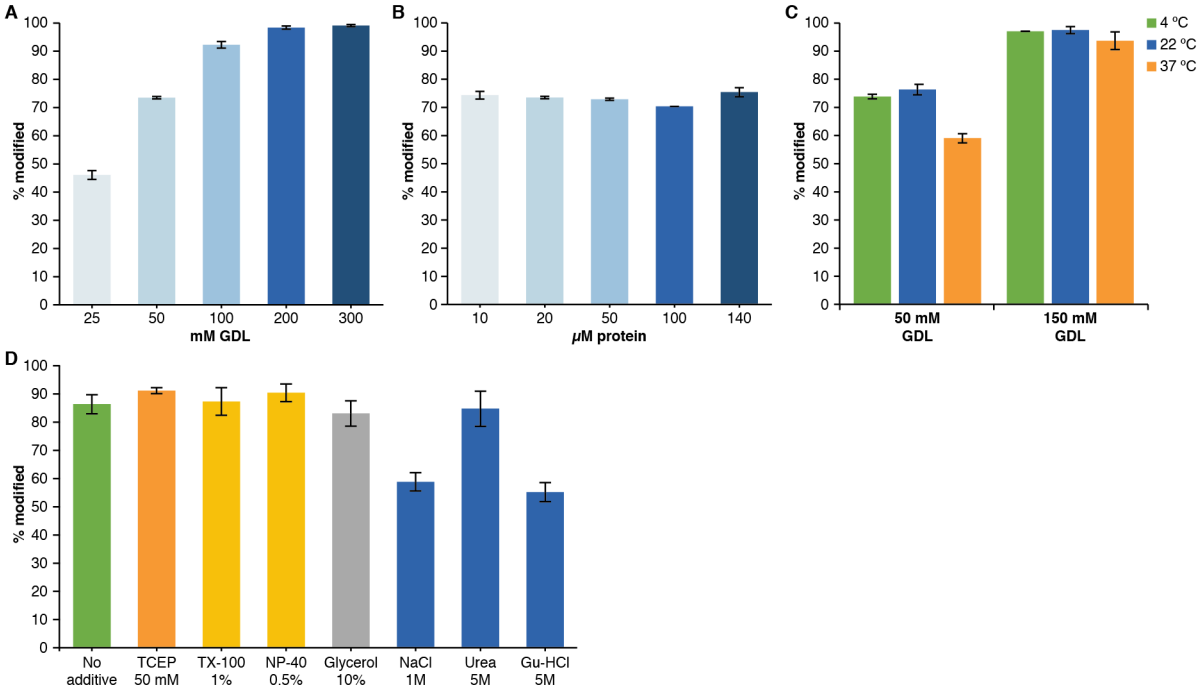


Figure S18. **Parameters influencing GDL-labelling yields.** 20 μM GSS-H_6-AffiEGFR was reacted in 500 mM HEPES, pH 7.5 (NaOH) for 1 h at room temperature (22 ºC), unless otherwise specified. **(A)** Varying GDL labelling reagent concentration with fixed protein substrate concentration (20 μM). **(B)** Varying protein substrate concentration with fixed, suboptimal GDL labelling reagent concentration (50 mM). **(C)** Varying temperature with categorically fixed GDL labelling reagent concentrations. **(D)** 20 μM GSS-H_6-AffiEGFR in 100 mM HEPES (pH 7.5) was mixed with additives, and 75 mM GDL (suboptimal) were added to the reactions. The reactions were monitored by MALDI-TOF MS, and are reported as the relative m/*z* ratio of the TIC for the acylated (m_1) and non-acylated species (m_2) ([$m_{1/z}$/($m_{1/z}$+$m_{2/z}$)]*100%). Additive percentages are reported as v/v. Tris(2-carboxyethyl)phosphine (TCEP), Triton X-100 (TX-100), Nonoxynol-40 (NP-40), Guanidine (Gu-).

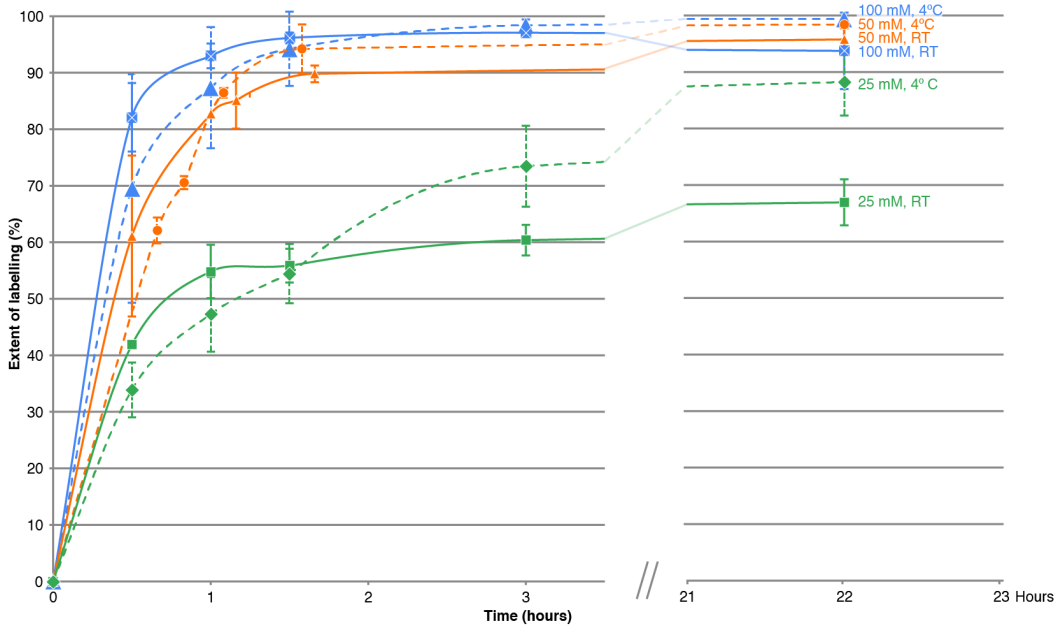


Figure S19. **Extent of labeling of PP13 (GSSHHHHHHGGTYSAHFGPLTWVAKPQGG) with various GDL concentrations over time.** 250 µM PP13 was labeled with 25, 50 and 100 mM GDL in 200 mM HEPES (pH 7.5, NaOH), at both 4 ºC (dashed lines) and RT (solid lines). The extent of labeling was analyzed using reflectron, positive mode MALDI-TOF mass spectrometry and the extent of labeling was calculated as the XIC for the adduct divided by the sum of the XIC for the adduct and native substrate. The experiment was performed in triplicate and error bards show the standard deviation form the mean (n=3). Points were connected with Excel smooth line function for helping visual interpretation and do not indicate continuous data, nor curve fitting.

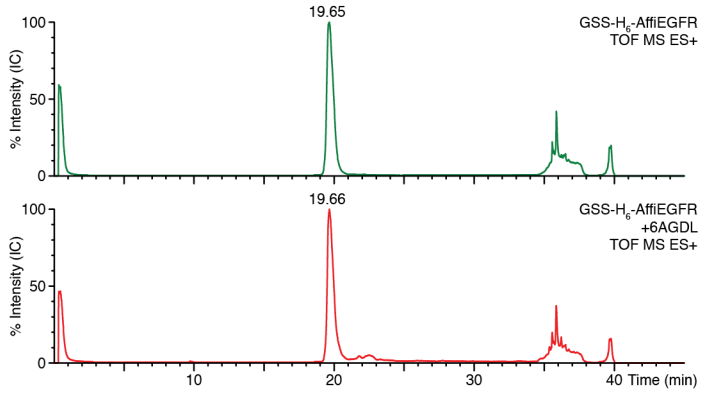


Figure S20. **LC-MS elution chromatogram (TIC) for the unlabelled (top) and 6AGDL-labelled (bottom) GSS-H_6-AffiEGFR protein.** Both unlabelled and labelled protein elute at around 19.7 min. Flow rate 0.35 mL min-1. Solvent A: water containing 0.1% formic acid (v/v); Solvent B: acetonitrile containing 0.1% formic acid (v/v). The column was eluted using a linear gradient from 0 to 100% of solvent B over a period of 40 minutes.

Note: Intact, labelled protein was chromatographically indistinguishable from unlabelled material by LC-MS (C18 revers phase, Figure S20). Modified, smaller peptides co-eluted slightly later.

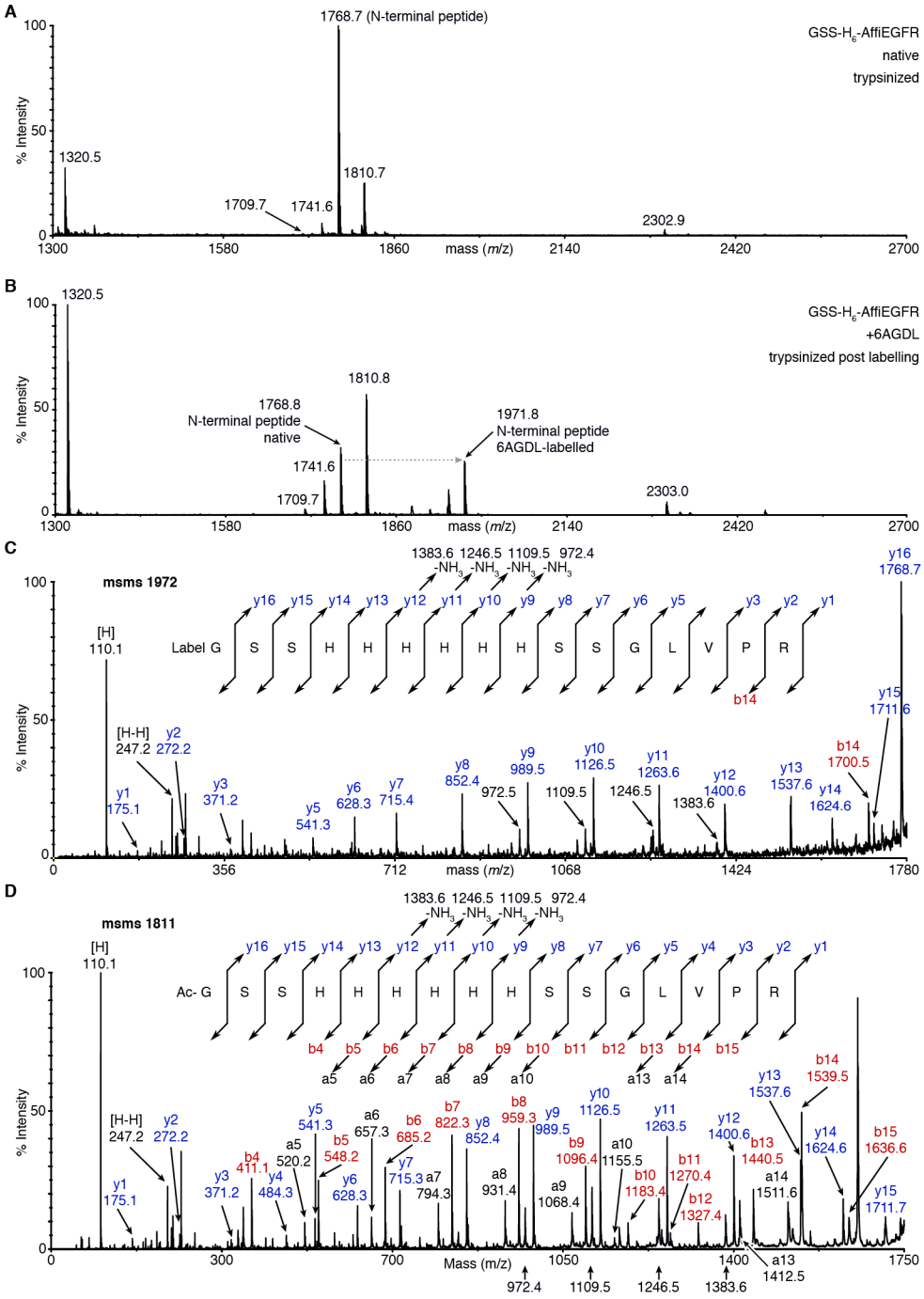


Figure S21. **MALDI TOF MS and MALDI TOF-TOF MS/MS of 6AGDL-labelled, trypsin-digested GSS-H$_6$-AffiEGFR. (A)** MALDI-TOF spectrum (positive mode, reflectron) of unlabelled and **(B)** 6AGDL-labelled material. The peak at *m/z* 1769 (+1) is the unmodified N-terminal fragment (GSS-H$_6$.SSGLVPR) and the peak at *m/z* 1972 (+1) is a singly-labelled 6AGDL species. **(C)** MALDI TOF-TOF MS/MS spectrum (positive mode, reflectron) of the peak at *m/z* 1972 showing the N-terminal position of the label. **(D)** MALDI TOF-TOF MS/MS spectrum (positive mode, reflectron) of the peak at *m/z* 1811 corresponding to the host-acetylated N-terminal peptide, *i.e.* which does not react with 6AGDL.

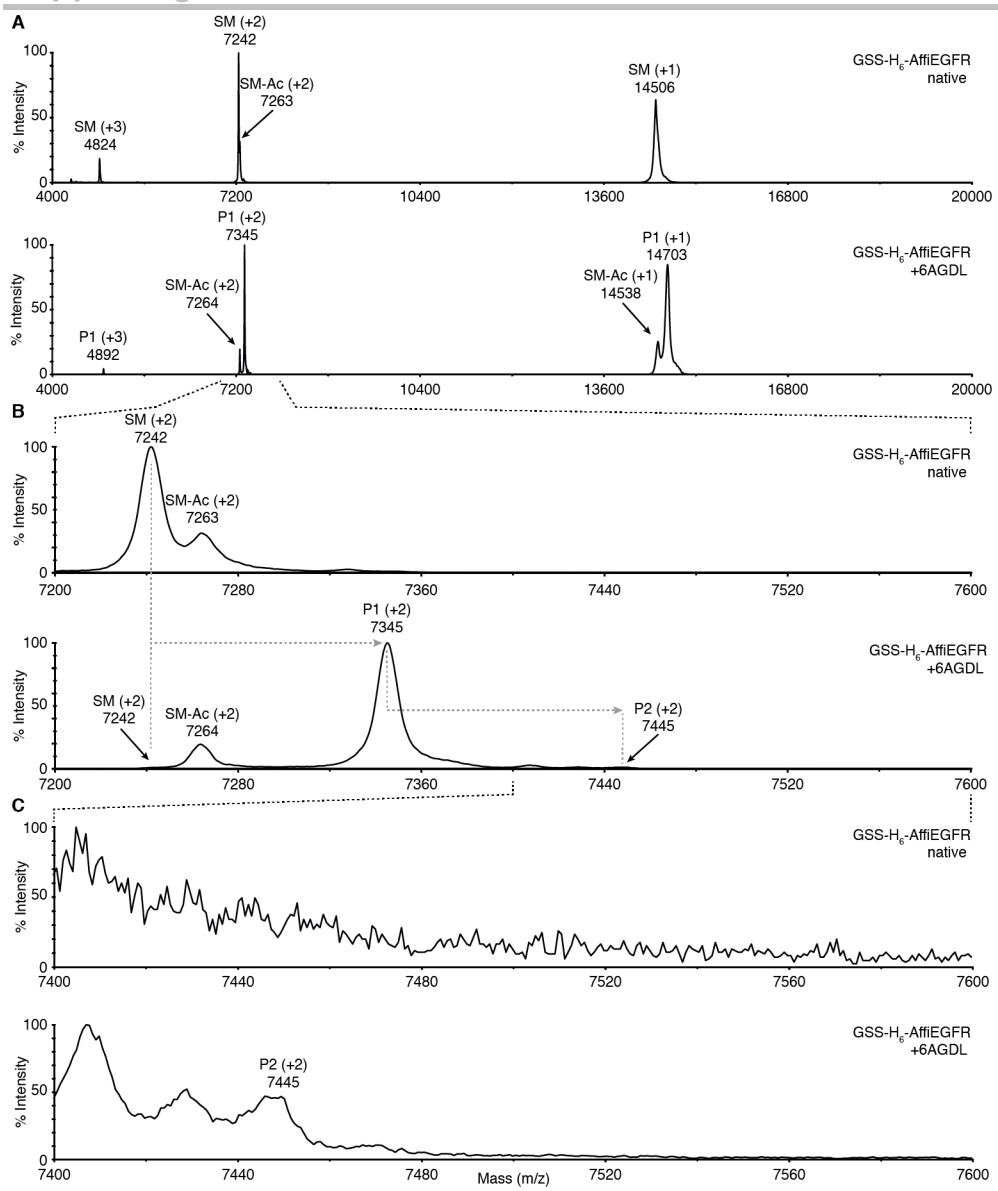


Figure S22. **Labelling of GSS-H_6-AffiEGFR using 6AGDL.** 20 μM GSS-H_6-AffiEGFR protein was labelled with 100 mM 6AGDL solution at pH 7.5 HEPES (NaOH), 1h, RT conditions. The protein prior and after labelling was analysed using linear, positive mode MALDI-TOF mass spectrometry with **(A)** showing the *m/z* 4000-20000 range with singly (+1), doubly (+2) and triply (+3) charged species observed **(B)** zoom in on the *m/z* 7200-7600 region showing the most abundant doubly charged species and **(C)** zoom in on the *m/z* 7400-7600 region showing minor doubly labelled, doubly charged species with *m/z* 7445. Number in the brackets (+x) indicates the charge state of the species observed. Theoretical *m/z* values for the observed peaks are: SM (+1)- 14429, SM(+2)-7215, SM(+3)-4810, SM-Ac(+1)-14471, SM-Ac(+2)-7236, P1(+1)-14632, P1(+2)-7317, P2(+2)-7418.

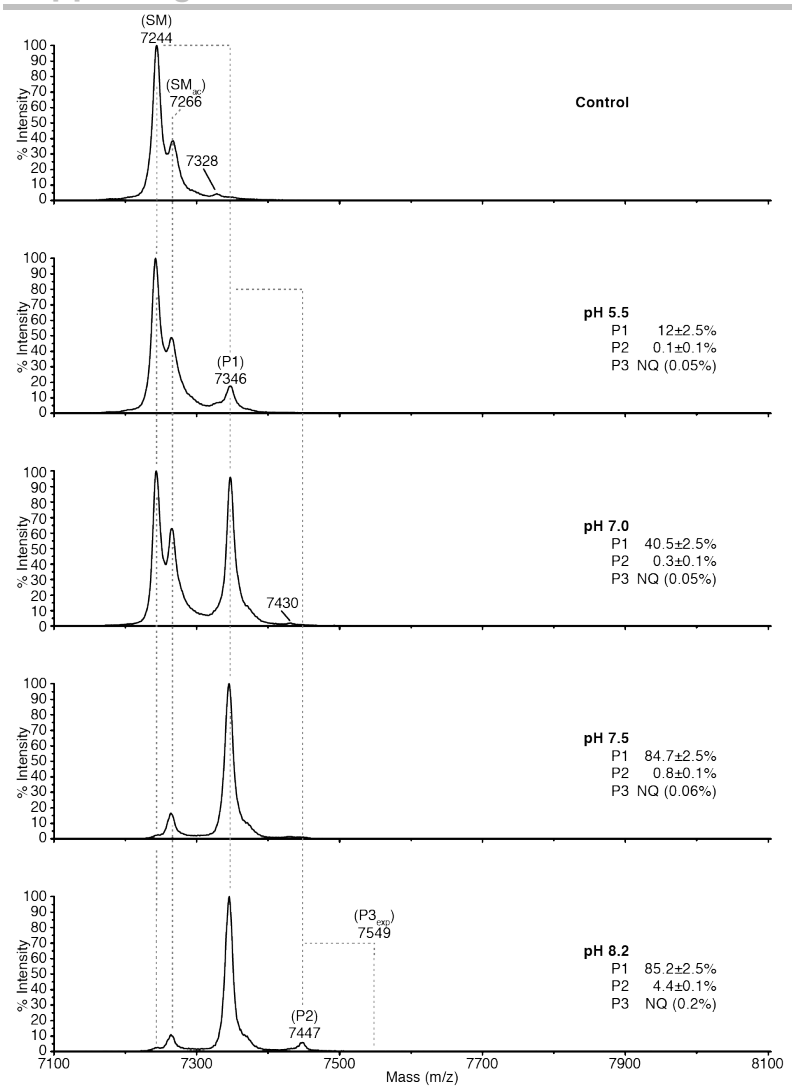


Figure S23. **Labelling of MALDI TOF MS of GSS-H_6-AffiEGFR.** 20 µM of GSS-$_{H6}$-AffiEGFR was reacted with 100 mM of 6-AGDL in either 0.5 M sodium acetate buffer at pH 5.5, 0.5 M sodium phosphate buffer at pH 7.0, 0.5 M HEPES (NaOH) buffer at pH 7.5, or 0.5 M HEPES (NaOH) buffer at pH 8.2 RT, for 1 hour respectively. The reaction was terminated by acidification by adding TFA to a final concentration of 0.8 M. A 10 µL aliquot was taken and desalted using ZipTip C18 resin, eluting in 10 µL of 70% (v/v) ACN, 0.1 % (v/v) TFA. Eluates were mixed 1:1 with CICCA matrix mixture and spotted onto a MALDI steel target prior to detection in Linear positive mode on a MALDI TOF MS. The second charge state is shown. The extent of labelling was determined by calculating the ratio between singly/doubly/triply labelled protein (accounting for acetylated and sodiated species) at the secondary charge. The expected mass for a triply activated protein was not observed. NQ – not reliably quantifiable, estimates provided only.

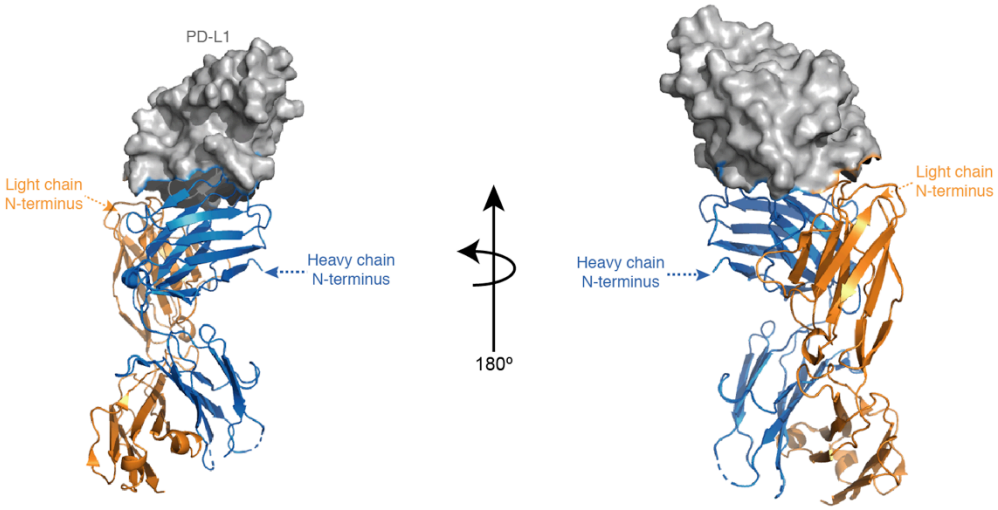


Figure S24. **Crystal structure of PD-L1 with atezolizumab fab.** The N-termini are indicated by arrows (PDB 5XXY).

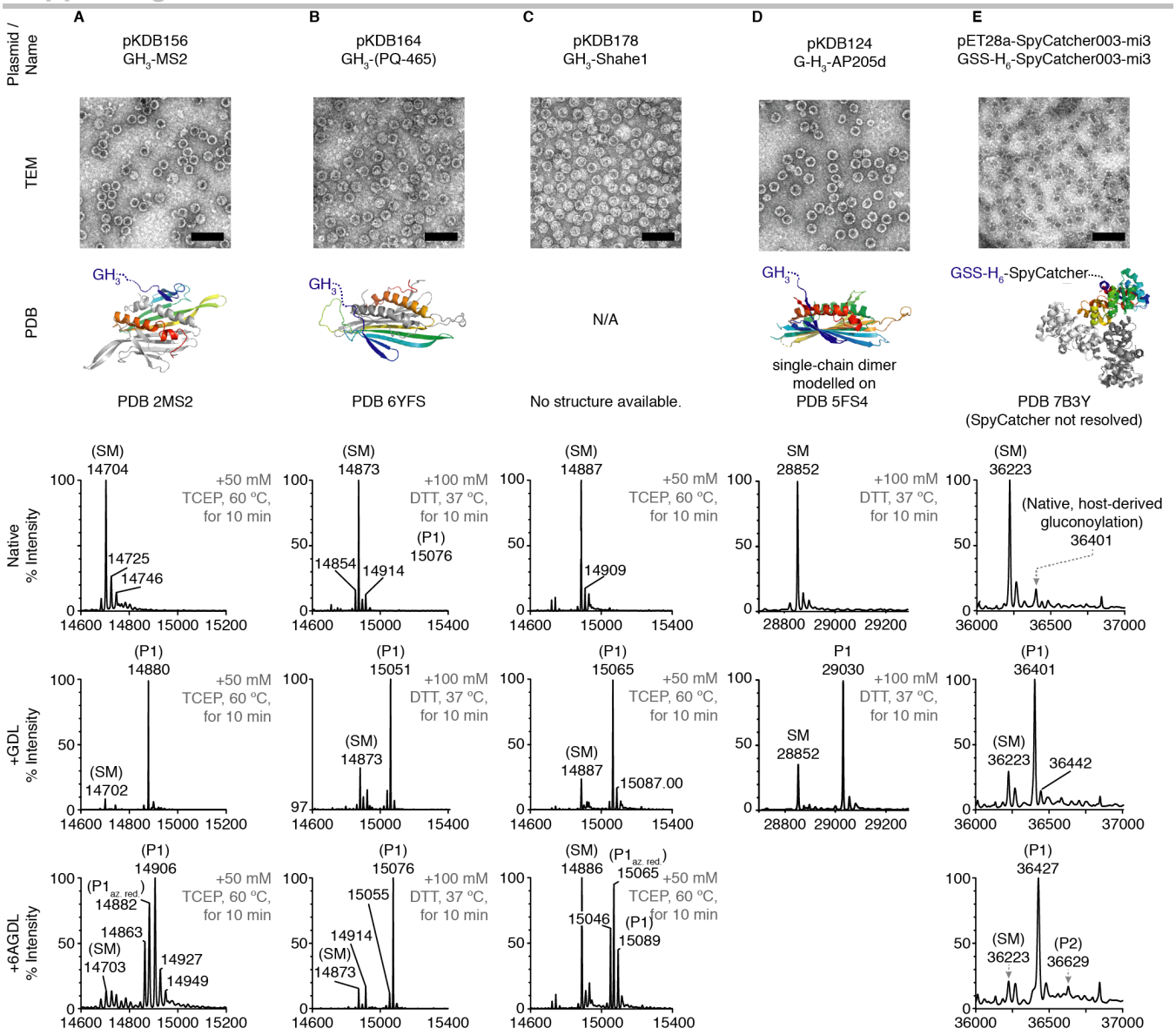


Figure S25. **(Azido-) gluconoylation of various Gly-(Ser$_2$)-His$_n$ virus-like and protein nanoparticles variants. (A)** GH$_3$-tagged HiPerX MS2 variant VLP, **(B)** GH$_3$-tagged PQ-465 VLP, **(C)** GH$_3$-tagged Shahe1, **(D)** GH$_3$-tagged single-chain dimer AP205 (only +GDL), and **(E)** GSSH$_6$-tagged SpyCatcher-mi03. Structures for the parent subunits are provided where available. Expression and protein purification via IMAC, followed by spin filtration over 100-300 kDa MWCO yielded assembled particles, as confirmed by TEM (scale bars 100 nm). The protein particles were labelled with 100 mM (6A-)GDL, for 1 h at RT, at 1-2 mg/mL (VLPs), and at 5-6 mg/mL (GSSH$_6$-SpyCatcher-mi03) in 200-500 mM HEPES/NaOH buffer pH 7.5. For VLPs, disulphide bonds are either reduced with TCEP, or DTT as indicated - for all VLP, including native and GDL-treated samples. We speculate that the lower than expected 6AGDL-adduct mass for (A), and (C) is due to the reduction of the azide group with TCEP (expected with Δ-25.89, annotated as P1$_{az.\ red.}$). It appears that DTT does not obviously reduce the label azido group of GH$_3$-(PQ-465) under the conditions tested here. GSSH$_6$-tagged SpyCatcher-mi03 showed some native, host-derived gluconoylation because it was expressed in standard *E. coli* BL21(DE3), which is *pgl* deficient.[20] See Figure S26 for a tryptic peptide of SpyCatcher-mi03, showing N-terminal modification.

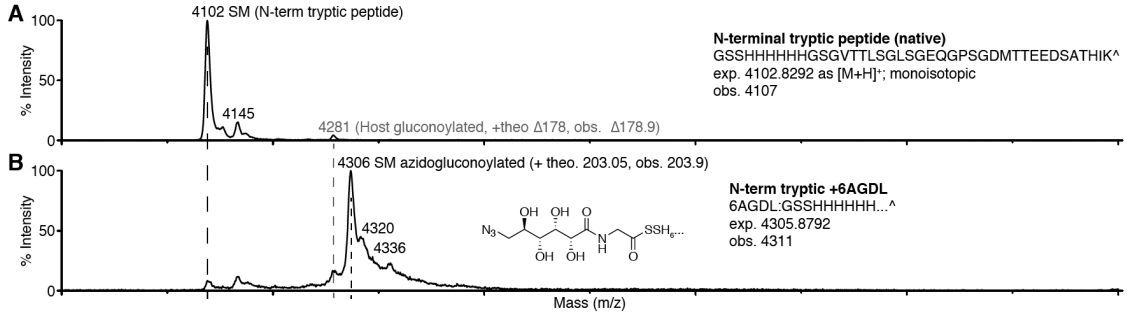


Figure S26. N-terminal tryptic peptide of **SpyCatcher-mi3. (A)** Native N-terminal SpyCatcher-mi03 tryptic peptide. **(B)** SpyCatcher-mi03 was **l**abelled with 100 mM 6AGDL at pH 7.5, spin filtered in 50 mM borate pH 8.2, and trypsinized for 1 h at 37 ºC. The reaction was acidified and C18 ZipTip purified. Spectra were acquired in linear mode with MALDI MS. Some native host-derived gluconoylation is observed at 4281, since the expression host was *pgl* negative. The recalitricant modification at 4145 is likely N-terminally acetylated material.

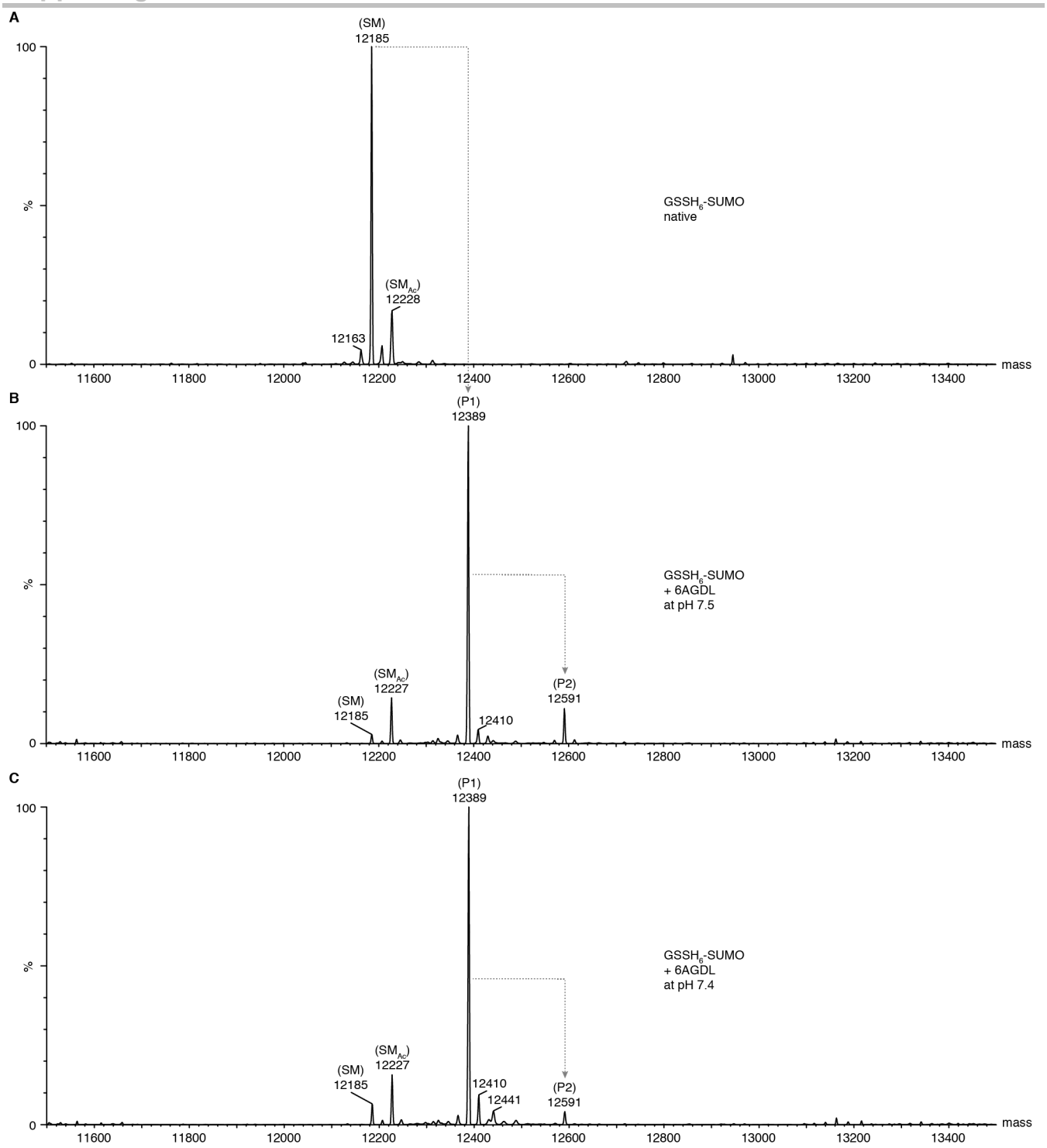


Figure S27. **Labelling with 6AGDL. (A)** Native GSS-H_6-SUMO. **(B)** Labelling with 100 mM 6AGDL at pH 7.5, and **(C)** at pH 7.4 for 1 h. Spectra were acquired via LC-MS/Q-TOF.

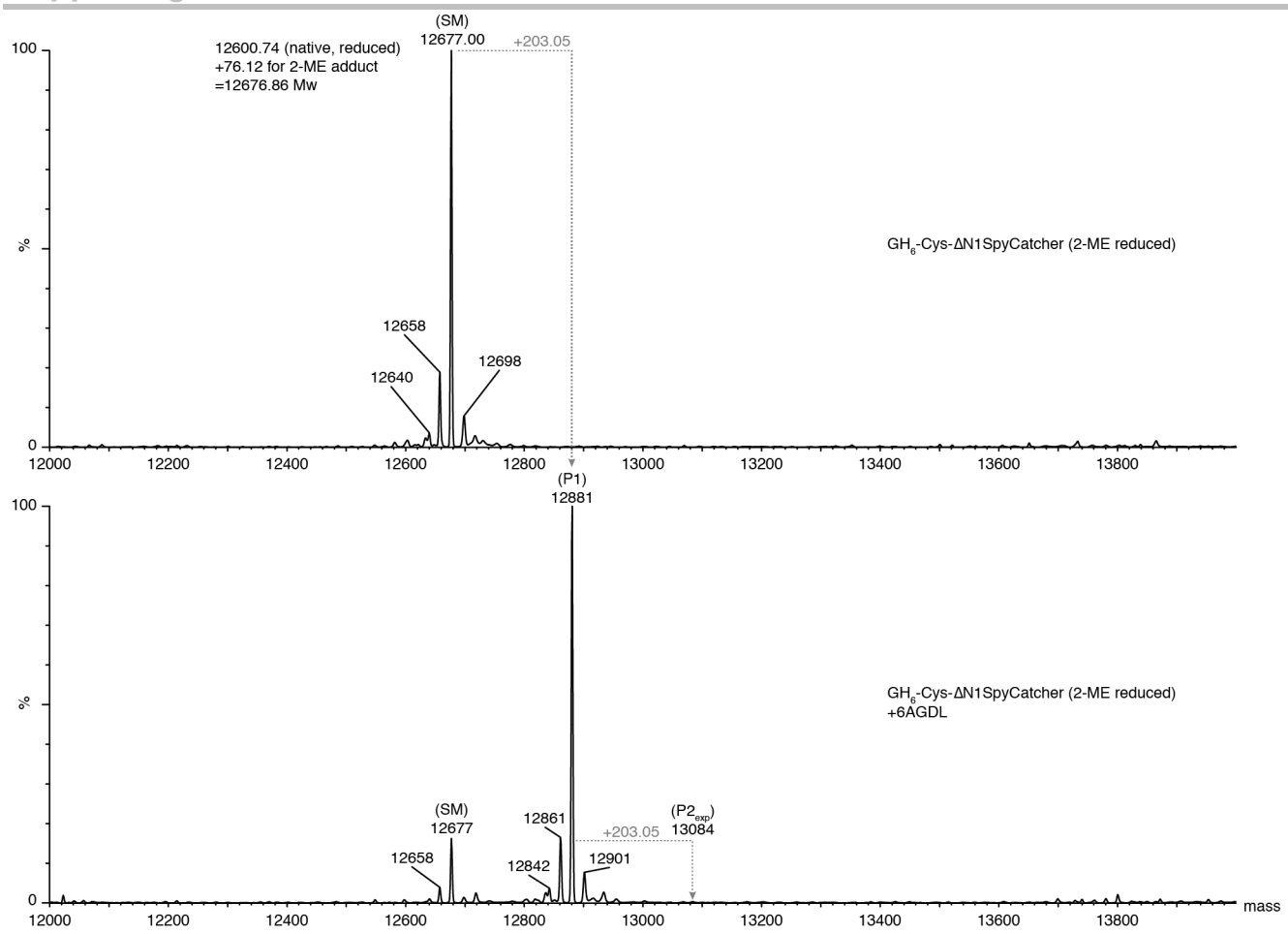


Figure S28. **GH_6-Cys-SpyCatcher (monomer).** Top: Native. Bottom: Labelling with 100 mM 6AGDL at pH 7.5 for 1 h. Spectra were acquired via LC-MS/Q-TOF.

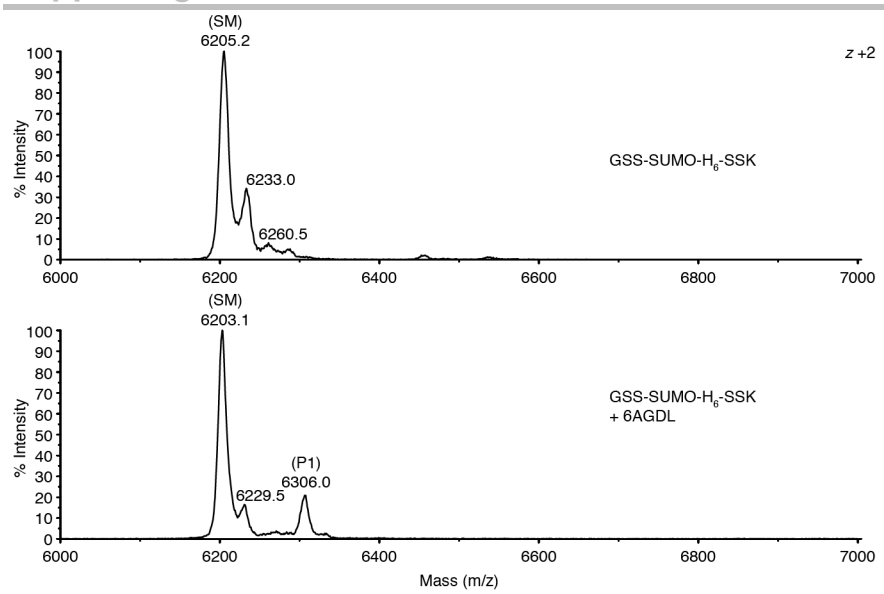


Figure S29. **GSS-SUMO-H6-SSK.** Top: Native. Bottom: labelling with 100 mM 6AGDL at pH 7.5, for 1 h. Spectra were acquired via MALDI TOF MS.

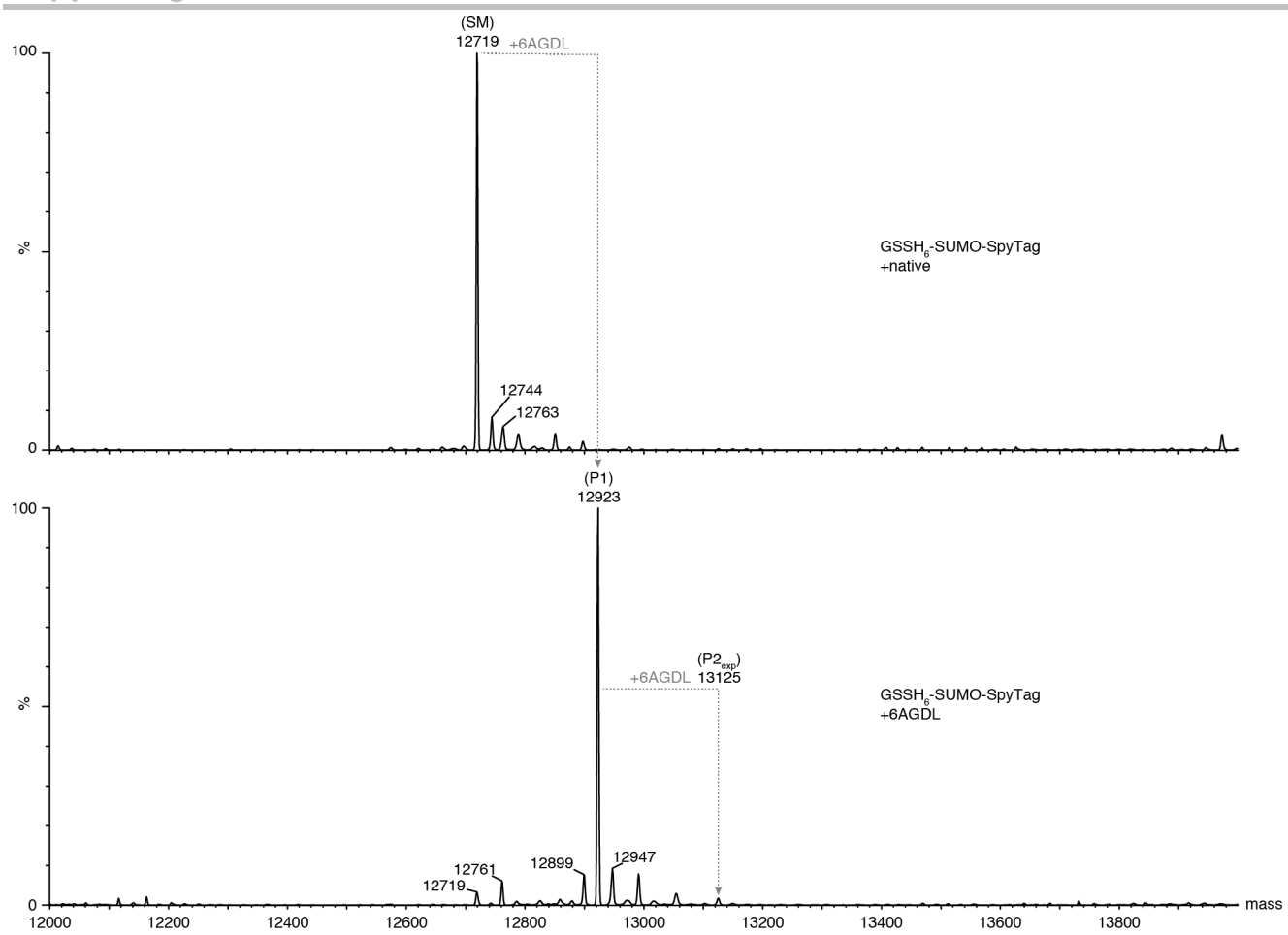


Figure S30. **$GSSH_6$-SUMO-SpyTag** labelling with 100 mM 6AGDL at pH 7.5 for 1 h at room temperature. Spectra were acquired via LC-MS/Q-TOF.

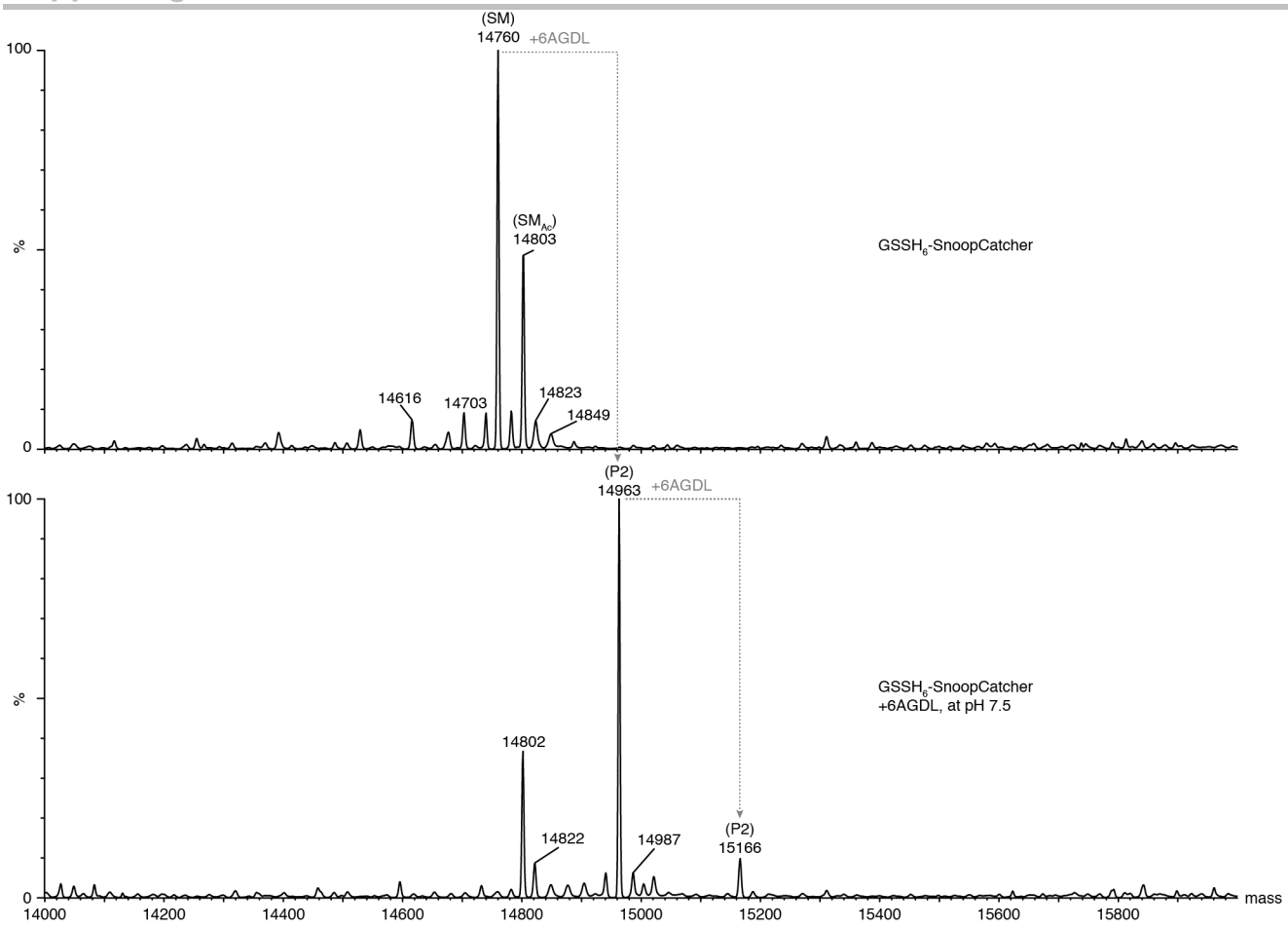


Figure S31. **GSS-H$_6$-SnoopCatcher.** Top: native. Bottom: Labelling with 100 mM 6AGDL for 1h, at RT, pH 7.5. Spectra were acquired via LC-MS/Q-TOF.

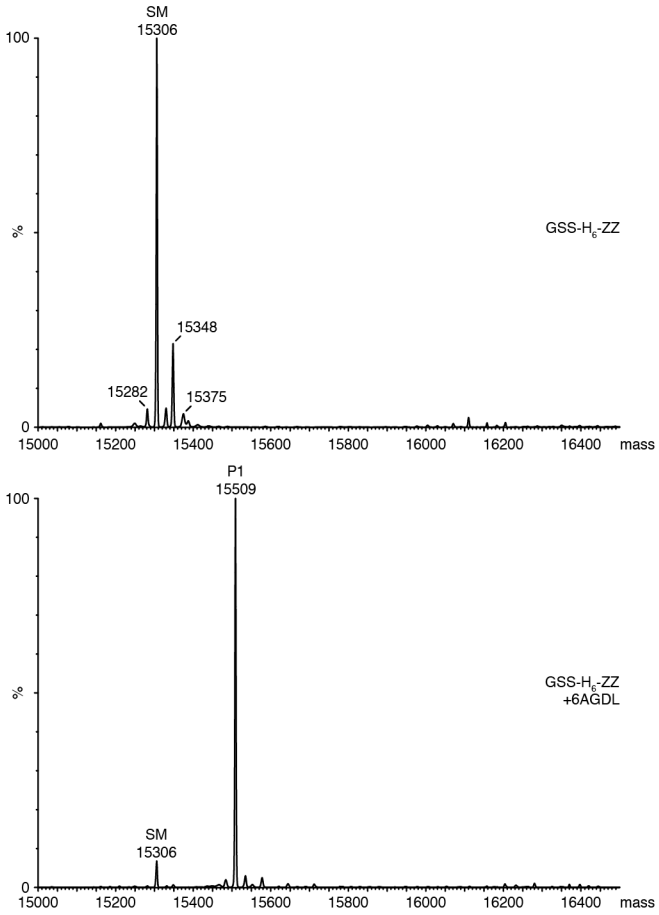


Figure S32. **GSS-H_6-ZZ.** Top: native. Bottom: Labelling with 100 mM 6AGDL for 1h, at RT, pH 7.5. Spectra were acquired via LC-MS/Q-TOF.

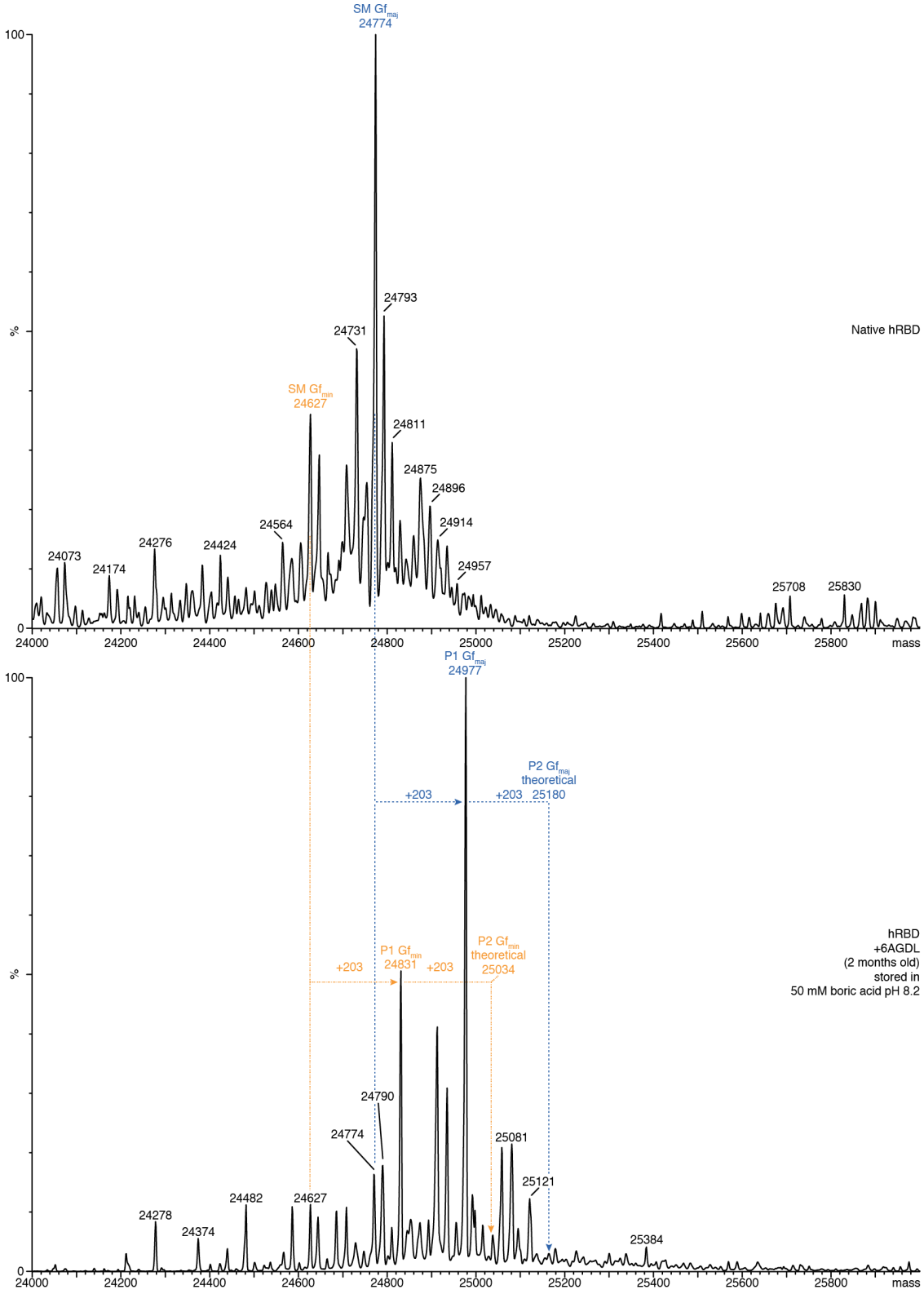


Figure S33. **Labelling of HEK-expressed glycoprotein.** (Top) G-H_6-hRBD, (bottom) G-H_6-hRBD activated with 6AGDL, dialysed into 50 mM boric acid pH 8.2, stored for 2 months at 4ºC in solution

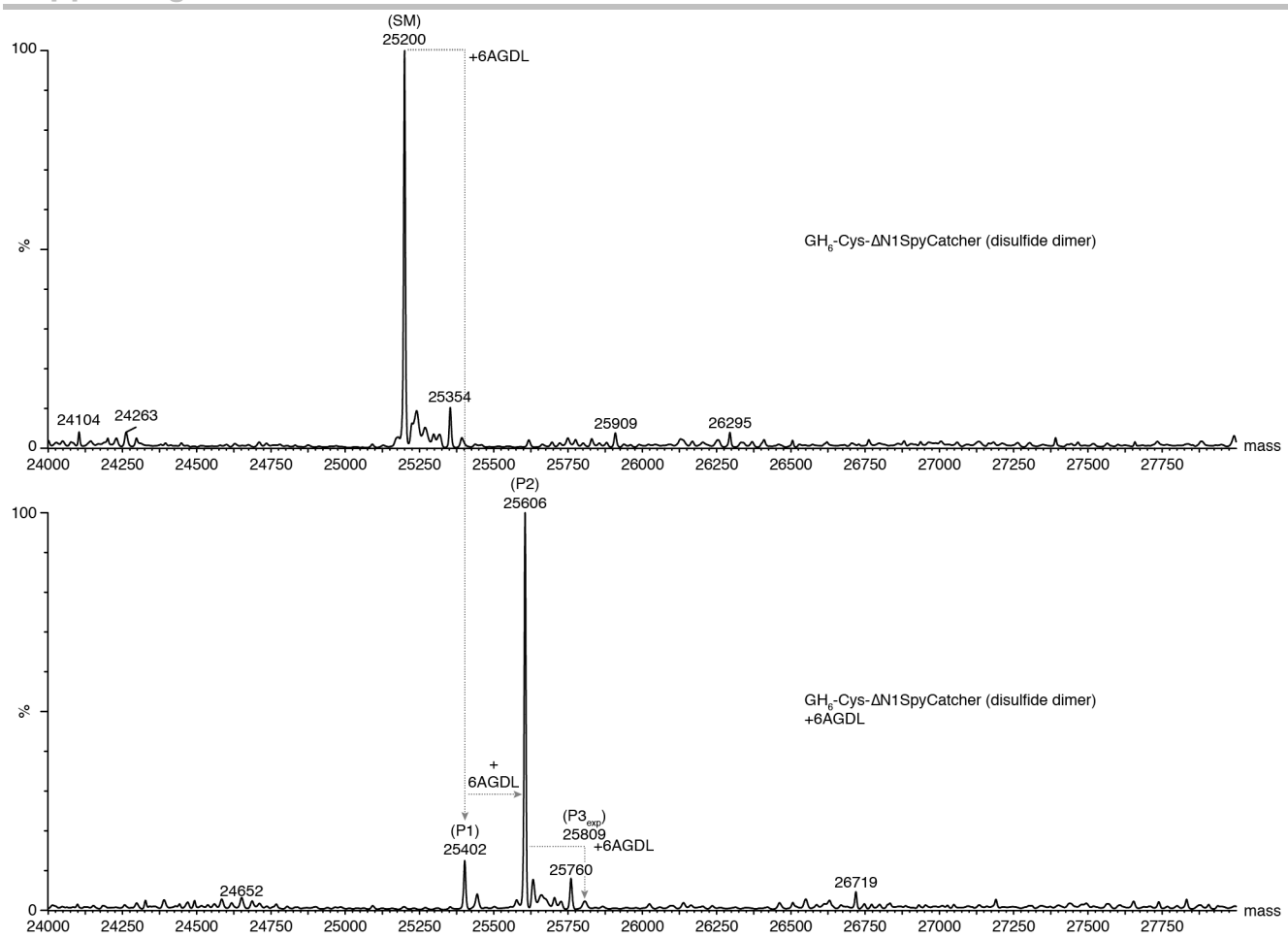


Figure S34. **Cys-SpyCatcher (dimer).** Top: native. Bottom: Labelling with 100 mM 6AGDL for 1h, at RT, pH 7.5. Spectra were acquired via LC-MS/Q-TOF.

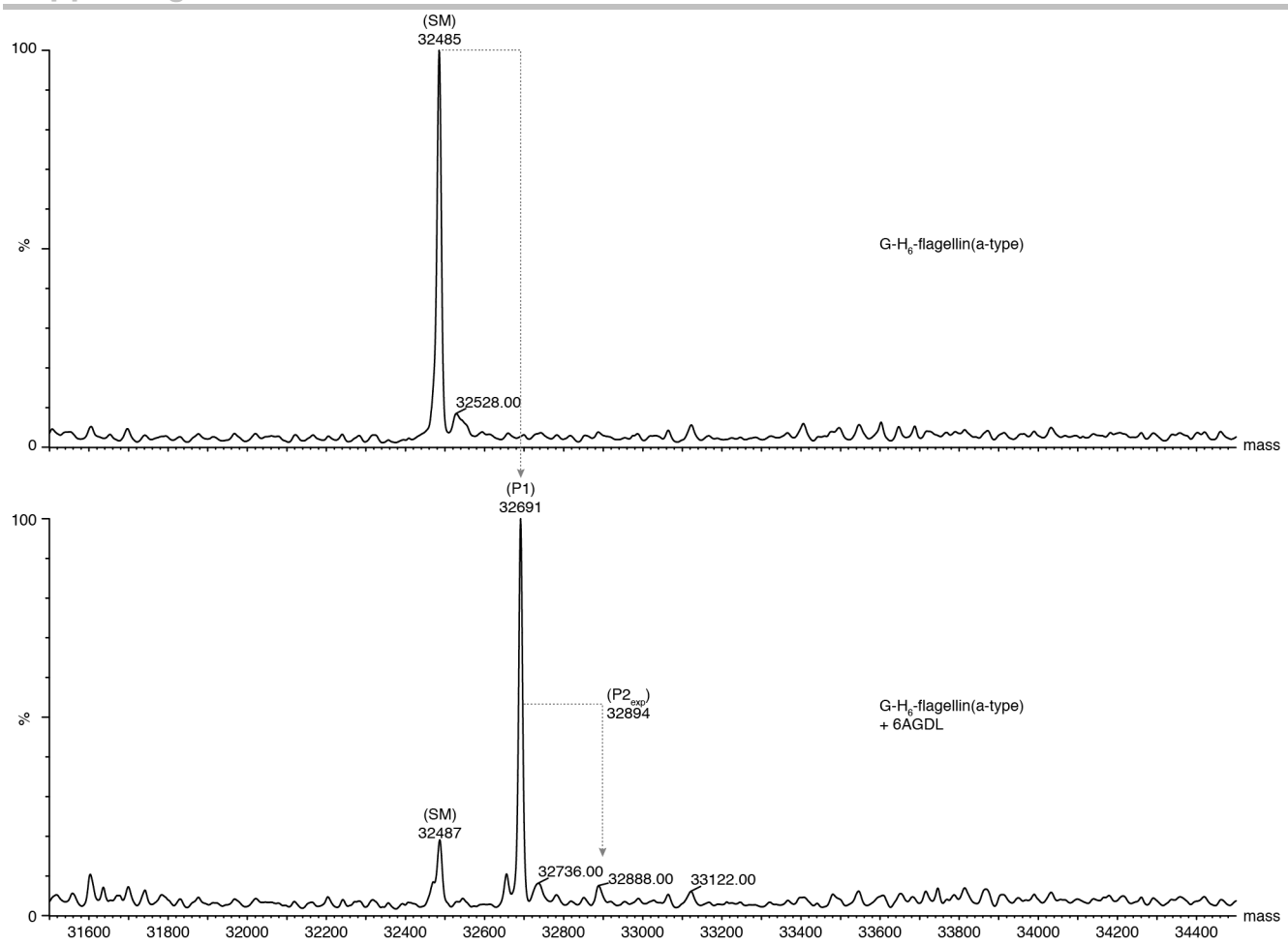


Figure S35. **Labelling of GH_6-flagellin** (a-type, pKB_259). Top: native. Bottom: Labelling with 100 mM 6AGDL for 1h, at RT, pH 7.5. Spectra were acquired via LC-MS/Q-TOF.

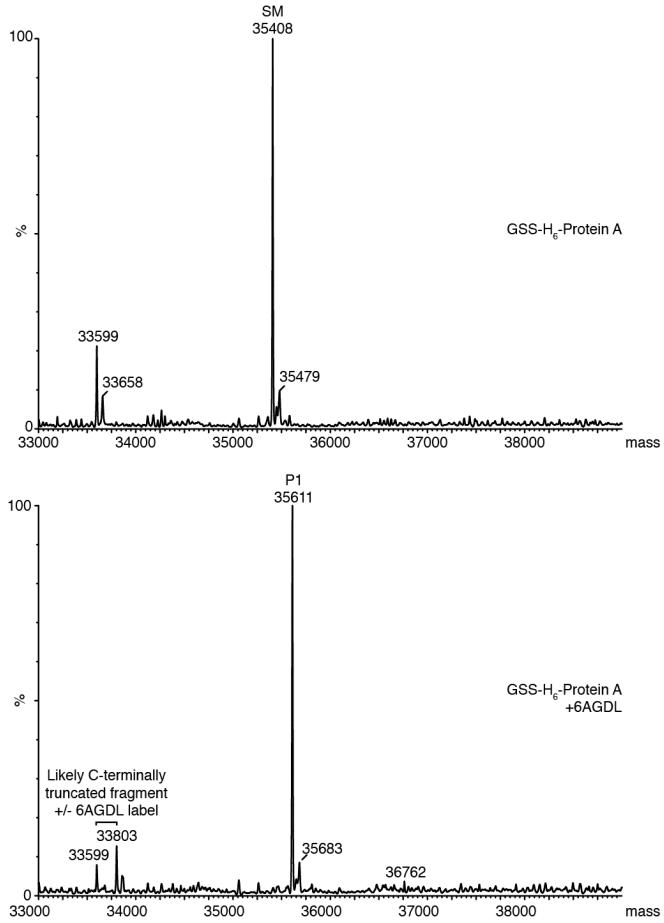


Figure S36. **GSS-H_6-Protein A.** Top: native. Bottom: Labelling with 100 mM 6AGDL for 1h, at RT, pH 7.5. Spectra were acquired via LC-MS/Q-TOF.

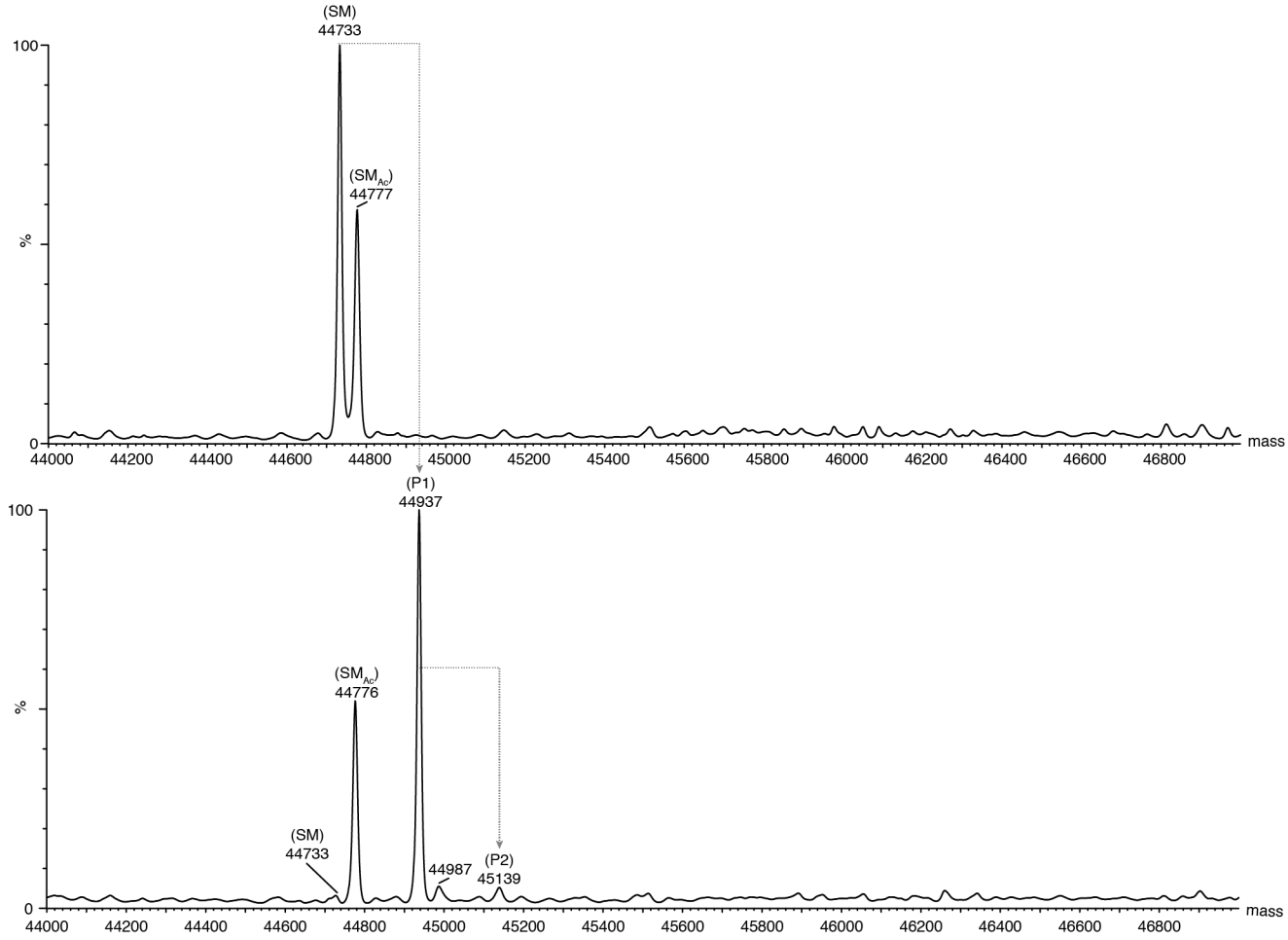


Figure S37. **GSS-H_6-SpyTag-MBP. Top: native.** Bottom: Labelling with 100 mM 6AGDL for 1h, at RT, pH 7.5. Spectra were acquired via LC-MS/Q-TOF.

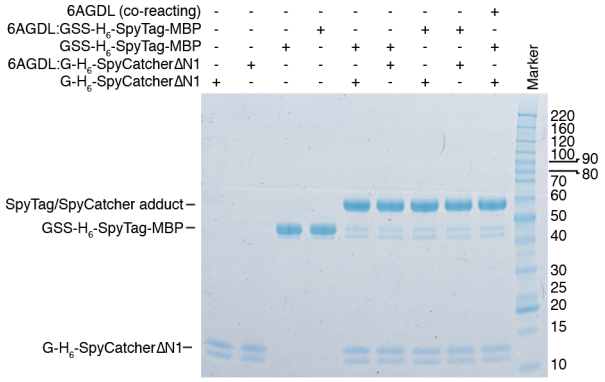


Figure S38. **Azidogluconoylation is compatible with Catcher/Tag chemistry.** G-H_6-SpyCatcherΔN1 and/or GSS-H_6-SpyTag-MBP protein was reacted for 1 h with 100 mM 6AGDL at room temperature in 200 mM HEPES pH 7.5 in order to azidogluconoylate the respective Catcher/Tag protein. After the reaction, azidogluconoylated Catcher and/or Tag protein was then directly mixed with the cognate partner and reacted for 1 h (i.e. in the presence of 6-azido-6-deoxy-gluconic acid). GH_6-SpyCatcherΔN1 and/or GH_6-SpyTag-MBP were also simultaneously co-reacted in the presence of 6AGDL for a total of 2h without prior azidogluconoylation (i.e. in the presence of the lactone). Reactions were stopped with SDS-PAGE loading buffer and analysed by SDS-PAGE (NuPAGE, Bis-Tris gel, 4-12% gradient, MES buffer), with Coomassie Blue staining. Expected Mw G-H_6-SpyCatcherΔN1 12.5 kDa, observed by SDS-PAGE ~11-13 kDa. Note, first generation SpyCatcher is known to run as a doublet in SDS-PAGE.[43] Expected Mw GSS-H_6-SpyTag-MBP 44.7 kDa; observed by SDS-PAGE ~45 kDa. The adduct of Catcher/Tag would be expected at 57.2 kDa without 6AGDL-modifications, observed at >56 kDa. Single azidogluconoylated SpyCatcher/SpyTag protein would be expected to be ~203.05 Da bigger, a difference which is too small to resolve by SDS-PAGE for proteins of the size of SpyTag-MBP (see Figure S37 LC/MS for azidogluconoylation of GSS-H_6-SpyTag-MBP).

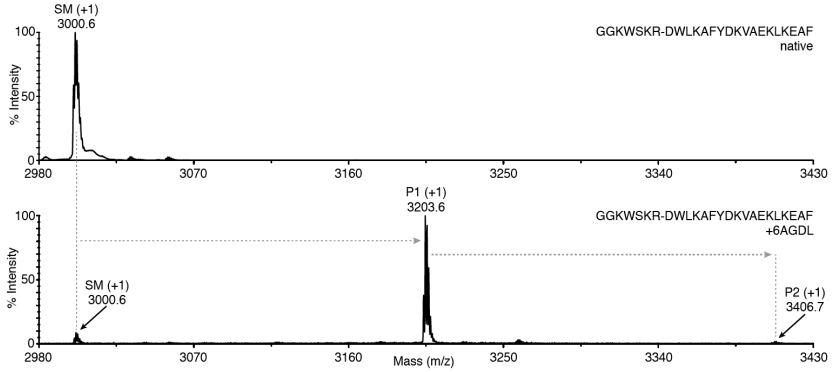


Figure S39. **Double labelling.** MALDI-TOF MS spectra showing the extent of labelling of GGKWSKR-Beltide-1 peptide (PP7) after labelling twice with 100 mM 6AGDL in 1 M HEPES (NaOH) pH 7.5 for 3 h each time at 4 ºC. SM(+1): PP7 starting material, singly charged, protonated; P1(+1): PP7 product labelled with one 6AGDL tag, singly charged, protonated; P2(+1): PP7 product labelled with two 6AGDL tags, singly charged, protonated.

Note: Figure S39. Single labelling GGKWSKR-Beltide-1 peptide (PP7) with 100 mM 6AGDL resulted in ~75 % labelling (data not shown). After repeatedly labelling the peptide once more ~91% mono-functionalized and ~2% bi-functionalized product could be obtained based on Ion Count (XIC) ratios of MALDI-TOF MS data.

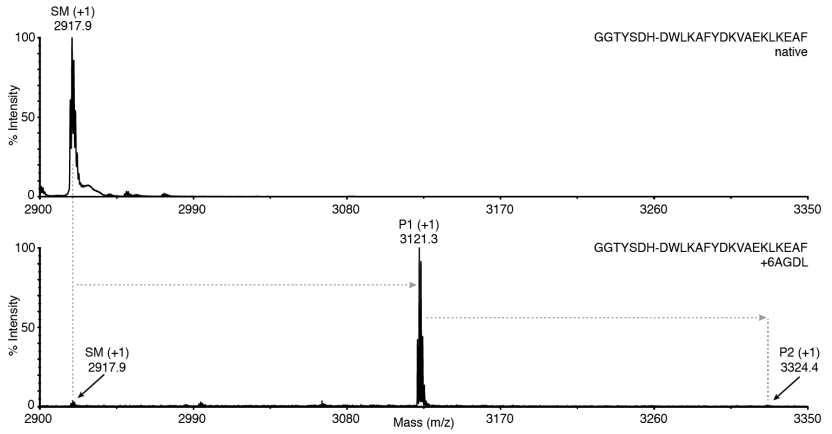


Figure S40. **Double labelling.** MALDI-TOF MS spectra showing the extent of labelling of GGTYSDH-Beltide-1 peptide (PP10) after repeated (twice) labelling with 100 mM 6AGDL in 1 M HEPES pH 7.5 for 3h and another 3h labelling after addition of secondary charge of 100 mM 6AGDL at 4 ºC. SM(+1): PP10 starting material, singly charged, protonated; P1(+1): PP10 product labelled with one 6AGDL tag, singly charged, protonated; P2(+1): PP10 product labelled with two 6AGDL tags, singly charged, protonated.

Note: Figure S40. Single labelling of GGTYSDH-Beltide1 (PP10) with 100 mM 6AGDL yielded ~67% mono-functionalized (data not shown). Labelling the singly-treated peptide once more with 100 mM 6AGDL for 3 h resulted in ~95% mono-functionalized and ~1% bi-functionalized products.

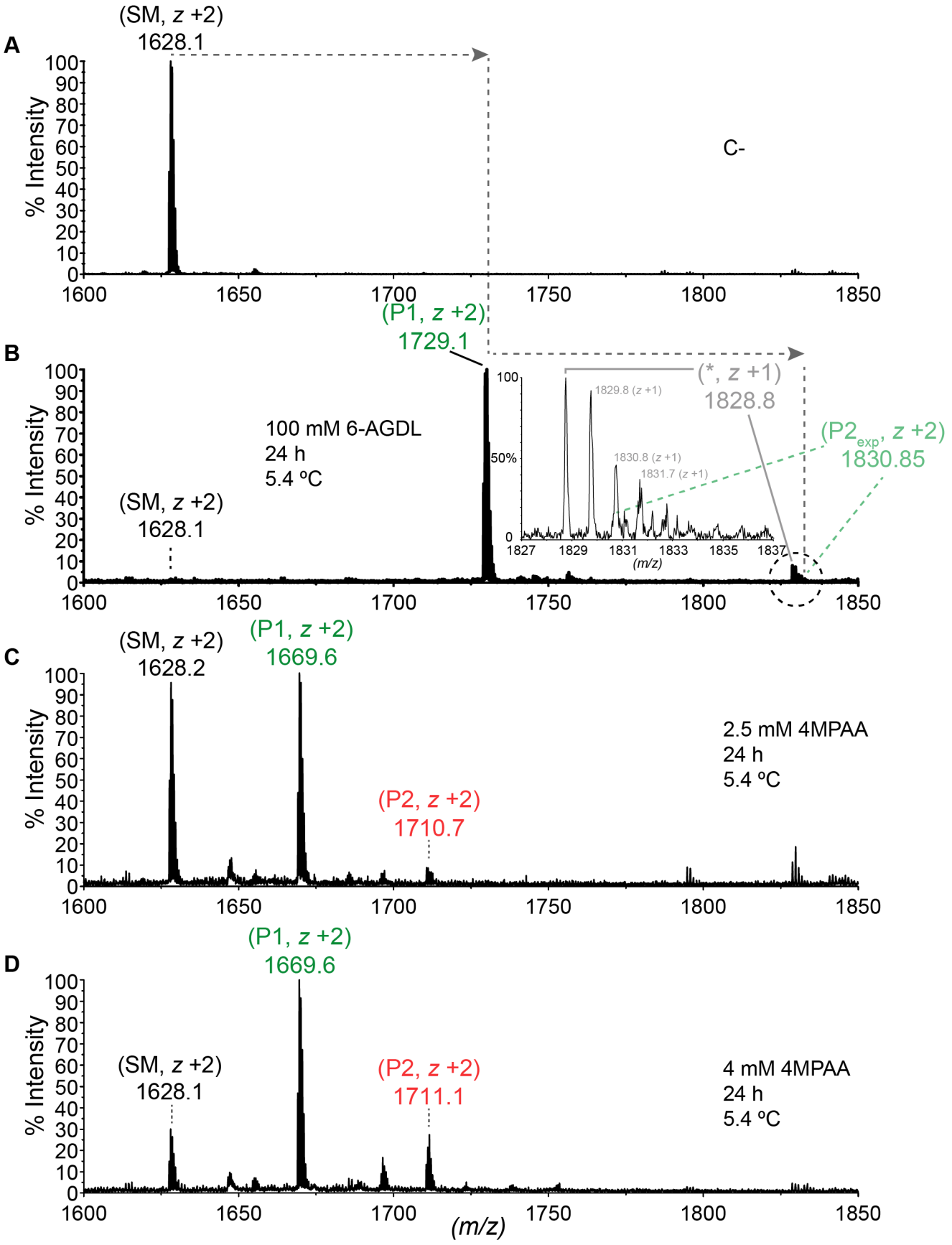


Figure S41. **Representative comparison between 4MPAA and 6AGDL labelled PP6.** 1 mM PP6 peptide was incubated with either **(A)** no label, **(B)** 100 mM 6AGDL, **(C)** 2.5 mM 4-MPAA, or **(D)** 4 mM 4-MPAA for 24 h at 5.4 ºC in 200 mM HEPES (NaOH), pH 7.5. Reactions were analysed by MALDI-TOF MS at charge number z +2. The inset for (B) shows a zoom in of the circled mass range. In this range, species singly-charged species at 1828.8(*) is observed. This * species is not a double-labelled peptide, which is doubly charged and expected at 1831.15. SM (1628.1+101.525+101.525), starting material; P1, mono-labelled species; P2, bi-labelled species. For the 6AGDL reaction, the position for which a dually-labelled species would be expected is indicated.

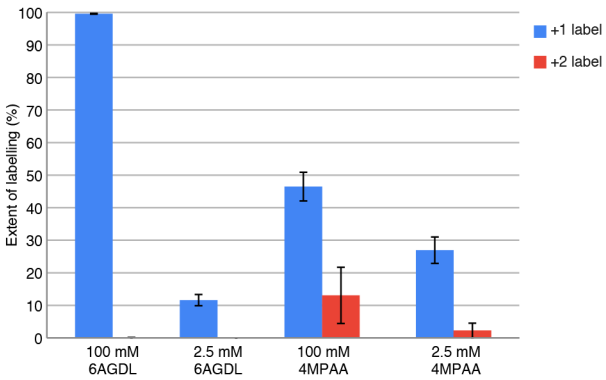


Figure S42. **Quantitative comparison between 4MPAA and 6AGDL labelled PP13.** 250 µM PP13 peptide was reacted with either 100 mM, or 2.5 mM 6AGDL, or 4MPAA at 4-5.4 ºC in 200 mM HEPES (NaOH), pH 7.5, for 2 h. Reactions were analysed by MALDI-TOF MS at charge number *z +1*, and a representative result is shown in Figure S41. Error bars show the standard deviation from the mean (n=3).

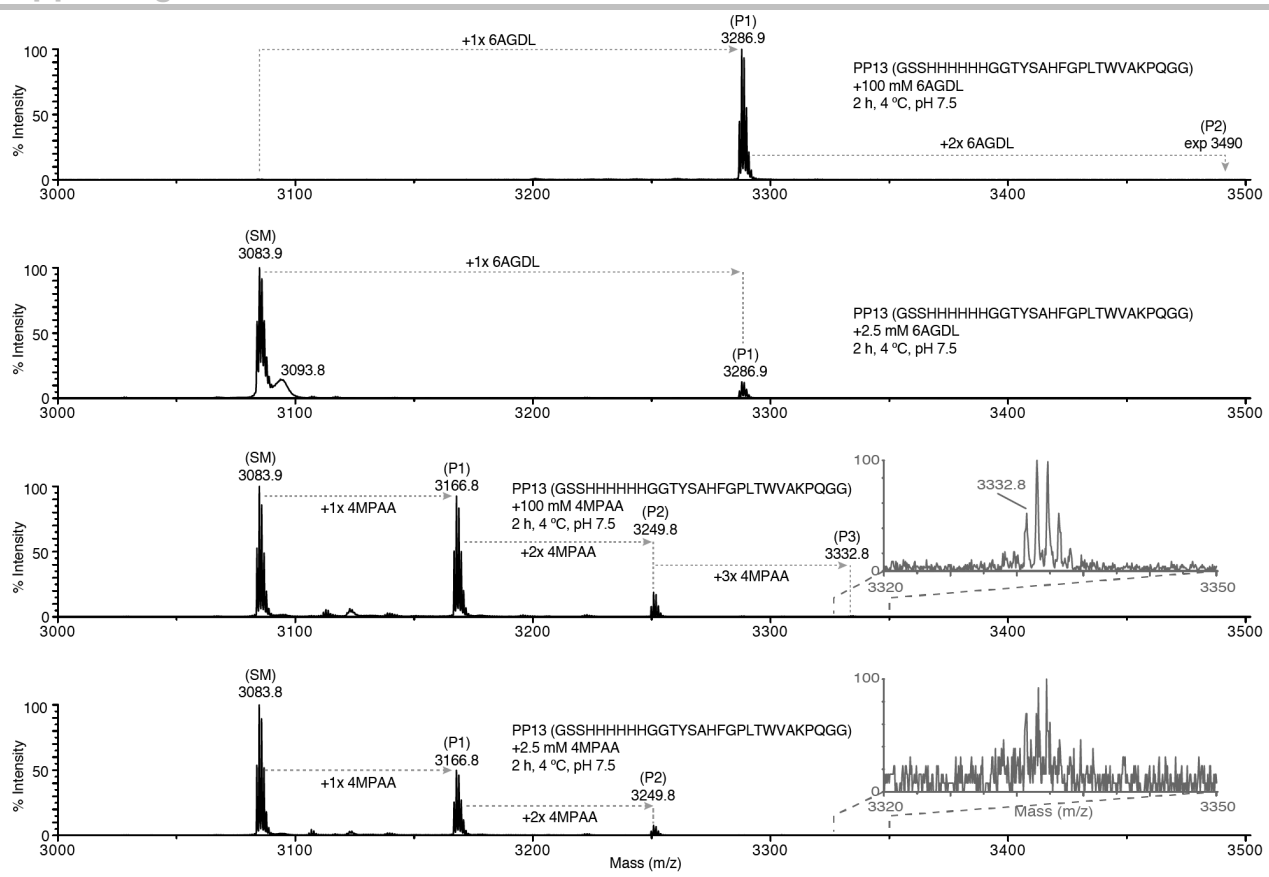


Figure S43. **Representative comparison between 4MPAA and 6AGDL.** 250 µM PP13 peptide was reacted with either 100 mM, or 2.5 mM of 6AGDL, or 4MPAA at ~4 ºC in 200 mM HEPES (NaOH), pH 7.5, for 2 h. Reactions were analysed by MALDI-TOF MS at charge number *z +1*. The insets for the 4MPAA reactions show a zoom in of the indicated mass range. Strikingly, we detect triple functionalization, likely on a Ser, Thr, or His residue at m/z/ 3322.8 (theoretical expected for a triple adduct is 3332.6), as the particular substrate has only two amine functional groups (N^{α}- and Lys25).

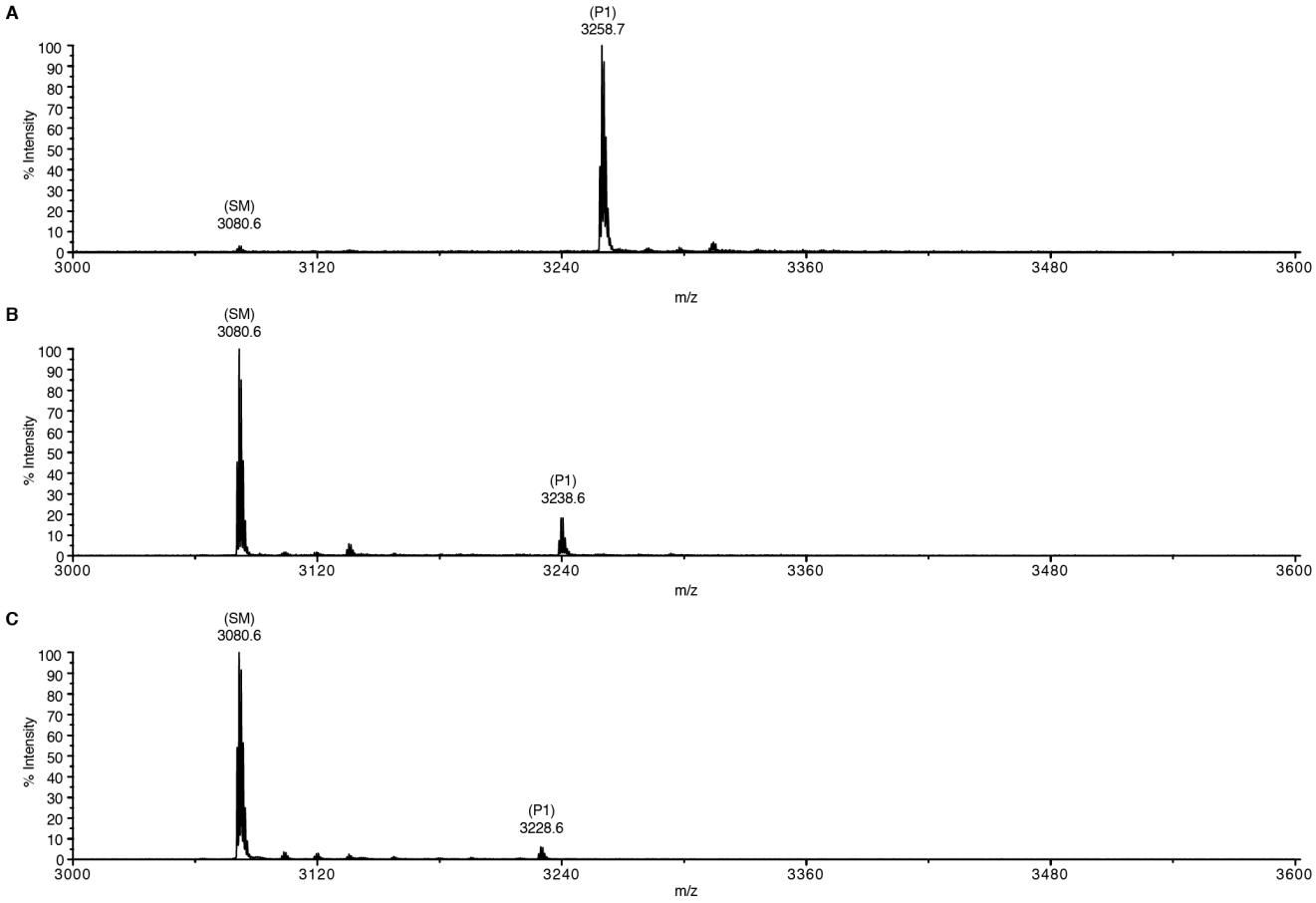


Figure S44. **Reaction of PP1 peptide with various lactones at 100 mM in 200 mM HEPES (NaOH) at pH 7.5 for 1 h. (A)** GDL, **(B)** D-glucorono-6,3-lactone (CAS 32449-92-6), and **(C)** D-ribono-1,4-lactone; followed by C18 ZipTipping with 0.1% TFA and MALDI MS.

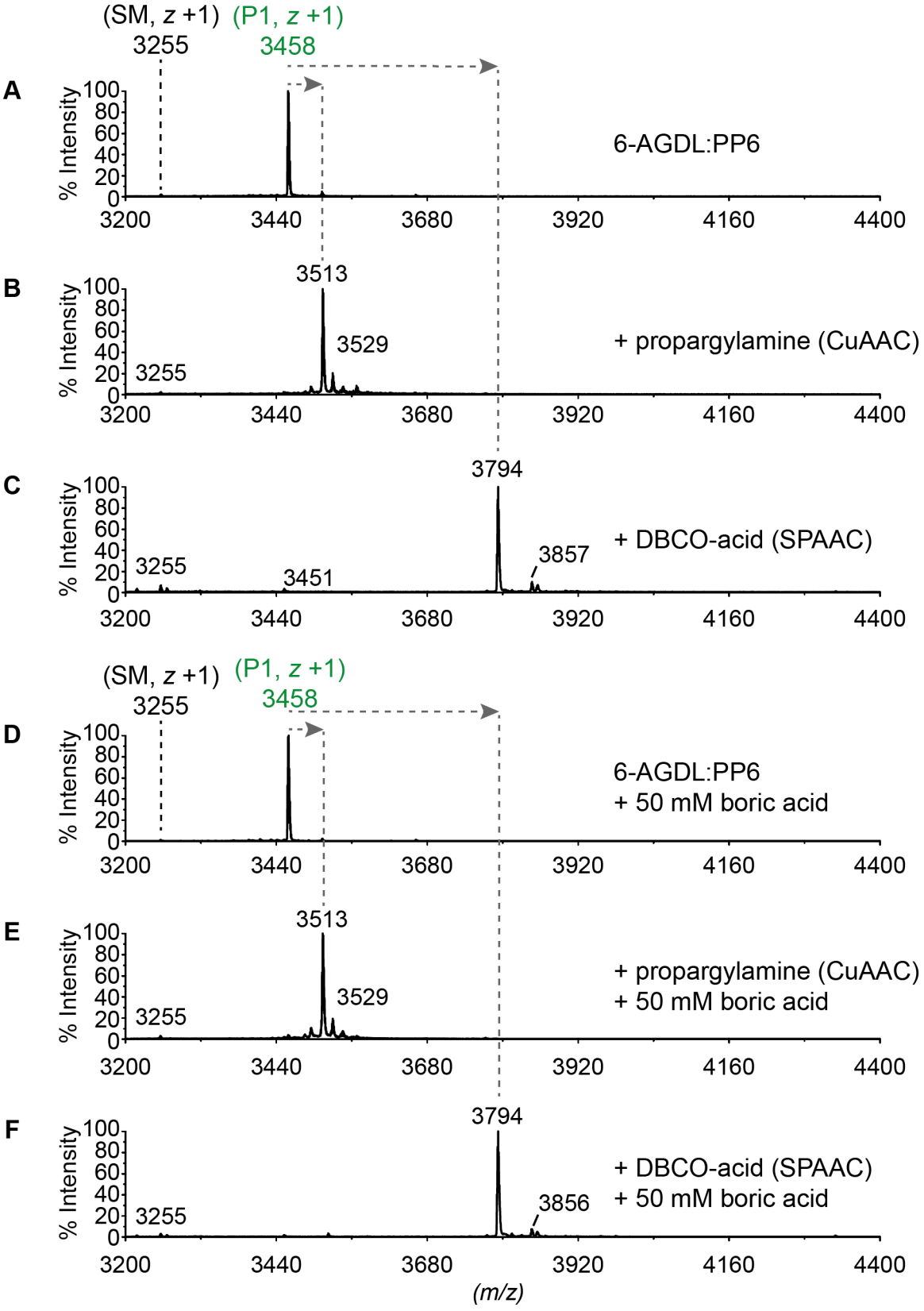


Figure S45. **Click-chemistry labelling of azidogluconoylated PP6 via CuAAC or SPAAC.** Reaction of 2-fold molar excess of propargylamine (CuAAC, via THPTA), or 5-fold molar excess of DBCO-acid (SPAAC) over the azidated peptide after 15, and 30 minutes at 37 ºC in 100 mM potassium phosphate buffer (pH 8.0), in the presence or absence of 50 mM boric acid.

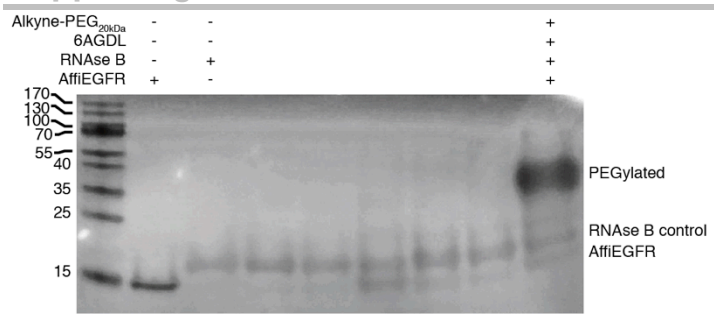


Figure S46. **PEGylation of 6AGDL-labelled GSS-H_6-AffiEGFR via CuAAC.** Azidated protein was contacted with excess 20kDa-PEG-alkyne in the presence of THPTA, $CuSO_4$, sodium ascorbate. RNASeB was included as an internal control.

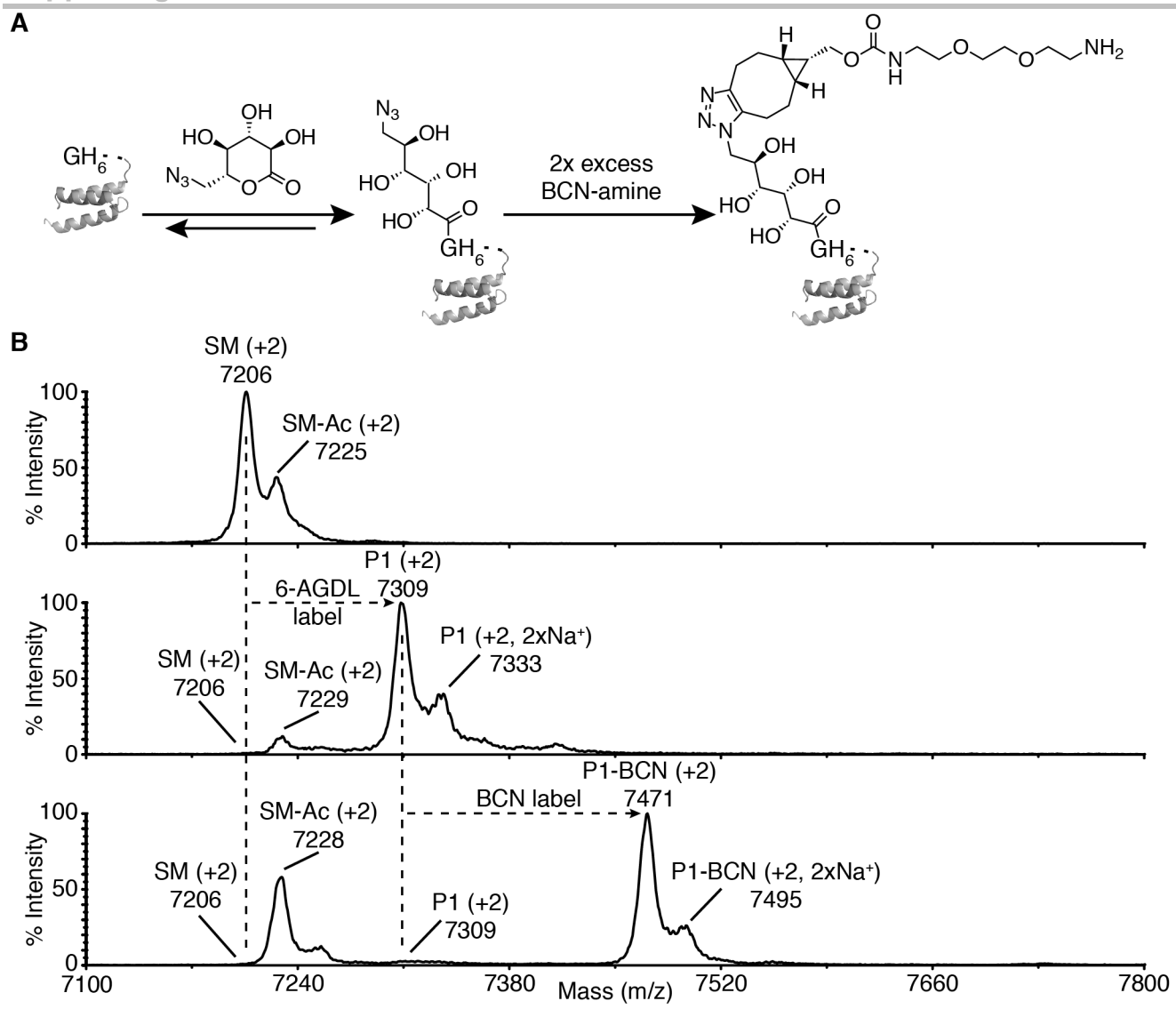


Figure S47. **Two-step mono-functionalization of a model protein with 6AGDL and metal-free click chemistry. (A)** Schematic representation of the two-step mono-functionalization of GSS-H_6-AffiEGFR. **(B)** MALDI-TOF MS spectra showing labelling of GSS-H_6-AffiEGFR with 6AGDL, and then followed by BCN-amine. Top panel shows GSS-H_6-AffiEGFR prior 6AGDL labelling, the middle panel shows mono-functionalised GSS-H_6-AffiEGFR after incubation with 6AGDL, and the bottom panel shows the species after metal-free click labelling of 6AGDL:PP6 with BCN-amine. Spectra were acquired in positive linear mode. SM(+2): GSS-H_6-AffiEGFR starting material, doubly charged, protonated; SM-Ac(+2)-GSS-H_6-AffiEGFR acetylated starting material, doubly charged, protonated; P1(+2): GSS-H_6-AffiEGFR product labelled with one 6AGDL tag, doubly charged, protonated; P1(+2, 2xNa^+): GSS-H_6-AffiEGFR product labelled with one 6AGDL tag, doubly charged, sodiated; P1-BCN (+2): P1 labelled with BCN-amine, doubly charged, protonated; P1-BCN (+2, 2xNa^+): P1 labelled with BCN-amine, doubly charged, sodiated.

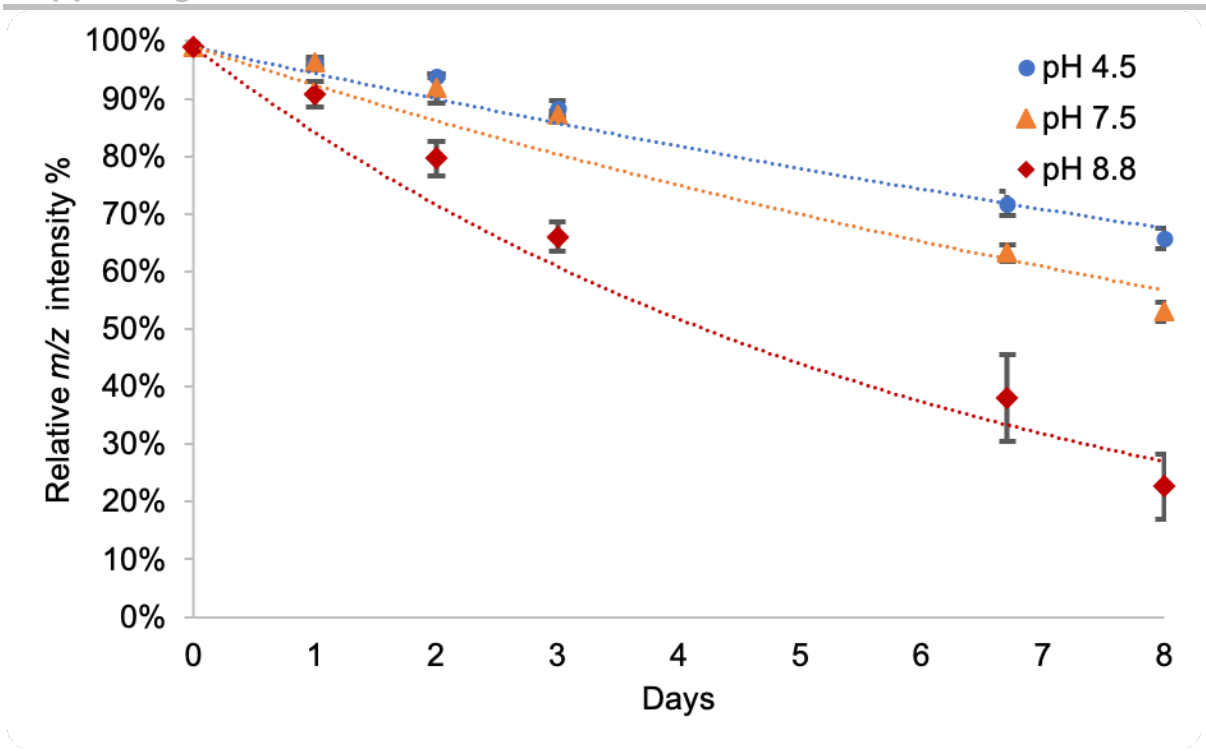


Figure S48. **Reversibility of the 6-azido-6-deoxy-gluconoyl modification as a function of pH.** Change in the relative *m/z* intensity ($[m_1/z/(m_1/z+m_2/z)]*100\%$, *z* +2) during storage in solution at room temperature at different pH values over time as measured by MALDI-TOF MS. The initial concentration of 6AGDL labelled GSS-H_6-AffiEGFR was ~20 µM. pH 4.5 [100 mM acetic acid (NH_3)], pH 7.5 [50 mM HEPES (NaOH)], and pH 8.8 [100 mM NH_3 (acetic acid)]. Curves are based on three measurements and exponential trendline fitting; error bars represent 1 standard deviation from the arithmetic mean.

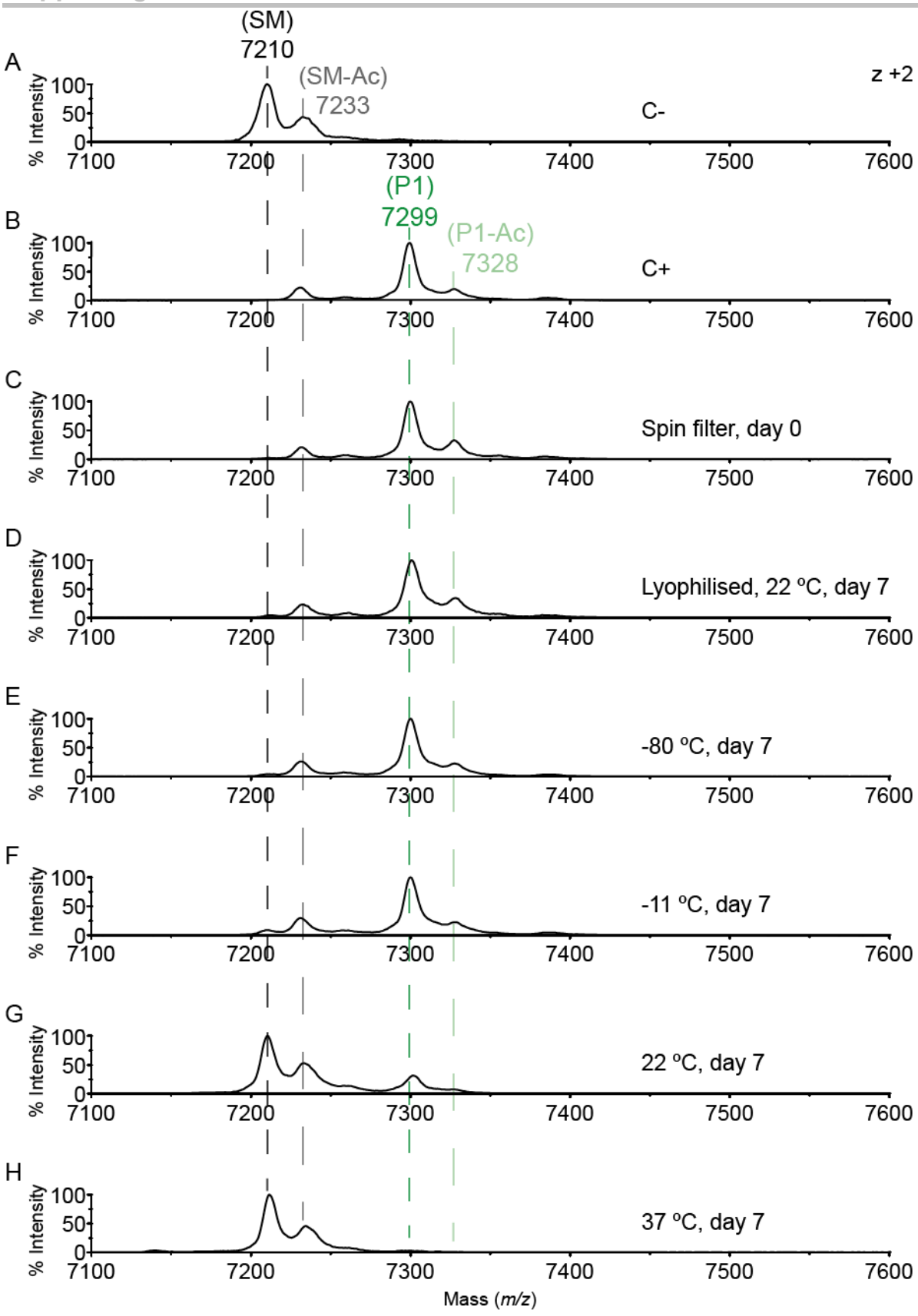


Figure S49. **Reversibility of gluconoylation as a function of physical state and temperature.** GSS-H_6-AffiEGFR was labelled with GDL and excess label was removed by spin-filtering into 50 mM potassium phosphate buffer (KOH), pH 7.5. **(A)** Control protein without label, **(B)** conjugate after labelling, **(C)** conjugate after spin-filtering (day 0). The conjugate was either (D) lyophilised and stored at room temperature, **(E)** frozen at -80 ºC, kept in solution at **(F)** -11ºC, **(G)** room temperature (22 ºC), or at **(H)** 37 ºC for 7 days. The initial labelled protein concentration was ~105 µM. The extent of labelling was determined by MALDI-TOF MS at the secondary charge number (z +2) for all conditions after C18 ZipTip purification.SM, starting material; P1, mono-labelled species; -Ac, acetylated

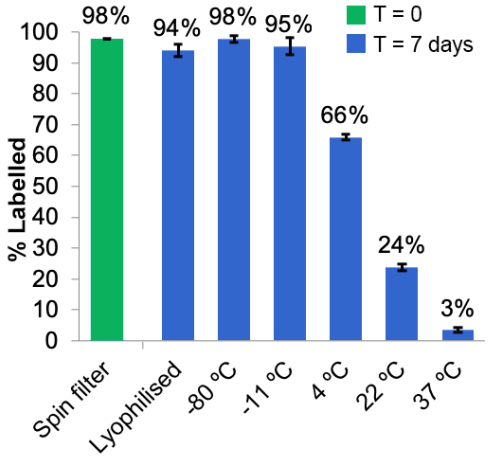


Figure S50. **Stability of GDL labelling after storage for 7 days.** GSS-H_6-AffiEGFR was labelled with GDL to near-completion in 1h at room temperature (22 ºC). Excess GDL was removed via repeated spin-filtering into 50 mM pH 7.5 potassium phosphate (KOH) buffer ("Spin filter", which served as the Control). The spin-filtered reaction was aliquoted into different tubes that were either lyophilised (lyophilisate stored at 22 ºC), stored in solution at different temperatures. -80 ºC (frozen), at -11 ºC (not frozen), 4 ºC, 22 ºC or 37 ºC. The remaining GDL label after 7 days of storage was determined by MALDI-TOF MS and calculated as the ratio of the total ion count of labelled protein divided by the sum of labelled and unlabelled protein. Error bars represent 1 standard deviation from the arithmetic mean of three independent experiments. See Figure S49 for exemplarily data.

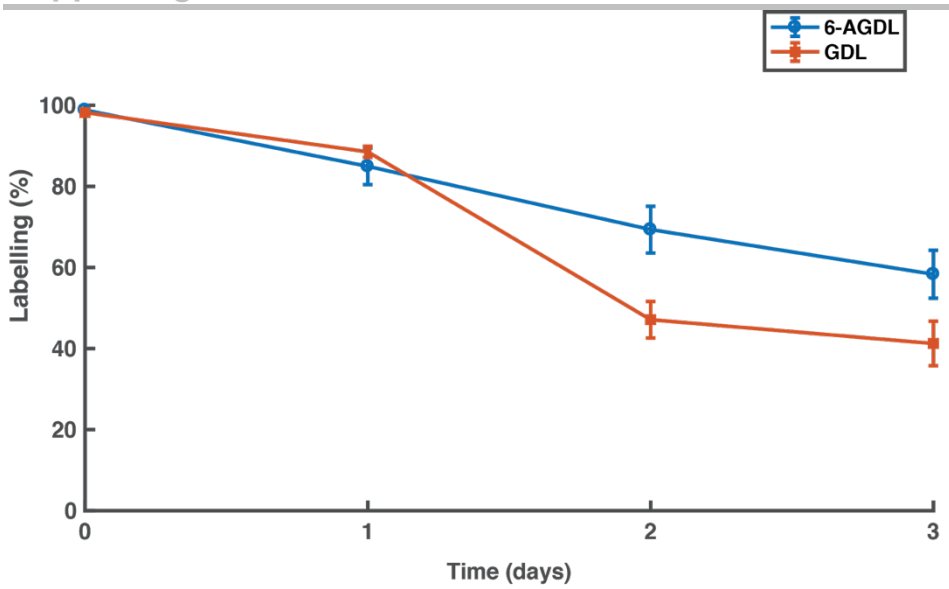


Figure S51. **Comparison between the reversibility of 6AGDL-labelled, and GDL-labelled GSS-H_6-AffiEGFR at pH 7.5** and 37 ºC. 20 µM GSS-H6-AffiEGFR was reacted with either 100 mM GDL or 6AGDL in 500 mM HEPES (NaOH) pH 7.5 for 1 h at RT. Excess label was removed using spin filters and washing repeatedly with 500 mM HEPES (NaOH) pH 7.5. Reversibility was assessed by taking samples at the indicated times, acidification, C18 ZipTip desalting and MALDI-TOF MS analysis. The average mean is plotted and error bars are reported as one standard deviation from the arithmetic mean (n=3).

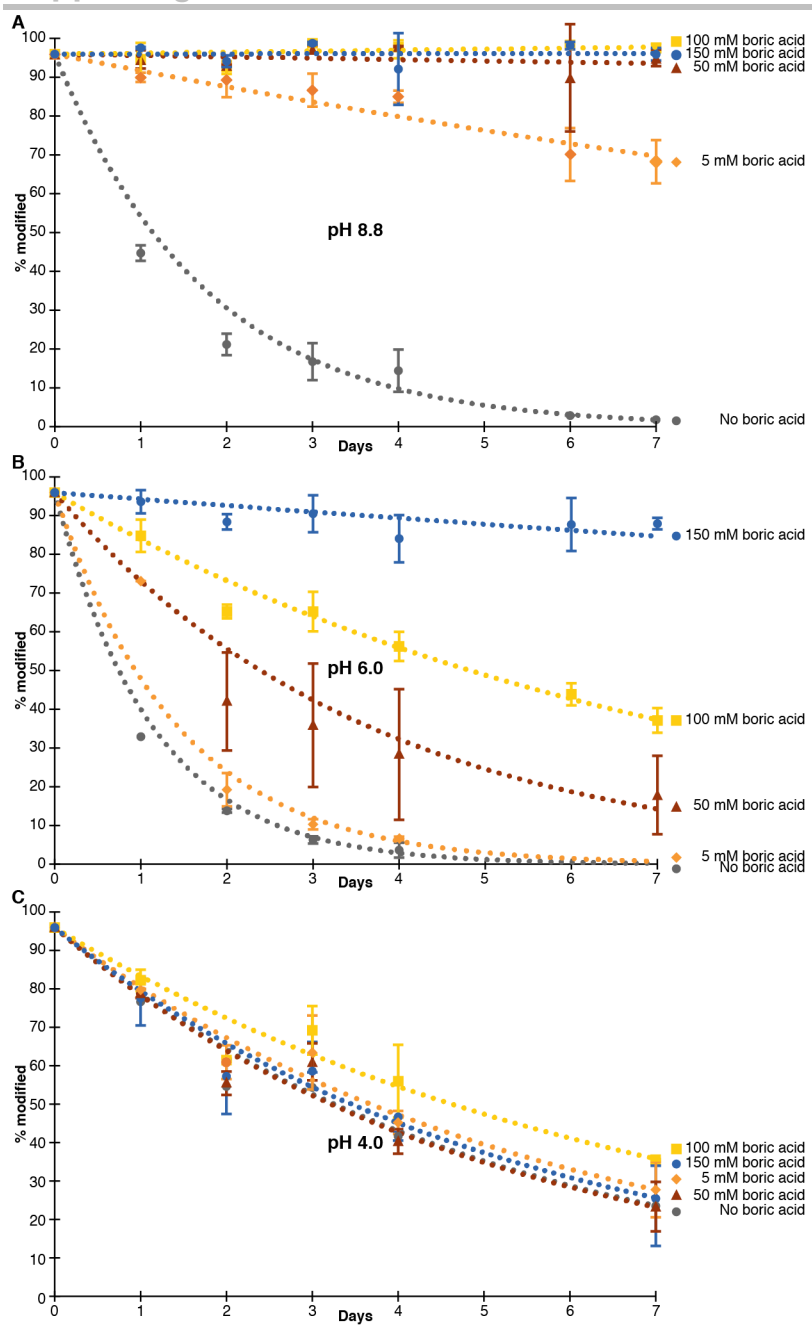


Figure S52. **Reversibility in the presence of boric acid. (A)** at pH 8.8 **(B)** at pH 6.0 **(C)** and pH 4.0. 6AGDL-labelled PP6 peptide was produced as before, extra label was removed by C18 reverse phase purification, and the labelled peptide was stored in different buffers supplemented with various concentrations of boric acid at 37 ºC. Change in the relative *m/z* intensity ($[m_1/z/(m_1/z+m_2/z)]*100$) is reported over time and monitored by MALDI-TOF MS TIC count.

Note: Reversibility was dose- and pH-dependently inhibited. Over a period of 7 days at 37 ºC with >5 mM boric acid at pH 8.8 (ammonium acetate), little reversibility took place, whereas non-boric acid treated samples quickly reversed to their unlabelled state (Figure S52A). Supplementation with boric acid still modulated reversibility at pH ≥6 (sodium citrate, Figure S52B), whereas no effect was observed in acidic condition (pH 4, sodium acetate, Figure S52C). Least boric acid supplementation was required in alkaline conditions (pH 8.8) with 5 mM boric acid supplementation providing a better protective effect as compared to 100 mM boric acid at pH 6. It is known that boric acid can form esters with carbohydrate derived diols.[53] We did not observe the formation of esters directly in MALDI TOF MS spectra, likely because the esters are susceptible to hydrolysis and highly acidic conditions (0.1% v/v TFA) were used to remove buffer reagents prior to spotting. The pH dependence of the protective effect suggests that formation of pH sensitive boronate esters may play a critical role in protecting the polyol label. It is possible that the protective boronic esters are formed with (i) the polyols of the labels, or (ii) with one or both of the two N-terminal Ser hydroxyl groups of PP6 (GSS-H_6-Beltide-1), akin to inhibition of serine proteases with boronic esters.[114] However, we find that supplementation with boric acid also provides protective for gluconoylated peptides that do not contain Ser residues at their N-termini *i.e.* PP1 (G-H_6-Belitde-1) (data not shown).

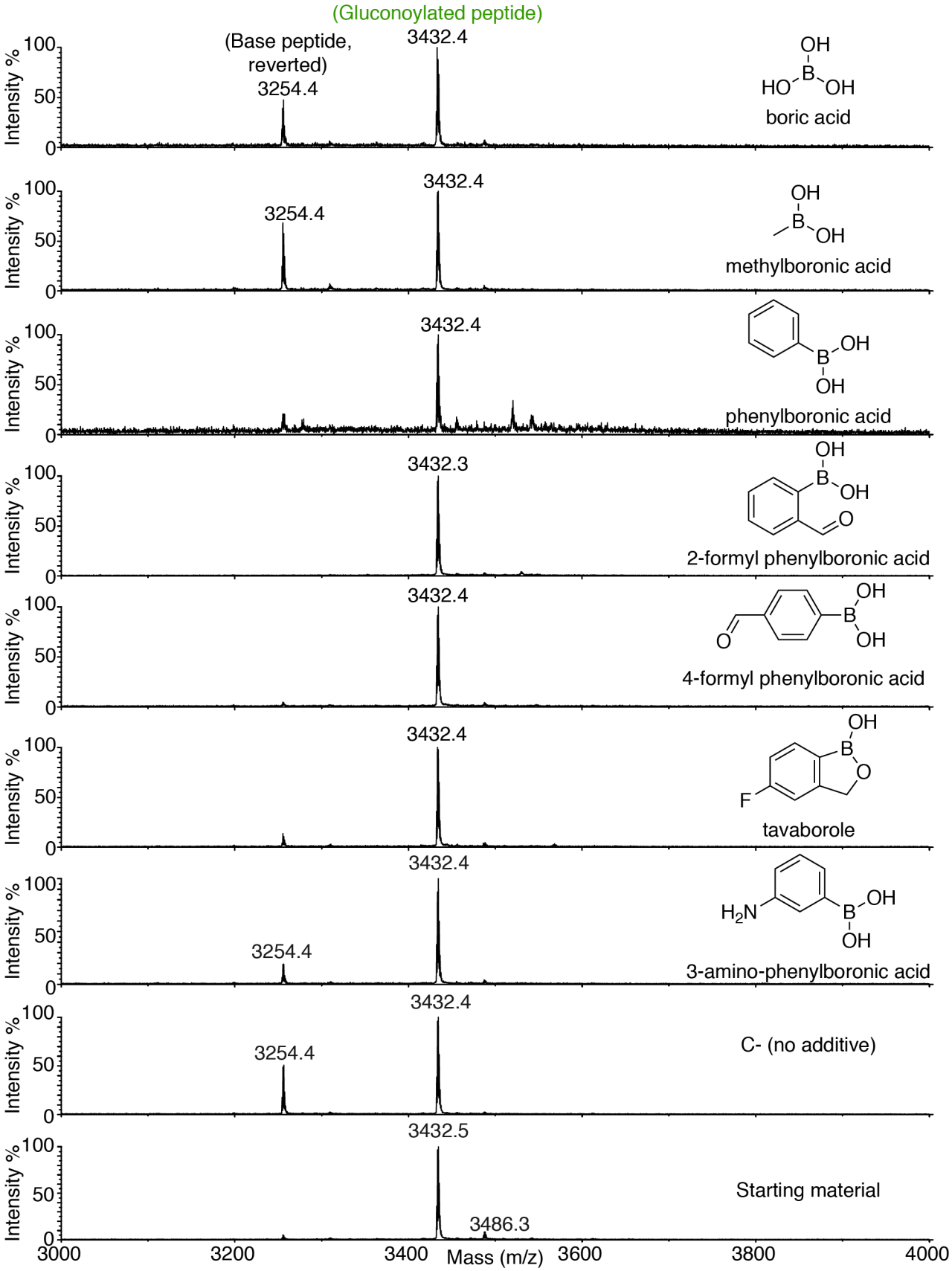


Figure S53. **Reversibility of gluconoylated peptide (PP6) in the presence of boron containing compounds.** 50 µM peptide PP6 labelled with GDL was incubated in 50 mM HEPES pH 7.5 (NaOH), in the presence or absence of boron-containing additives at 5 mM. To accelerate the reversibility reaction, the mixtures were incubated for 15h overnight at 50 ºC in a thermo cycler with heated lid. Next day, take 15 µL sample aliquots were acidify with 5 µL of 50% (v/v) acetic acid, purified by C18 ZipTip C18 and analysis by MALDI TOF MS.

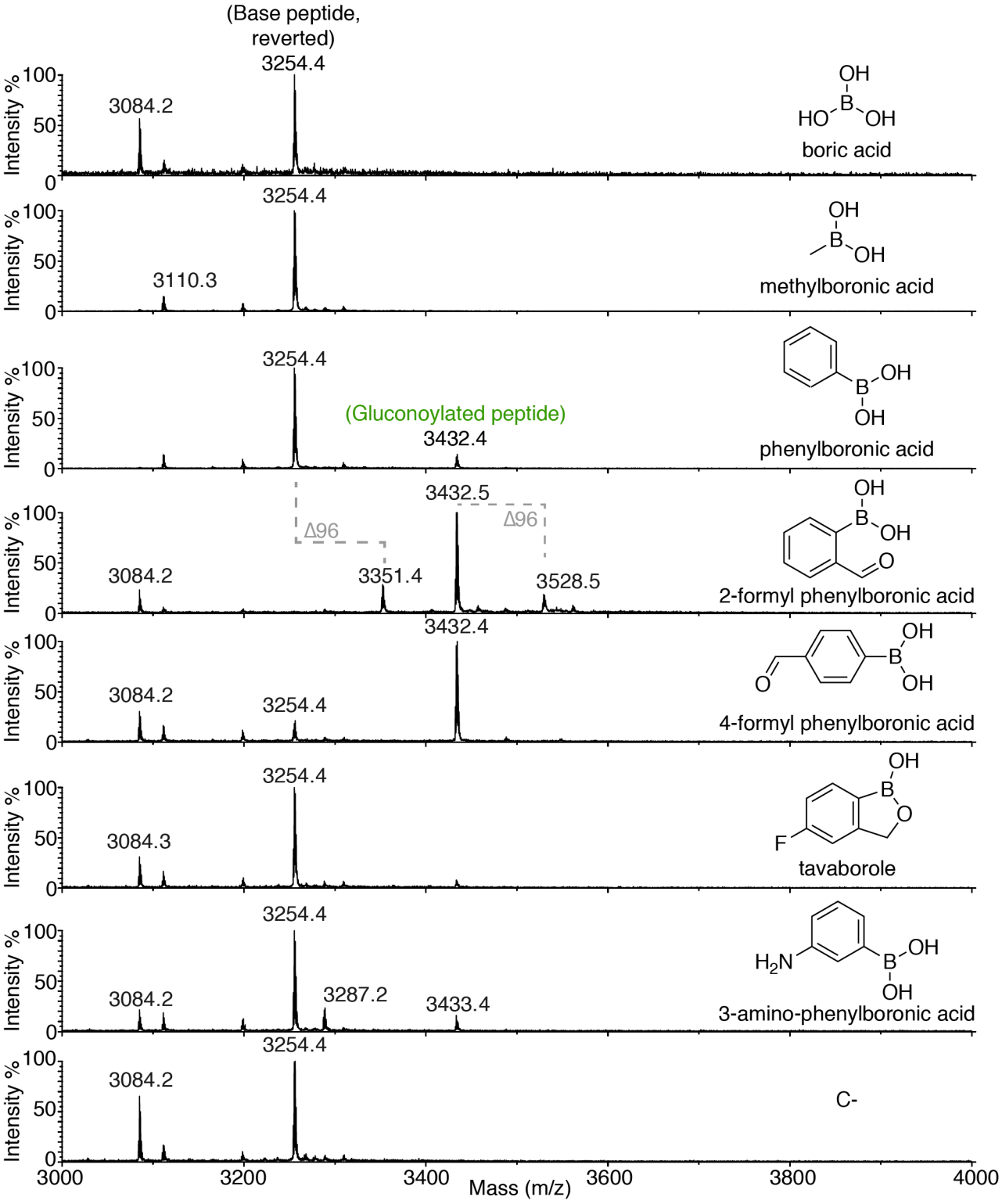


Figure S54. **Reversibility of gluconoylated peptide (PP6) in the presence of 5 mM 2-formylpheylboronic acid, after 5 days at 50 ºC at pH 7.5.** See legend of Figure S53 for experimental details.

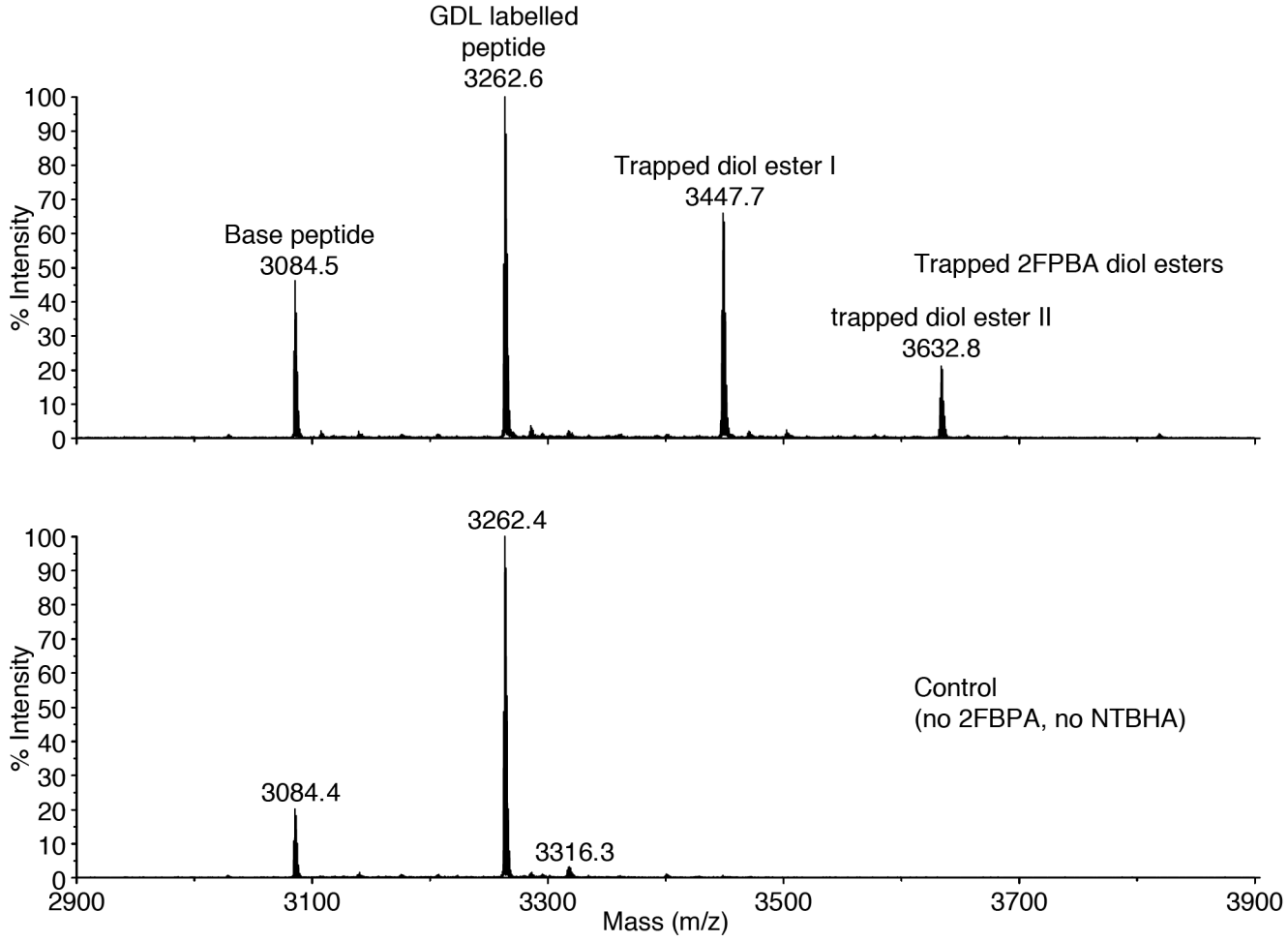


Figure S55. **Trapping of GDL diol derived 2-formylpheylboronic acid ester.** Gluconoylated peptide (PP13) was incubated with 2.5 mM 2-formylpheylboronic acid prior to addition of N-tert-butylhydroxylamine hydrochloride (NTBHA) to 2 mM. The **2-formylpheylboronic acid** concentration was reduced to 2 mM by the addition of NTBHA. The reaction with all components was incubated for 120 min at RT. Reactions were acidified, purified by C18 ZipTip and analysed by MALDI TOF MS.

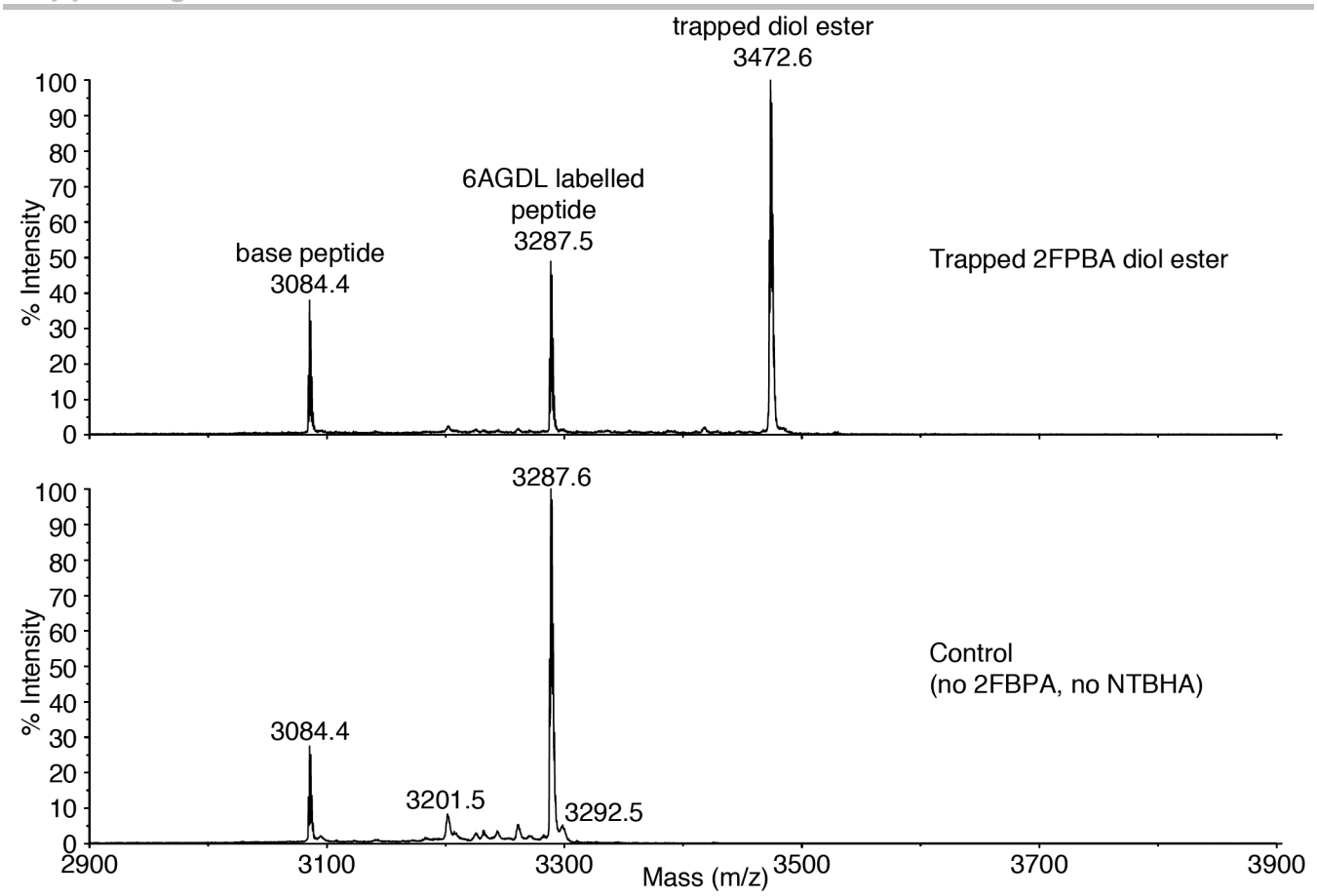


Figure S56. **Trapping of 6AGDL diol derived 2-formylpheylboronic acid ester.** 6-azido-6-deoxy-gluconoylated peptide (PP13) was incubated with 2.5 mM 2-formylpheylboronic acid for 2 minutes prior to addition of NTBHA to 2 mM. The 2-formylpheylboronic acid concentration was reduced to 2 mM by he addition of NTBHA. The reaction with all components was incubated for 150 min at RT. Reactions were acidified, purified by C18 ZipTip and analysed by MALDI TOF MS.

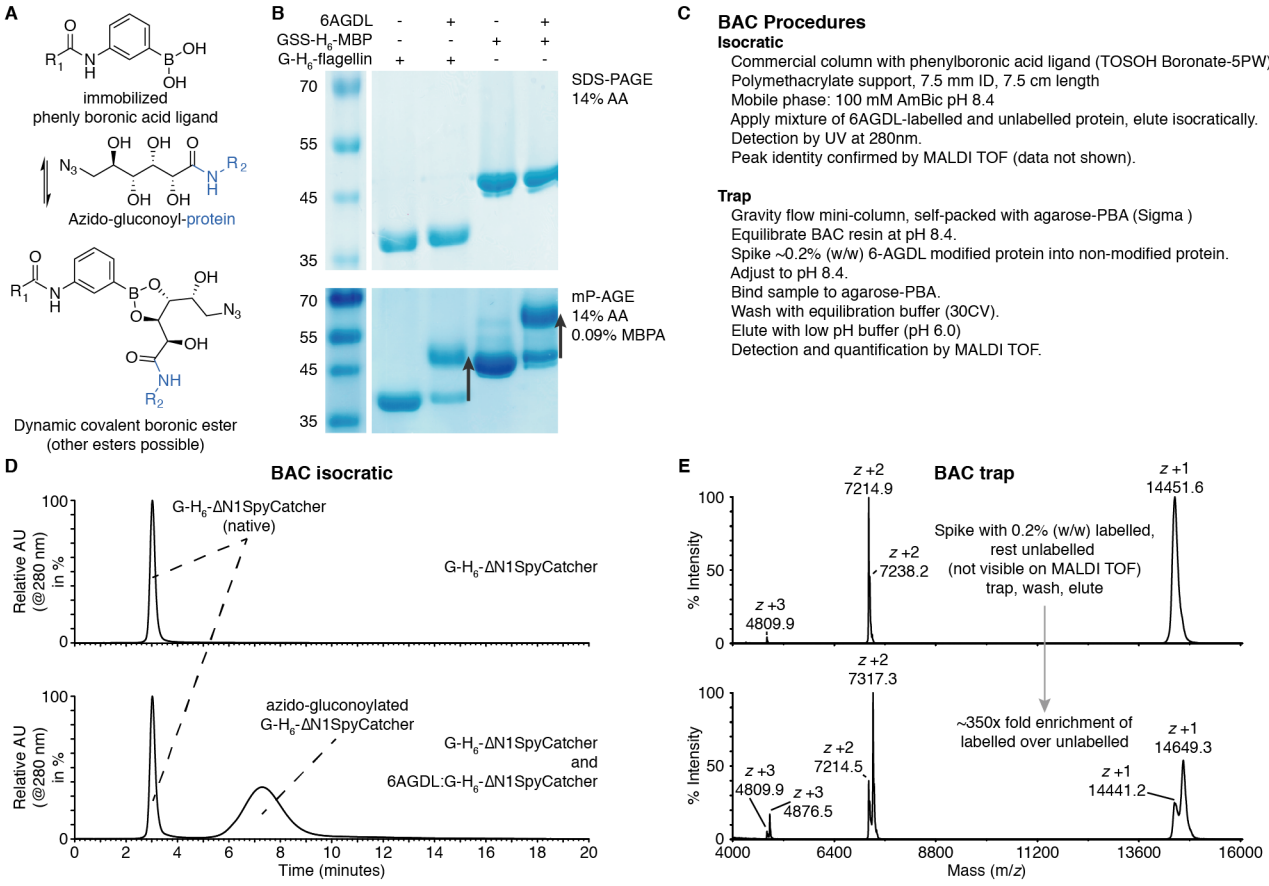


Figure S57 **Azidogluconoylation analysis with a modified SDS-PAGE system.** (A) Possible structure of a dynamic covalent boronic acid ester in equilibrium with free protein. (B) Standard SDS-PAGE (top) shows a single band for both untreated and azidogluconoylated (+203.5 Da) protein, as the molecular weight difference is to small to observe. Leveraging the dynamic covalent equilibrium between 3-methacrylamidophenylboronic acid (MPBA) and the diols of the azido-polyol label, mP-AGE (bottom) azidogluconoylated material migrates dramatically slower due to diol affinity to the gel-incorporated boronic acid, whereas the unmodified fraction of the protein remains unaffected and migrates true to size. mP-AGE only requires addition of commercially available, polymerizable MPBA to SDS-PAGE resolving. (C) Procedure for non-denaturing BAC. (D) Isocratic elution of G-H6-ΔN1-SpyCatcher yields a single sharp peak (top). Admixing with 6AGDL-labelled G-H6-ΔN1-SpyCatcher and elution gives an earlier sharp elution for unlabelled protein as before, and a later broad peak of azidogluconoylated G-H6-ΔN1-SpyCatcher. Identities were confirmed by MALDI TOF MS (not shown). (E) ~0.2% (w/w) 6AGDL modified protein was spiked into non-modified protein solution. Agarose-PBA was equilibrate at pH 8.4, the sample was bound to the agarose-PBA, washed several times with binding buffers and finally eluted with slightly acidic pH step buffer, yielding ~350x fold enrichment over non-labelled as measured by MALDI TOF MS IC.

Note: We propose that Gly-HisTags may be used in tandem, not only to resolve labelled from unlabelled substrate via BAC, but also to remove common IMAC impurities, which are likely poorly (6A)GDL reactive due to unsuitable N-termini (Table S5).

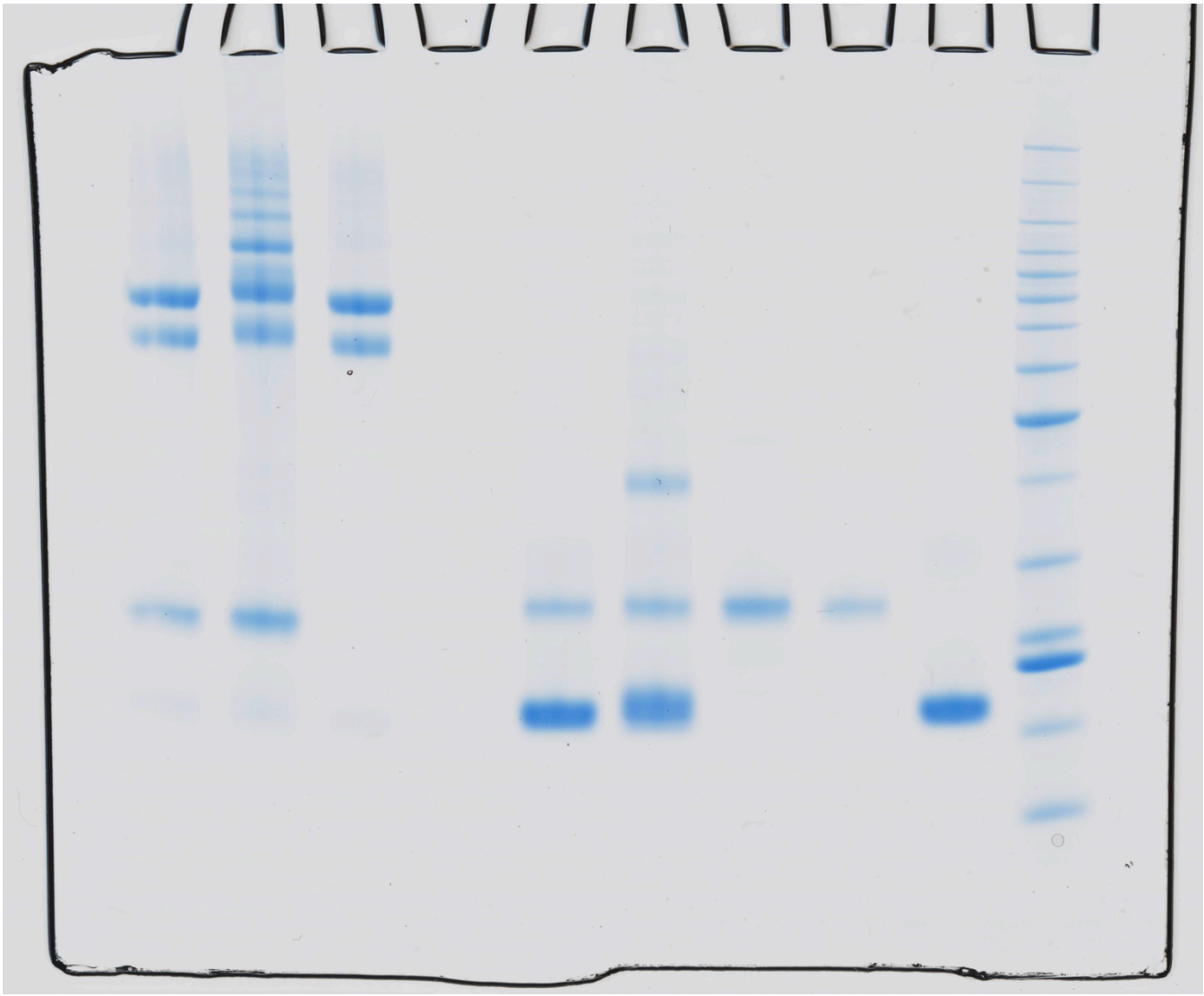


Figure S58. **Uncropped gel.** Loaded: 4 µg Qβ equivalent.

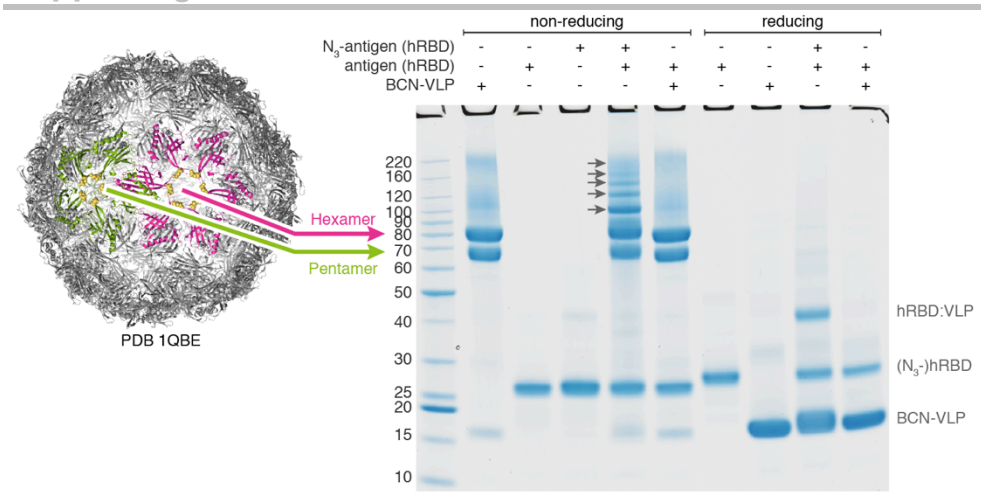


Figure S59 **Higher load gel (11.5 µg Qβ equivalent).** Grey arrows indicate coupling of hRBD(s) to the disulfide-linked penta/hexamer Qβ-subunits.

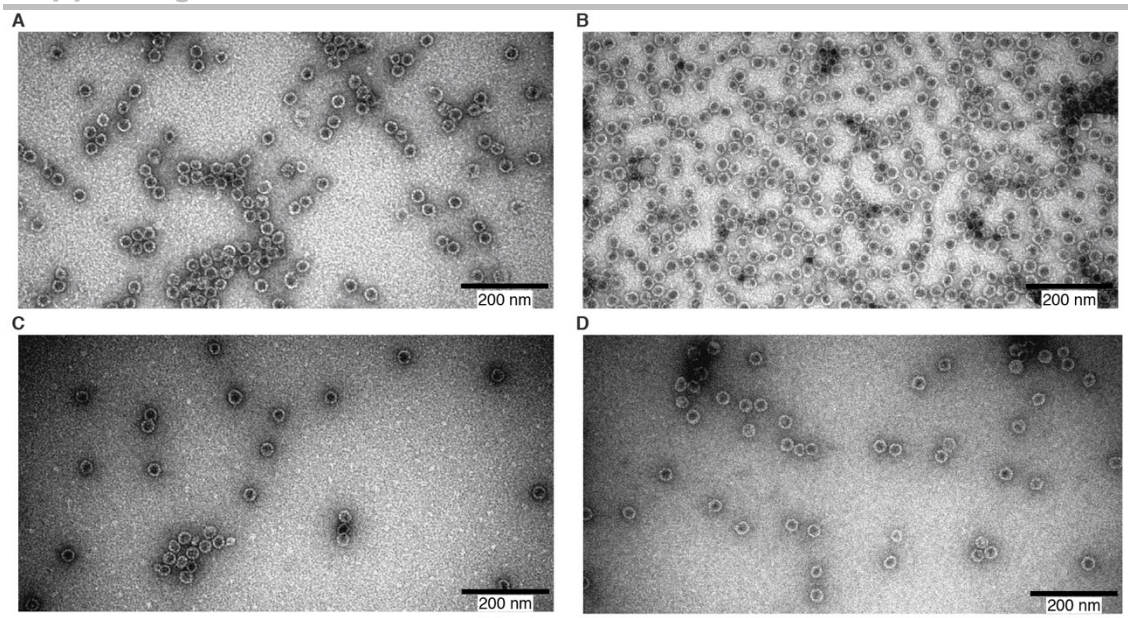


Figure S60 **TEM analysis of Qβ-RBD conjugates used for immunization. (A)** Qβ-BCN-PEG8-NHS+6AGDL-hRBD, **(B)** Qβ-BCN-PEG8-NHS+6AGDL-yRBD, **(C)** Qβ-BCN-NHS+tetrazine-NHS-hRBD, **(D)** Qβ-BCN-NHS+tetrazine-NHS-yRBD. Images were acquired after size-exclusion chromatography into the void.

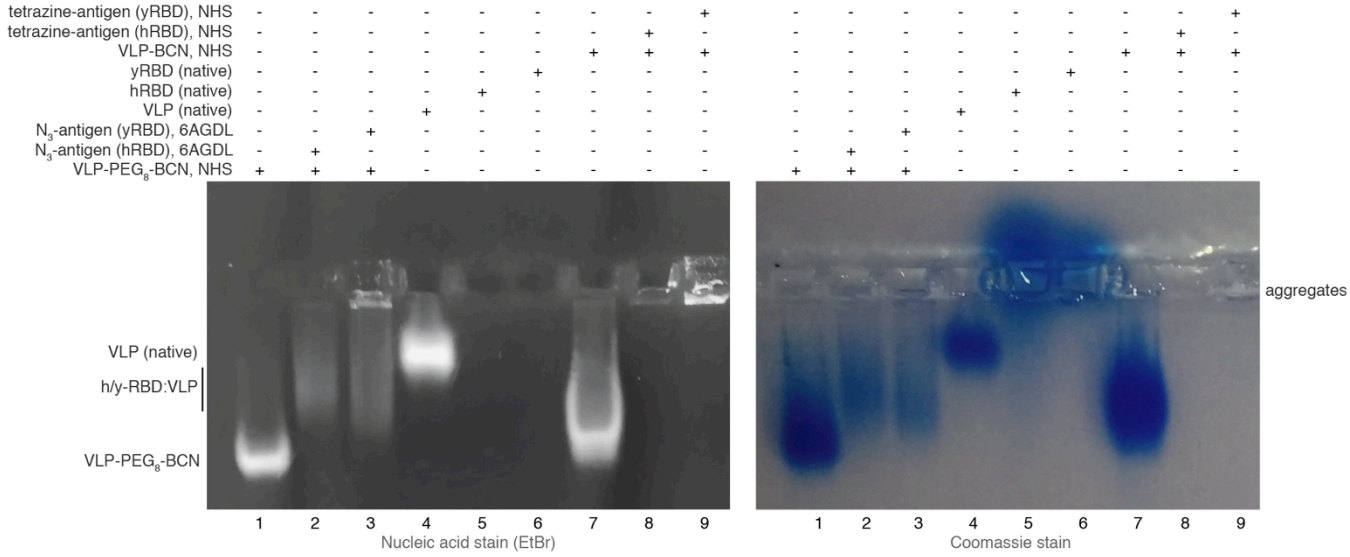


Figure S61 **NAGE analysis of Qβ-RBD coupling products**, stained for nucleic acid (likely RNA) content of VLPs by ethidium bromide (EtBr, left), and protein content by Coomassie blue (right). Lane 1 – Qβ labelled with BCN-PEG8-NHS, 2 – Qβ-BCN-PEG8-NHS/6AGDL-hRBD conjugate, 3 – Qβ-BCN-PEG8-NHS/6AGDL-yRBD conjugate, 4 – unlabelled Qβ, 5 – unlabelled hRBD, 6 – unlabelled yRBD, 7 – Qβ labelled with BCN-NHS, 8 – Qβ-BCN-NHS/tetrazine-NHS-hRBD conjugate, 9 – Qβ-BCN-NHS/tetrazine-NHS-yRBD conjugate. VLP formulations were analysed after size-exclusion chromatography into the void.

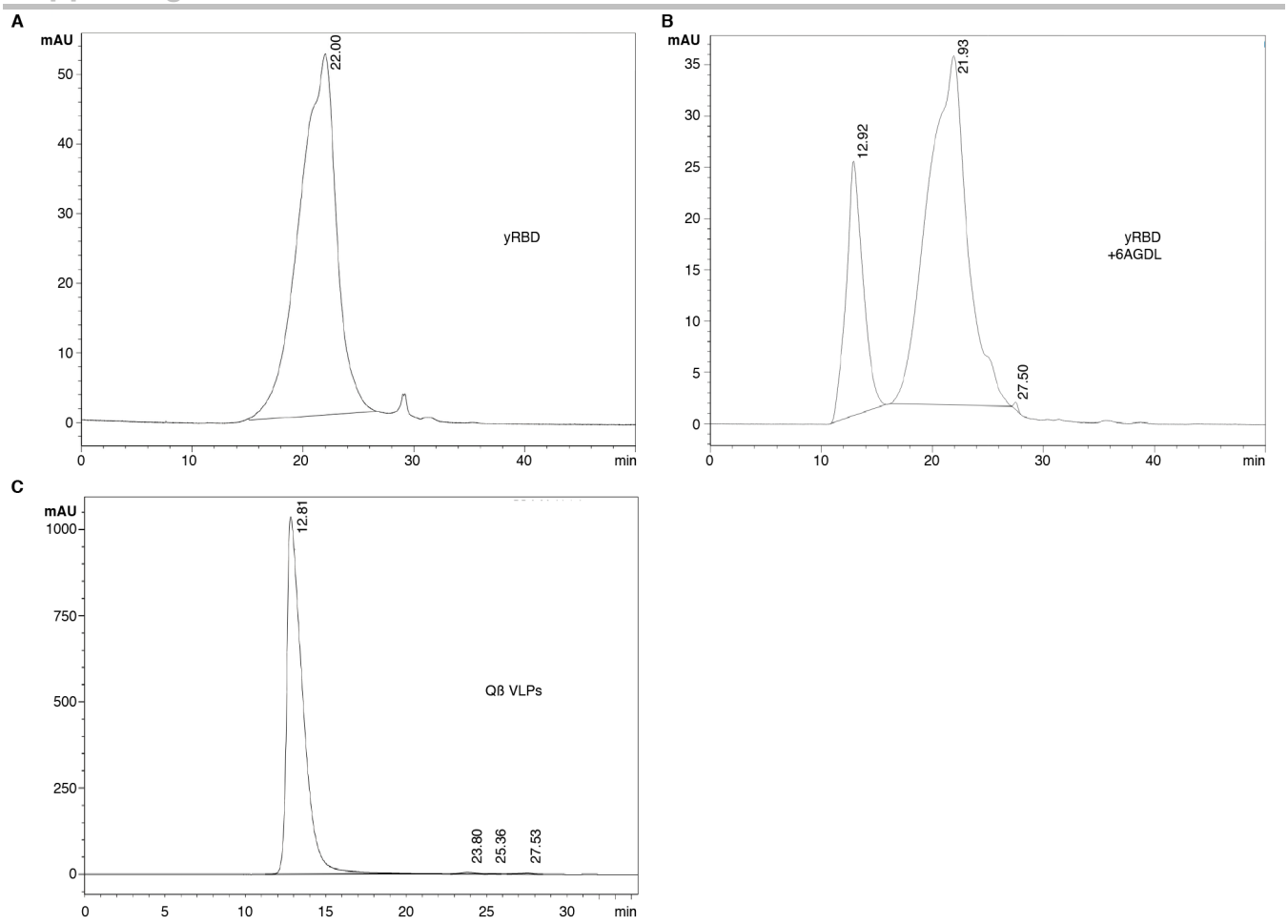


Figure S62 **Analytical gel filtration analysis** of **(A)** unlabelled yRBD, **(B)** 6AGDL-labelled yRBD, and **(C)** Qβ VLPs. Note that Qβ VLPs are eluted into the void on a Superdex 200 (>600 kDa, void volume 12-15 mL).

Note: Previously reported yRBD oligomerisation[115] was also observed for 1x freeze-thawed activated yRBD (Figure S62), and may be aggravated by the linker or method. Unlabelled hRBD also partially aggregated after weeks, 4 ºC, possibly due to RBD's hydrophobicity.

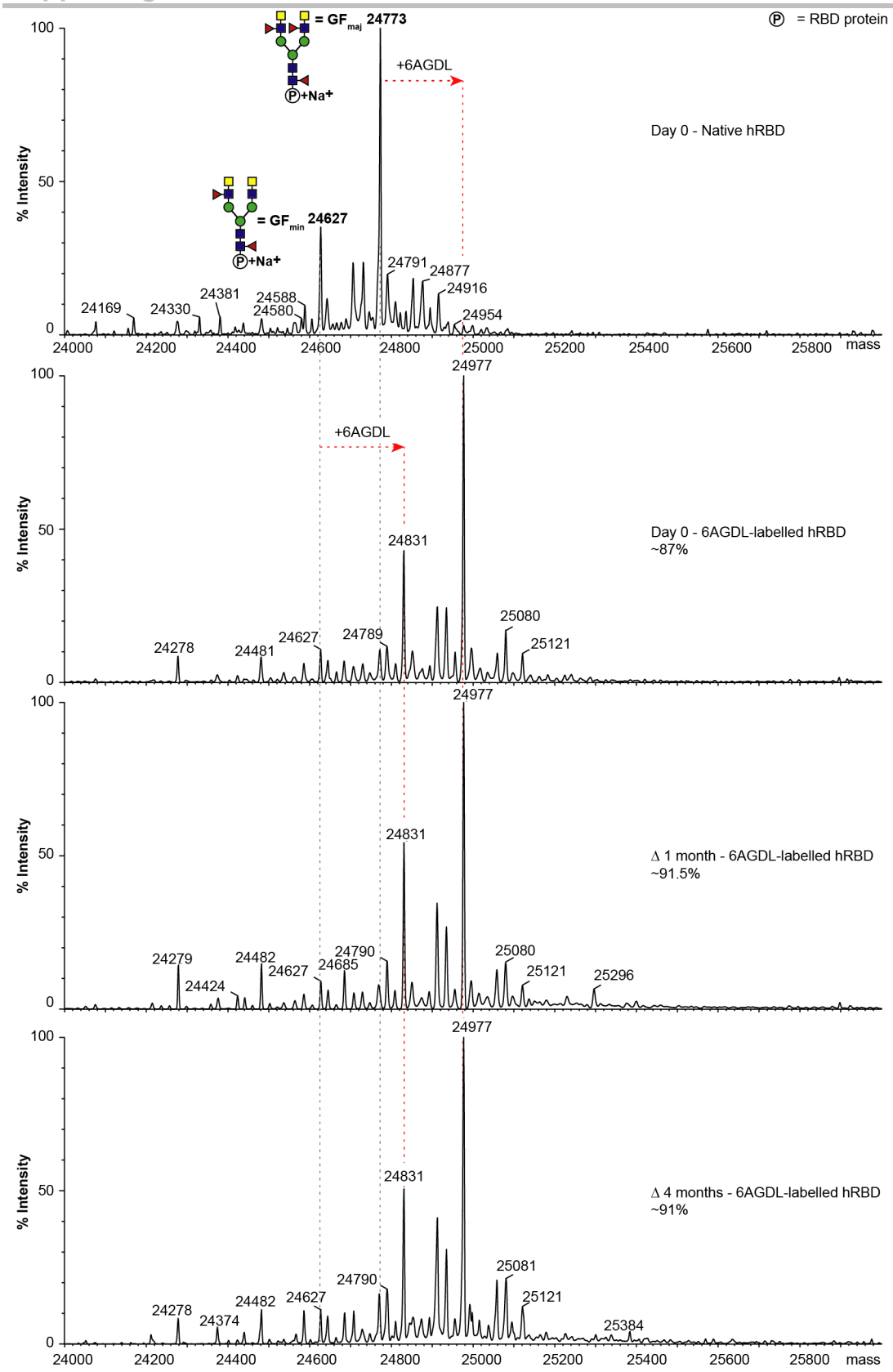


Figure S63 **6AGDL label stability for hRBD over time.** hRBD was labelled with 6AGDL. Excess label was removed by spin-filtration into 50 mM boric acid (pH 8.2). The conjugate was stored at 4 ˚C and ~1.5-2 mg/mL. Aliquots were taken on day 0, and after 1, and 4 months. The extent of azidogluconoylation was monitored by LC/MS Q-TOF. The extent of 6AGDL-labelling was estimated by calculating the ratio of the relative mass intensities ([m1/(m1+m2)]*100%) for the labelled (m1, 24977) and unlabelled (m2, 24773) major glycoform (GF_{maj}). We speculate that the small increase in labelling yield over time may be due to the limitation of using MS for quantification, and/or because unlabelled protein may be precipitating. For simplicity, only the two major glycoforms are annotated.

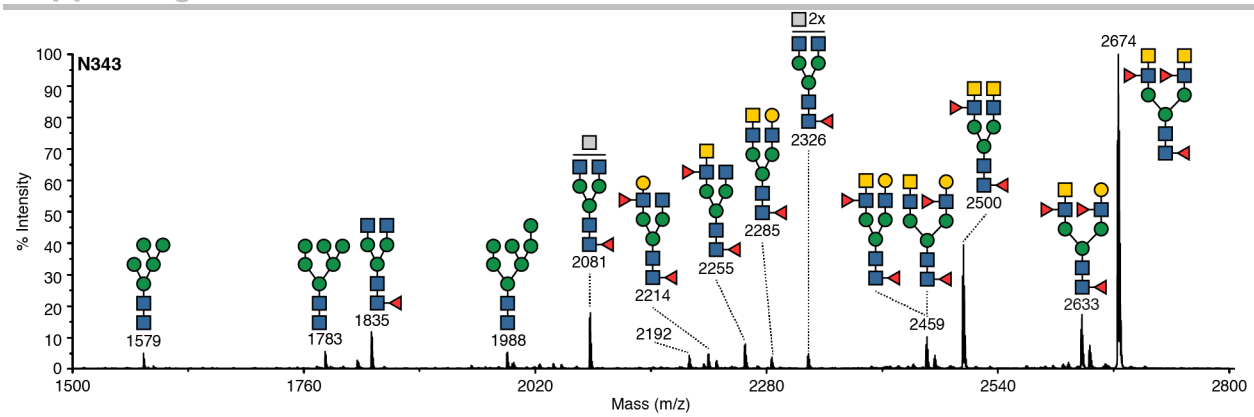


Figure S64 **hRBD glycoform analysis.** MALDI-TOF MS spectrum of permethylated, singly charged sodiated $[M+Na]^+$ N-glycans released by PNGase F from SARS-CoV-2 RBD expressed in HEK293 cells. The proposed structures are based on a previous report[70], and knowledge of mammalian N-glycan biosynthetic pathways.

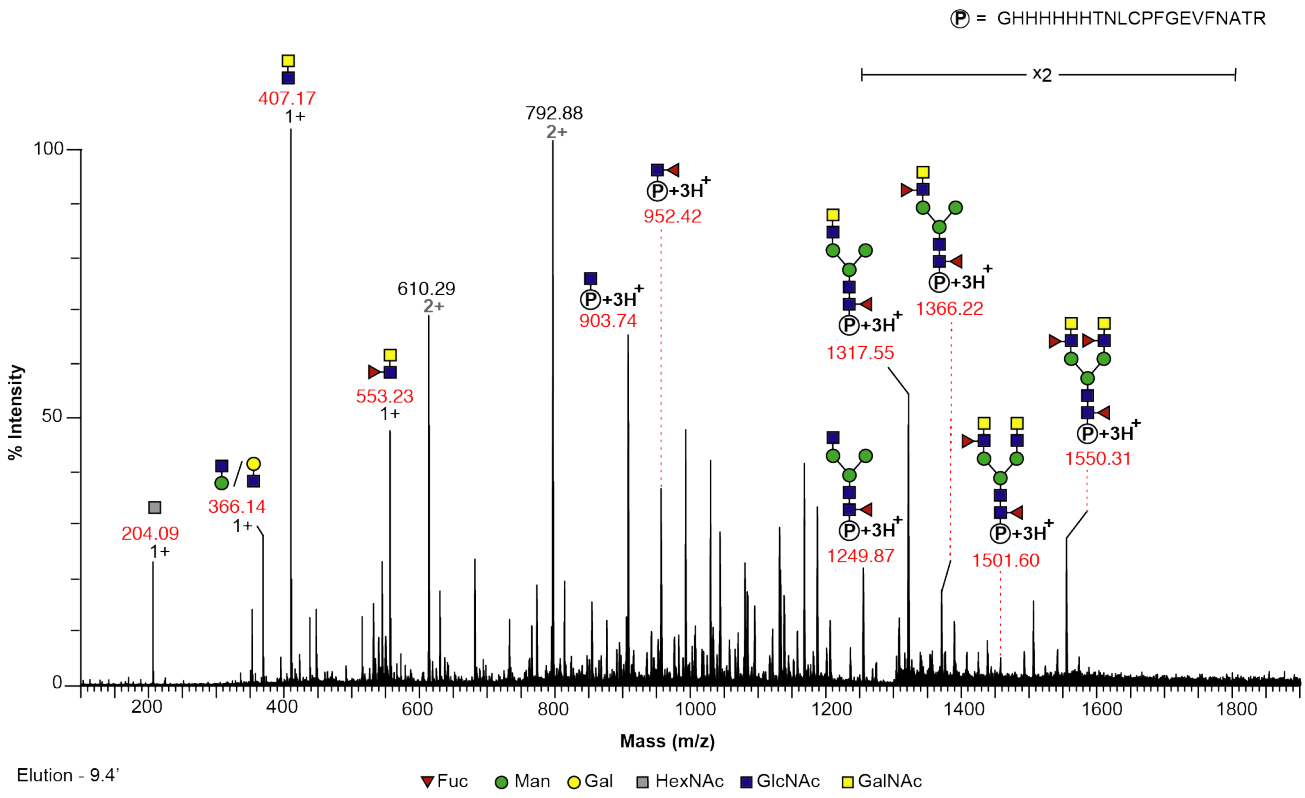


Figure S65 **Q-TOF MS spectrum of GH_6-333-TNLCPFGEVFNATR-346 glycopeptides** (LC elution time ≈ 9.4 min) derived from a trypsin digest of HEK 293 expressed SARS-CoV-2 RBD. For simplicity, only triply charged glycopeptide ions as well as singly charged low molecular weight reporter fragment ions have been annotated. Some signals arise from in-source fragmentation. For a more extensive list of signals observed, see Table S6.

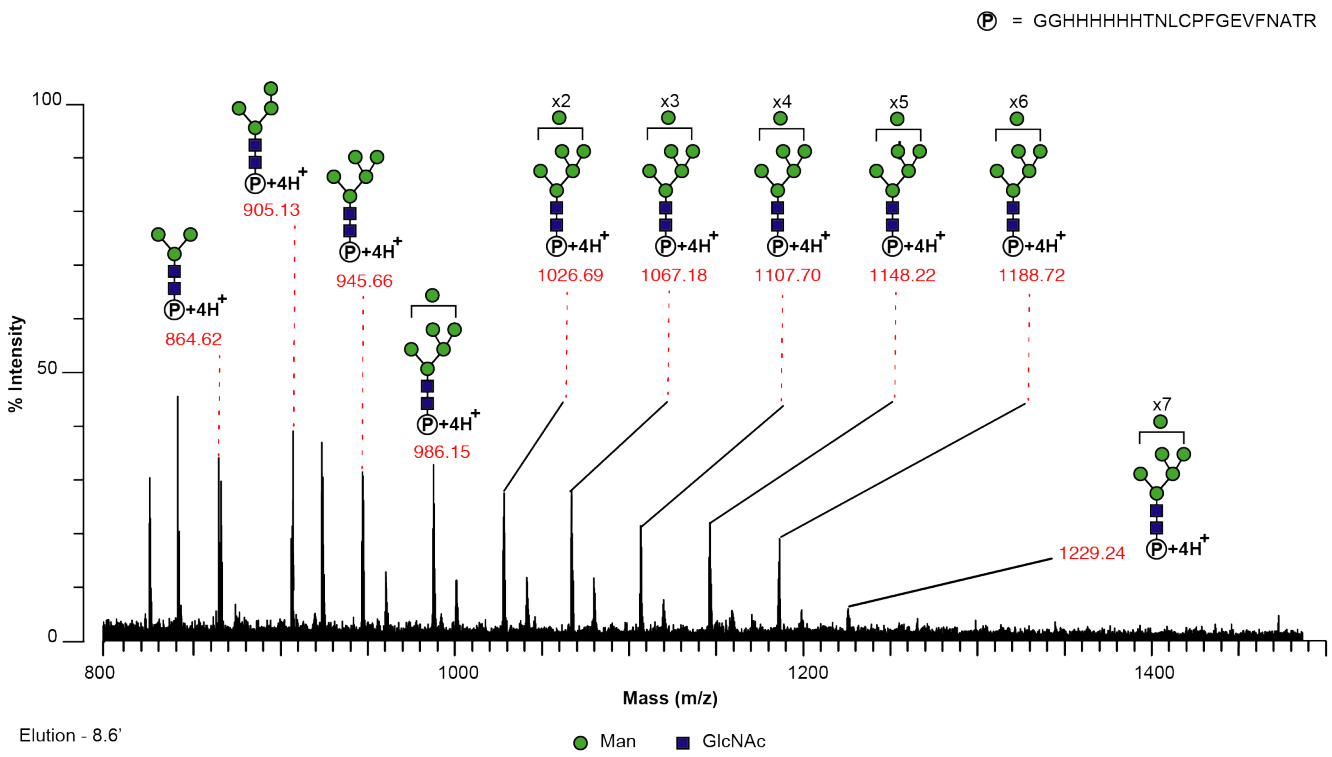


Figure S66 **Q-TOF MS spectrum of glycopeptides** (LC elution time ≈ 8.6 min) derived from trypsin digest of *Pichia pastoris* expressed SARS-CoV-2 RBD. This elution fraction predominantly contained non-phosphorylated GGH_6-333-TNLCPFGEVFNATR-346 glycopeptides. For simplicity, only the most abundant monoisotopic 4+ glycopeptide ions have been annotated. Some signals arise from in-source fragmentation. For a more extensive list of signals observed, see Table S7.

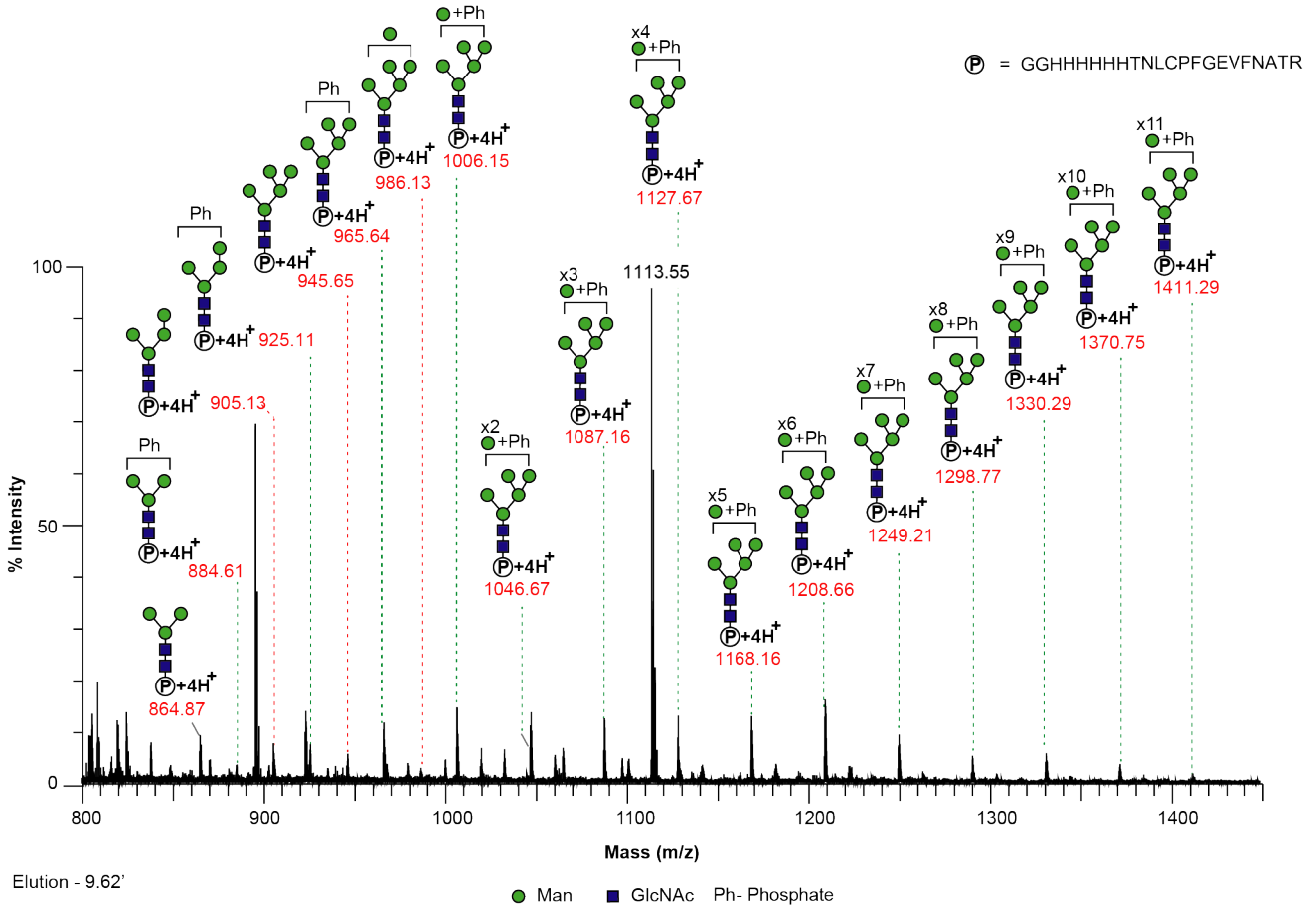


Figure S67 **Q-TOF MS spectrum of glycopeptides** (LC elution time ≈ 9.7 min) derived from a trypsin digest of *Pichia pastoris* expressed SARS-CoV-2 RBD. This elution fraction predominantly contained mono-phosphorylated GGH_6-333-TNLCPFGEVFNATR-346 glycopeptides. For simplicity, only the most abundant monoisotopic 4+ ions have been annotated. Some signals arise from in-source fragmentation. For a more extensive list of signals observed, see Table S8.

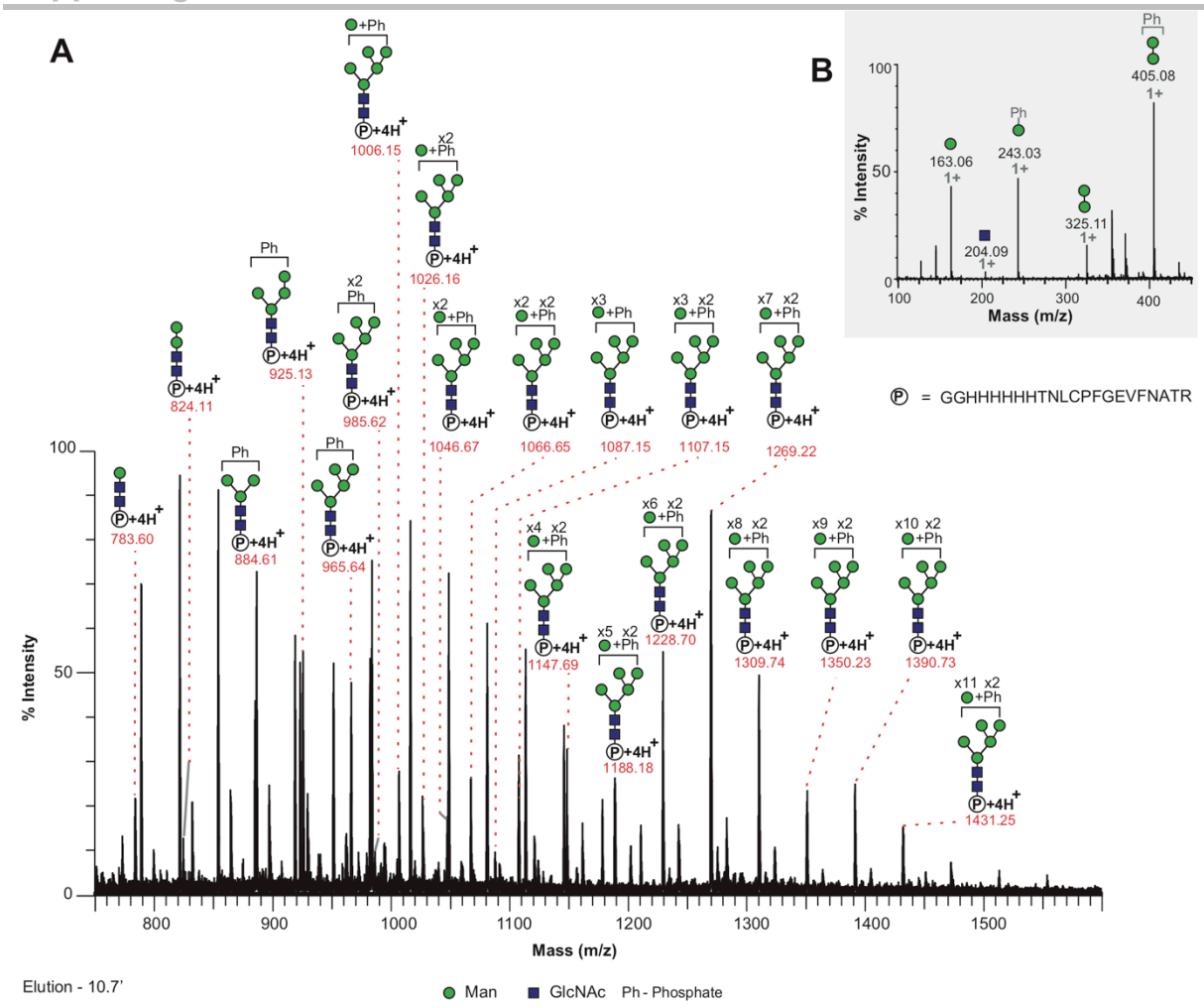


Figure S68 **QTOF MS spectrum of glycopeptides** (LC elution time ≈ 10.8 min) derived from a trypsin digest of *Pichia pastoris* expressed SARS-CoV-2 RBD. This elution fraction predominantly contained di-phosphorylated GGH_6-333-TNLCPFGEVFNATR-346 glycopeptides. **(A)** For simplicity, only the most abundant monoisotopic 4+ glycopeptide ions, **(B)** as well as low-molecular weight singly charged reporter fragment ions have been annotated. Some signals arise from in-source fragmentation. For a more extensive list of signals observed, see Table S9.

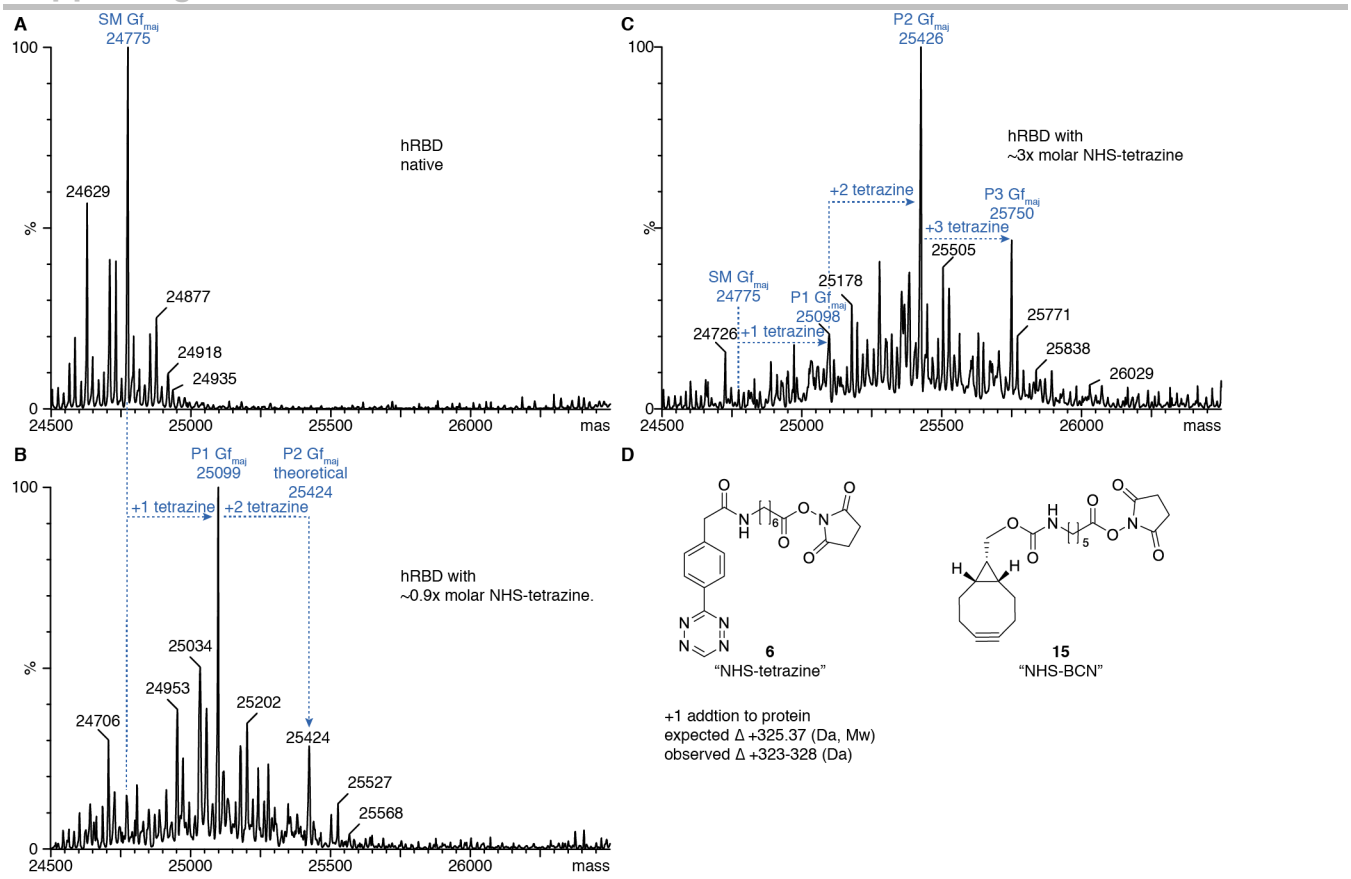


Figure S69. **hRBD activation with NHS-tetrazine.** Labelling was conducted at pH 7.5, room temperature.

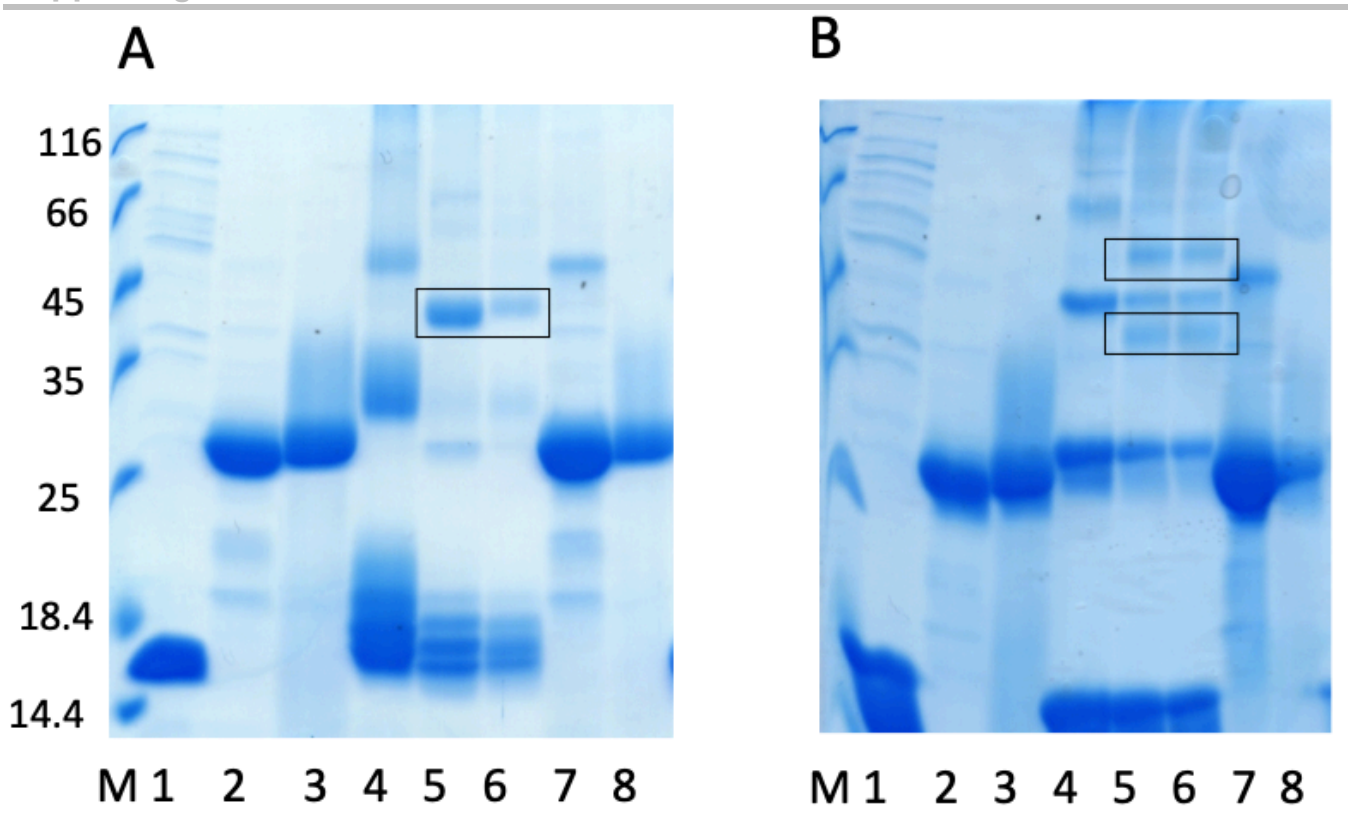


Figure S70 **Reducing SDS-PAGE (15%) analysis of Qβ-RBD coupling products.** All lanes marked with "M" contain molecular weight marker. The conjugate products are framed. **(A)** Qβ-BCN-PEG8-NHS/6AGDL-RBD materials used for immunization. 1 – unlabelled Qβ, 2 – unlabelled hRBD, 3 – unlabelled yRBD, 4 - Qβ labelled with BCN-PEG8-NHS, 5 – purified Qβ-BCN-PEG8-NHS/6AGDL-hRBD conjugate after void-chromatography on a S200 column, 6 – purified Qβ-BCN-PEG8-NHS/6AGDL-yRBD conjugate after chromatography on S200 column, 7 - hRBD labelled with 6AGDL, 8 - yRBD labelled with 6AGDL. **(B)** Qβ-BCN-NHS/tetrazine-NHS-RBD materials used for immunization. 1 – unlabelled Qβ, 2 – unlabelled hRBD, 3 – unlabelled yRBD, 4 - Qβ labelled with BCN-NHS , 5 – purified Qβ-BCN-NHS/tetrazine-NHS-hRBD conjugate after chromatography on S200 column, 6 – purified Qβ-BCN-NHS/tetrazine-NHS-yRBD conjugate after chromatography on S200 column , 7 - hRBD labelled with tetrazine-NHS, 8 - yRBD labelled with tetrazine-NHS. The conjugate products are framed.

Note: Some small amounts of free RBD can be observed after size-exclusion, which is likely aggregated and thus co-elutes into the void (Figure S70).

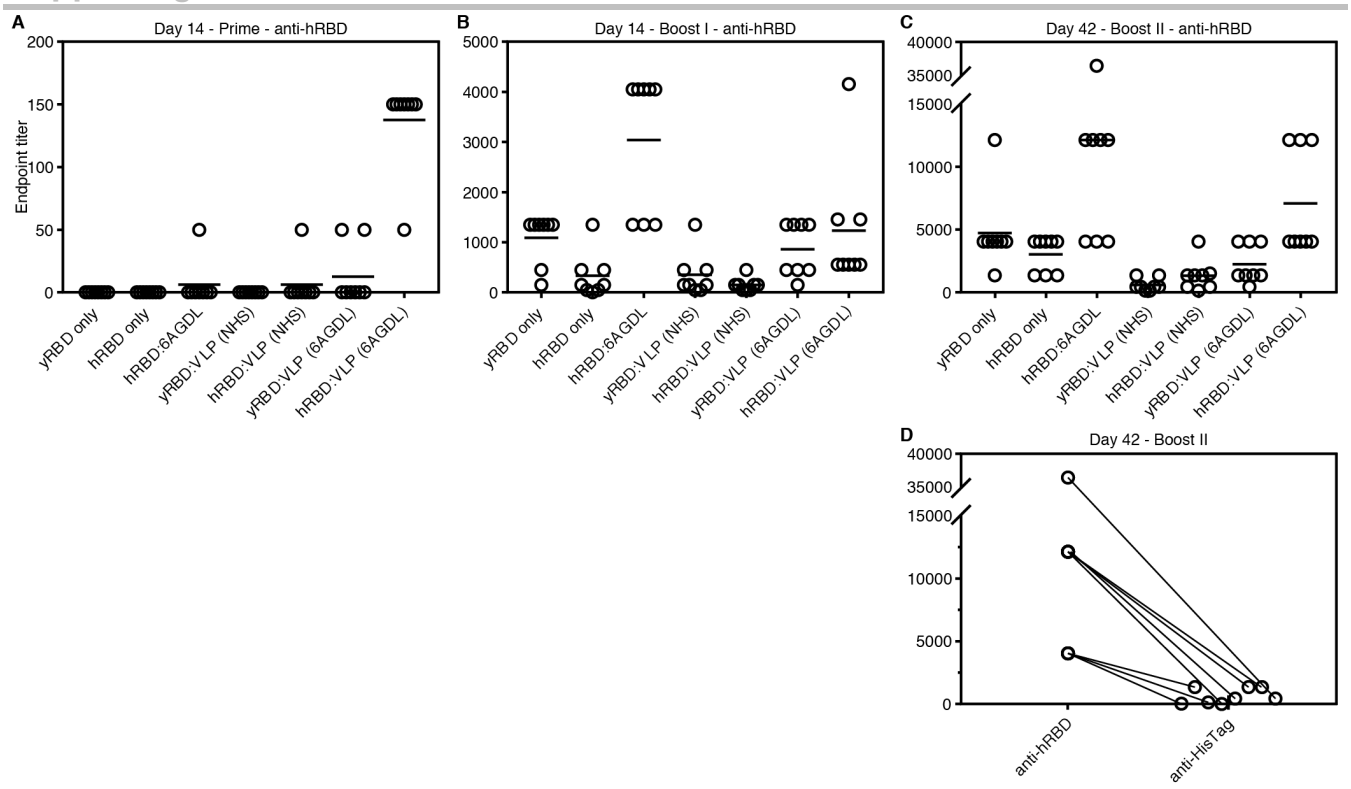


Figure S71 **ELISA endpoint titers. (A)** After prime (anti-hRBD), **(B)** first boost (anti-hRBD), and **(C)** second boost (anti-hRBD). Individual values plotted per animal per circle. The geometric mean for each group is indicated by a vertical line. **(D)** Endpoint titers after second boost for anti-hRBD and anti-HisTag (unrelated HisTagged protein) for the activated hRBD group; lines connect values for each individual animal endpoint.

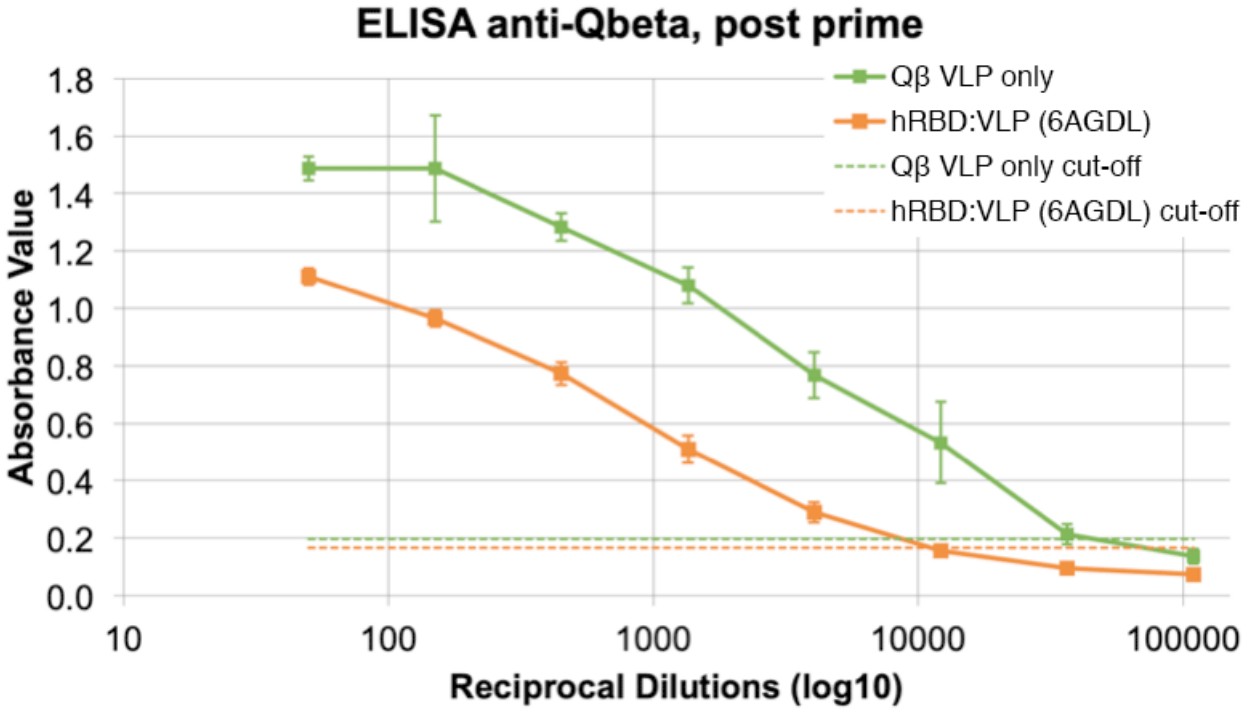


Figure S72 **ELISA against Qβ after priming.**Mice were primed with Qβ VLP alone, or with hRBD conjugated to Qβ via 6AGDL [hRBD:VLP (6AGDL)].

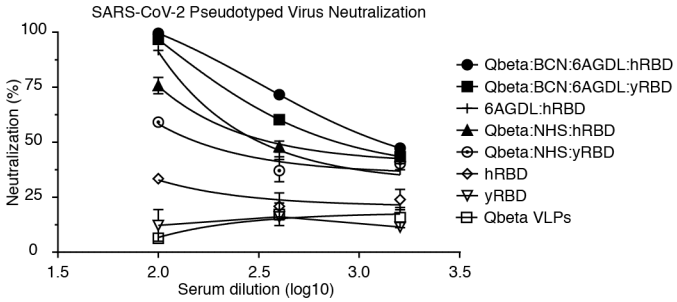


Figure S73 **Pseudovirion neutralization assay**. Pooled serum samples from each immunized group were pre-incubated with SARS-CoV-2 spike pseudotyped MLV and used for infection of HEK293T-ACE2 cells. The percentage of infected cells (expressing GFP reporter) was quantified by flow cytometry and the normalized neutralization was calculated as a ratio from infection control (without serum). The neutralization percentage curve was analyzed using nonlinear regression function and the neutralization titres (NT50, 70, and 90) are represented as the serum dilutions resulting in 50, 70 and 90% neutralization in Table S10.

Note: Neutralizing activity of immunized mice sera was analysed using SARS-CoV-2 pseudotyped-MLV in HEK293T-ACE2 cells. As a positive control for neutralization, the anti-SARS CoV-2 S specific antibody (Sino Biological) providing complete neutralization at dilution 1:200, was used. MLV particles produced without SARS-CoV-2 S protein encoding construct was used as a negative control. Importantly, the HEK293T (ACE2 negative) cells were not susceptible to SARS-CoV-2 pseudotyped-MLV infection (<0.1%), whereas the HEK293T-ACE2 cells showed 11 % of infection according to flow cytometry analysis. The mice sera were pooled for each group (n=8), pre-incubated with SARS-CoV-2 MLV and used for HEK293T-ACE2 cell infection. Three serum dilutions were tested (1:100, 1:400, 1:1600).

Sera from 6AGDL-activated monomeric hRBD also showed possible improved performance over native monomeric protein in VNT, which however did not reach statistical significance (Figure 6), but appears more pronounced in PVNT assays. Antigen aggregation alone may have caused increased immunogenicity, however we cannot rule out an adjuvant effect for the 6AGDL-linker presented here, as a patent application claims adjuvant properties for gluconoylated antigen[116], and linear polyols play an important role in MR1 metabolite antigen presentation.[117] Further studies will be required to rule out, or conversely elucidate a mechanism for (azido)-gluconoyl adjuvant activity.

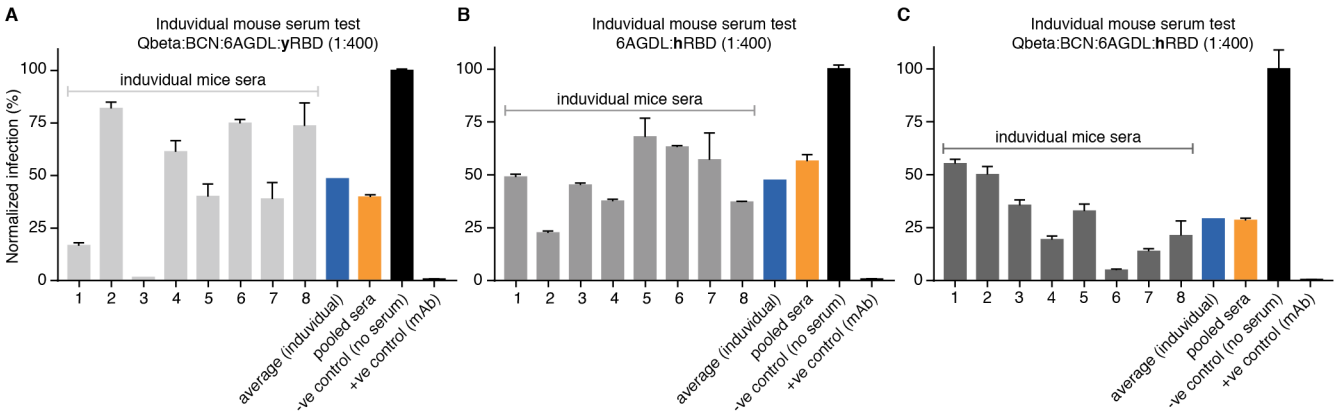


Figure S74 **Pseudovirion neutralization assay.** SARS-CoV-2 pseudotyped-MLV infection inhibition obtained by virus pre-incubation with individual mouse serum (n=8) of selected groups. Individual average value represents the average normalized % of infection of eight tests (n=8), the pooled sera represent the % of infection from polled sample (n=8). Error bars represent the standard deviation of two experiments.

Supporting Tables

Table S1. Estimated extent of labelling of proteins.

		MALDI-TOF MS (Average, SD, Uncertainty by TIC)						
Protein	N-ter. sequence	Mono-	Di-	Mw (kDa)	# Lys	# Cys	2PCA	Spectra
GSS-H_6-SUMO	GSS-H_6	94.9 ± 3.4% (2.1%)	1.1 ± 1.2% (0.1%)	12.2	9	0	+	Figure S1
GSS-H_6-SUMO(H23A)	GSS-H_6	88.6 ± 6.3% (3.4%)	1.6 ± 0.4% (0.1%)	12.1	9	0	+	Figure S2
GS-SUMO-H_6-SSK	GSSDSEVNQ	23.5 ± 2.8% (2.5%)	0.3 ± 0.2% (0.3%)	12.4	10	0	+	Figure S3
GSS-H_6-AffiEGFR-	GSS-H_6	93.8 ± 2.2% (4.5%)	1.6 ± 0.4% (0.2%)	14.4	10	0	+	Figure S4
G-H_6-SnoopCatcher	GSS-H_6	94.7 ± 1.7% (2.7%)	0.7 ± 0.1% (0.6%)	14.8	9	0	+	Figure S5
G-H_6-Cys-ΔN1SpyCatcher	G-H_6-DC	95.3 ± 2.3% (1.0%)	1.6 ± 0.2% (0.2%)	12.7	8	1	+	Figure S6
NbSyn2-H_6	GQLVESGGG	10.4 ± 1.5% (8.6%)	0.5 ± 0.3% (0.8%)	16.4	6	4	+	Figure S7
RNase B	KETAAAKFE[118]	n.d.	n.d.	~15	10	8	+	Figure S8

Model proteins (~20 µM) were incubated with 125 mM GDL in 200 mM HEPES pH 7.5 at RT (22 °C) for 1 h. The N-terminal amino-acid sequence is reported as found in literature or as the DNA-encoded amino acid sequence after Met-cleavage. The extent of GDL-labelling is calculated as the ratio between the IC for the labelled species divided by the sum of the IC for both unlabelled species and labelled label species. Labelling is reported as the arithmetic mean ± 1 standard deviation from the mean (MALDI-TOF MS, n=3) or from a single experiment (LC-MS Q-TOF, n=1). Uncertainty (MALDI-TOF MS) is reported as the IC ratio for the *m/z* value of a hypothetically labelled species in a mock reaction containing no GDL. The number of expected Lys and Cys residues is reported. The availability of the α-amino group of the N-terminus was assessed by 2PCA-labelling (MALDI-TOF MS) and is reported as "+" for substrates that demonstrated *m/z* shift to the expected single adduct. For experimental data see Supplementary information.

Note: Using MALDI-TOF MS in linear mode, some Total Ion Current (TIC) signal could be observed for non-label exposed samples where singly labelled protein would be expected (+178 m/z ≈ 1% of total adduct mass). This is due to limitations in resolving power for MALDI TOF-MS instrumentation in the linear (non-reflectron) mode, where target proteins of the investigated size (12-17 kDa) and associated expression host-derived target protein modifications, such as acetylation (+42 *m/z*) can create wide and "tailing" peaks. All species were most abundantly observed with charge number +2, which increases apparent resolution but still leaves ambiguity. In an attempt to quantify this "uncertainty", we calculated pseudo labelling as the ion current ratio for a hypothetical labelled species in a mock reaction containing no GDL.

Table S2 Estimated extent of labelling for various peptides.

Name	Beltide-1 variant	Sequence	Mono (%)	Mono StDev	Di (%)	Di StDev	Figure
PP6	Yes	**GSSHHHHHH**DWLKAFYDKVAEKLKEAF	83	9.6	0.3	0.03	Figure S10A
PP2	Yes	**GHHH**DWLKAFYDKVAEKLKEAF	76	9.3	0.0	0.01	Figure S10B
PP1	Yes	**GHHHHHH**DWLKAFYDKVAEKLKEAF	75	3.8	0.1	0.02	Figure S11A
PP3	Yes	**GHHHH**DWLKAFYDKVAEKLKEAF	73	10.5	0.1	0.02	Figure S11B
PP4	Yes	**GGHHHHHH**DWLKAFYDKVAEKLKEAF	69	9.1	0.0	0.02	Figure S12A
PP7	Yes	**GGKWSKR**DWLKAFYDKVAEKLKEAF	53	0.7	0.1	0.01	Figure S12B
PP5	Yes	**HHHHHH**DWLKAFYDKVAEKLKEAF	46	7.1	0.0	0.00	Figure S13A
PP8	Yes	**GASGSKG**DWLKAFYDKVAEKLKEAF	46	1.5	0.0	0.00	Figure S13B
PP9	Yes	**GG**DWLKAFYDKVAEKLKEAF	41	3.3	0.0	0.00	Figure S14A
EMP1		GGTYSCHFGPLTWVCKPQGG	36	2.8	0.1	0.03	Figure S14B
PP10	Yes	**GGTYSDH**DWLKAFYDKVAEKLKEAF	34	3.0	0.1	0.04	Figure S15A
PP11	Yes	DWLKAFYDKVAEKLKEAFDELY***K*HHHHHH**	18	6.5	0.1	0.01	Figure S15B
PP12	Beltide-1	DWLKAFYDKVAEKLKEAF	15	6.1	0.0	0.00	Figure S16A
Aprotinin		RPDFCLE…	6	0.1	2.3	0.09	Figure S16B
GluFib		EGVNDNEEGFFSAR	3	0.9	0.0	n.d.	Figure S17

Peptides were reacted at 50 µM in potassium phosphate buffer (pH 7.5) with 50 mM GDL for 1 h at room temperature (22 ºC). The reactions were stopped by addition of acetic acid, followed by ZipTip C18 clean up and MALDI TOF MS analysis. Yields from three separate experiments with single measurements are reported as TIC of labelled vs. the sum of all species. StDev is the standard deviation from the average mean. In bold, appended residues to base peptide Beltide-1.[19] Underlined, Lys residue highlight for attempted intra-chain His-proximity labelling. Note: 50 mM GDL was chosen as a sub-optimal labelling concentration to allow discerning differences in peptide reactivity. Higher labelling yields are achievable with 100 mM GDL, and / or repeated labelling

Table S3 Selectivity comparison of 6AGDL 4-methoxyphenyl 2-azidoacetate as assessed by MALDI-TOF MS.

Reagent			4MPAA			6AGDL			4MPAA			
Concentration			2.5 mM			100 mM			4 mM			
PP1	Time (h)	Temp. (ºC)	Single label	Double label	Unlabelled	Single label	Double label	Unlabelled	Single label	Double label	Triple label	Unlabelled
	1	5.4	10±2.7%	0.4±0.1%	90±2.7%	98±1.0%	0.3±0.1%	1.9±0.9%				
	18	5.4	59±1.4%	4.6±0.4%	36±1.5%	98±0.5%	0.7±0.0%	1.4±0.5%				
	24	5.4	60±1.5%	5.1±0.5%	34±1.9%	96±0.6%	0.9±0.1%	2.8±0.6%	67±4.3%	7.7±0.7%	0.5±0.1%	25±4.5%
	1	23	19±2.9%	0.5±0.1%	80±2.9%	98±0.4%	0.4±0.1%	1.2±0.3%				
	18	23	47±3.9%	3.0±0.4%	50±4.0%	88±2.8%	0.5±0.1%	12±2.8%				
	24	23	51±3.6%	3.4±0.5%	46±3.6%	82±3.5%	0.6±0.1%	17±3.6%	48±9.2%	0.4±0.4%	0.4±0.1%	48±10.0%
PP6	Time	Temp.	Single label	Double label	Unlabelled	Single label	Double label	Unlabelled	Single label	Double label	Triple label	Unlabelled
	1	5.4	8±1.7%	0.3±0.1%	92±1.7%	84±1.2%	0.3±0.1%	15±1.2%				
	18	5.4	38±0.5%	4.2±0.8%	58±1.4%	98±0.2%	0.8±0.2%	1.4±0.1%				
	24	5.4	40±3.1%	4.5±0.2%	55±3.3%	98±0.7%	1.0±0.1%	1.4±0.7%	53±7.8%	10.1±3.9%	0.7±0.3%	37±11.9%
	1	23	17±1.2%	0.6±0.2%	83±1.4%	90±1.9%	0.4±0.0%	9.2±1.8%				
	18	23	29±0.2%	1.3±0.1%	70±0.1%	90±1.5%	0.5±0.1%	9.9±1.4%				
	24	23	35±7.8%	2.1±1.3%	63±9.1%	91±3.9%	0.6±0.1%	8.3±4.0%	36±3.3%	2.5±0.2%	0.2±0.0%	62±3.5%

1 mM peptide was reacted with either 2.5 mM or 4 mM 4-methoxyphenyl 2-azidoacetate (4MPAA), or 100 mM 6AGDL for the indicated amount of time at the indicated temperature (ºC) in 200 mM HEPES pH 7.5. The extent of labelling is reported as the IC for the respective species divided by the sum of ICs for all species at a given charge state. Single experiment, triple measurement. Colour scaling was performed across time comparing all conditions for a single attribute and peptide. Green indicates desirable observation (high extent of mono-functionalization), and red indicates undesirable observation (either low mono-functionalization and/or significant double-labelling). See Figure S41 for exemplarily spectra.

Note: Both base peptides, PP1, and PP6 were modified using 4MPAA or 6AGDL at all temperatures tested. 100 mM 6AGDL treated peptide PP1 was quickly mono-functionalized (>98%) within 1 hour upon addition of labelling reagent at both 5.4 ºC and RT (23 ºC) with very little off-site labelling (~0.3%). In contrast, labelling with 2.5 mM 4MPAA showed only ~10% (5.4 ºC) or ~19% (RT) of mono-functionalized peptide. We did not observe the same extent of labelling as previously reported for labelling of 1 mM PP1 (G-H_6-Beltide-1) with 2.5 mM 4MPAA at 4 ºC over a period of 24 h [previously reported 92% mono-, 8% di-functionalised protein.[19] When we reacted PP1 with 2.5 mM 4MPAA for 24 h at 5.4 ºC, we instead obtained 60% mono-functionalization with ~5% dual-functionalization. We speculated that our preparation of 4MPAA may not have been as pure as the reported one. Upon increasing the concentration of 4MPAA to 4 mM, ~67% mono-functionalized peptide could be obtained for PP1, albeit at the cost of increasing offsite dual-functionalization (~7-8%) (24 h, 5.4 ºC) (Figure S41).

Compared to PP1, PP6 was less reactive with any labelling reagent tested after reacting for 1 h. Nevertheless mono-functionalised peptide was observed at over >80% at either temperature after 1 h incubation with 100 mM 6AGDL. Importantly, mono-functional labelling to ≥ 98% with little dual-functionalization (~0.8%) could still be obtained, when incubating with 100 mM 6AGDL at 5.4 ºC for prolonged periods (18 or 24 h) (Figure S41). In contrast, labelling of PP6 with either 2.5 mM or 4 mM 4MPAA for the same amount of time resulted in only ~40%, and ~53% when incubated at 5.4 ºC. Strikingly, significant dual-functionalisation was observed as a function of increasing 4MPAA concentrations (~5%, and ~10%). Our findings for 4MPAA are in somewhat in contrast to previous reports for labelling of PP1 where higher mono-functionalized but similar dual-functionalization were reported.[19] The highest concentration of 4MPAA tested by us was 4 mM, which gave ~67% mono-functionalization when tested with the highly-reactive G-H_6 N-terminus of PP1 (5.4 ºC, 24 h). Although higher concentrations of 4MPAA could have been used to possibly obtain higher mono-functionalized product yields, reacting with 4 mM already resulted in ~7-8% of duo-functionalized product, which is in close agreement to what was previously reported. We incubated our reactions overnight in a commercial fridge, which after calibrtion testing was found to be at ~5.4 ºC and not 4 ºC. It may be possible that the slight temperature difference may have caused the higher

extent of duo- to mono-labelling ratio. However, duo-functionalization for reactions incubated at room temperature for the same amount of time showed less double-labelling, casting doubt on a hypothetical negative-temperature dependence effect.

Another possibility may be that the ionization properties (ionization efficiency) of the native and differently-labelled peptide species are not the same. Therefore, our attempts at estimating the extent of labelling via MALDI-TOF MS may not give accurate values, possibly overestimating the extent of dual-labelling. However, we have used the same methodology, MALDI-TOF MS, to estimate the extent of labelling for both compounds. Hence the overestimation assumption of off-site labelling would likely be true for 6AGDL as well,-suggesting that 6AGDL is the more selective labelling reagent.

Thus we speculate that our observation of increased off-site labelling may be due use of MALDI-TOF MS for quantification, as opposed to the use of LC-MS with quantification via UV at 215 nm by Martos-Maldonado et al. 2018.[19] Chromatography may spread multipli- but differently-labelled species over time if the modifications are at different sites resulting in less sensitivity for UV detection.

Table S4. Reversibility of non-His-tagged Beltide-1 peptides in comparison to G-H_6-Beltide-1.

ID	Description	Day 0 (%)	Day 4 (%)	Day 5 (%)	Day 14 (%)
PP7	GGKWSKR-Beltide-1	84.8	80.8	82.1	64.4
PP10	GGTYSDH-Beltide-1	95.8	92.8	93.2	74.6
PP1	G-H_6-Beltide-1	94.1	70.1	61.1	26.9

The extent of 6AGDL labelling of 100 µM PP7, PP10 and PP1 peptide was followed over the course of 14 days at room temperature (~22 ºC), at pH 7.5 in 1 M HEPES (NaOH) buffer. The extent of labelling is reported as (TIC of mono-labelled)/(TIC of unlabelled + TIC of mono-labelled + TIC of duo-labelled) x 100%, where TIC is Total Ion Count from MALDI-TOF MS data. The data is based on a single experiment.

Table S5. Common IMAC contaminants from *E. coli* lysate.

Abbreviation	Protein	UniProt	Monomer mass (kDa)	N-terminus
DnaK	chaperone protein DnaK	P0A6Y8	69	GKIIGIDLGT
GroEL	chaperonin GroEL	Q6Q099	57	AAKDVKFGND
DnaJ	chaperone protein DnaJ	P08622	41	AKQDYYEILG
RplB	50S ribosomal protein L2	P60422	30	AVVKCKPTSP
CA	carbonic anhydrase 2	P61517	25	MKDIDTLISN
SlyD	FKBP-type peptidyl-prolyl cis-trans isomerase	P0A9K9	20	MKVAKDLVVS
Fur	DNA-binding transcriptional dual regulator	P0A9A9	17	TDNNTALKKA
GroES	cochaperonin GroES	P0A6F9	10	MNIRPLHDRV

Note: We speculate that (6A)GDL labelling after cell lysis and initial IMAC purification, followed by BAC, could remove HCPs. Common IMAC Ni-NTA contaminants are known.[119] With the exception of DnaK, all listed HCP are expected to be poor substrates for gluconoylation and hence would not benefit from subsequent BAC purification.

Table S6 N-glycan compositions of hRBD.

Mass observed (neutral)	Mass predicted	Ions observed	Composition
2708.22	2708.22	903.74 (3+)	HexNAc
2854.26	2854.28	952.42 (3+)	HexNAc-dHex
2911.23	2911.30	971.41 (3+)	HexNAc2
3235.38	3235.41	1079.46 (3+)	HexNAc2-Hex2
3381.42	3381.46	1128.14 (3+)	HexNAc2-dHex-Hex2
3397.41	3397.46	1133.47 (3+)	HexNAc2-Hex3
3543.57	3543.52	1182.19 (3+)	HexNAc2-dHex-Hex3
3584.64	3584.54	1195.88 (3+)	HexNAc3-dHex-Hex2
3600.54	3600.54	1201.18 (3+)	HexNAc3-Hex3
3746.61	3746.60	1249.87 (3+)	HexNAc3-dHex-Hex3
3892.62	3892.65	1298.54 (3+)	HexNAc3-dHex2-Hex3
3908.64	3908.65	978.16 (4+)	HexNAc3-dHex-Hex4
3949.64 3949.65	3949.68	988.41 (4+) 1317.55 (3+)	HeNAc4-dHex-Hex3
4095.76 4095.66	4095.73	1024.94 (4+) 1366.22 (+3)	HeNAc4-dHex2-Hex3
4111.68	4111.73	1028.92 (4+)	HeNAc4-dHex-Hex4
4152.68	4152.76	1039.17 (4+)	HeNAc5-dHex-Hex3
4241.72	4241.79	1061.43 (4+)	HeNAc4-dHex3-Hex3
4257.80	4257.79	1065.45 (4+)	HeNAc4-dHex2-Hex4
4298.80	4298.81	1075.7 (4+)	HeNAc5-dHex2-Hex3
4355.96	4355.83	1089.99 (4+)	HeNAc6-dHex-Hex3
4444.92	4444.87	1112.23 (4+)	HeNAc5-dHex3-Hex3
4460.88	4460.87	1116.22 (4+)	HexNAc5-dHex2-Hex4
4501.88 4501.80	4501.89	1126.47 (4+) 1501.60 (3+)	HexNAc6-dHex2-Hex3
4606.88	4606.92	1152.72 (4+)	HexNAc5-dHex3-Hex4
4647.96 4647.93	4647.95	1162.99 (4+) 1550.31 (3+)	HexNAc6-dHex3-Hex3

N-glycan compositions (LC elution time ≈ 9.4 min) observed for the glycopeptide GH_6-333-TNLCPFGEVFNATR-346 obtained by Q-TOF MS analysis of trypsin digested SARS-CoV-2 RBD protein expressed in HEK 293 cells. The main signals corresponding to monoisotopic glycopeptide masses are reported and do not represent and exhaustive list. Some signals arise from in-source fragmentation.

Table S7 Non-phosphorylated N-glycan compositions of yRBD.

Mass observed (neutral)	Mass predicted	Ions observed	Glycan composition
3454.48	3454.48	864.62 (4+)	HexNAc2-Man3
3616.52	3616.53	905.13 (4+)	HexNAc2-Man4
3778.64	3778.59	945.66 (4+)	HexNAc2-Man5
3940.60	3940.64	986.15 (4+)	HexNAc2-Man6
4102.76	4102.69	1026.69 (4+)	HexNAc2-Man7
4264.72	4264.74	1067.18 (4+)	HexNAc2-Man8
4426.80	4426.80	1107.70 (4+)	HexNAc2-Man9
4588.88	4588.85	1148.22 (4+)	HexNAc2-Man10
4750.88	4750.90	1188.72 (4+)	HexNAc2-Man11
4912.96	4912.96	1229.24 (4+)	HexNAc2-Man12

Non-phosphorylated N-glycan compositions (LC elution time ≈ 8.6 min) observed for the glycopeptide GGH_6-333-TNLCPFGEVFNATR-346 obtained by Q-TOF MS analysis of trypsin digested SARS-CoV-2 RBD protein expressed in *P. pastoris* yeast cells. Main signals corresponding to monoisotopic glycopeptide masses have been included in the table and do not represent and exhaustive list. Some signals arise from in-source fragmentation.

Table S8 Mono-phosphorylated N-glycan compositions of yRBD.

Mass observed (neutral)	Mass predicted	Ions observed	Glycan composition
3534.44 3534.45	3534.45	884.61 (4+) 707.89 (5+)	HexNAc2-Man3-Phos
3696.44 3696.40	3696.50	925.11 (4+) 740.28 (5+)	HexNAc2-Man4-Phos
3858.56	3858.55	965.64 (4+)	HexNAc2-Man5-Phos
4020.60	4020.61	1006.15 (4+)	HexNAc2-Man6-Phos
4182.68	4182.66	1046.67 (4+)	HexNAc2-Man7-Phos
4344.64	4344.71	1087.16 (4+)	HexNAc2-Man8-Phos
4506.64	4506.76	1127.67 (4+)	HexNAc2-Man9-Phos
4668.64	4668.82	1168.16 (4+)	HexNAc2-Man10-Phos
4830.64	4830.87	1208.66 (4+)	HexNAc2-Man11-Phos
4992.84	4992.92	1249.21 (4+)	HexNAc2-Man12-Phos
5155.08	5154.97	1289.77 (4+)	HexNAc2-Man13-Phos
5317.16	5317.03	1330.29 (4+)	HexNAc2-Man14-Phos
5479.00	5479.08	1370.75 (4+)	HexNAc2-Man15-Phos
5641.16	5641.13	1411.29 (4+)	HexNAc2-Man16-Phos

Mono-phosphorylated N-glycan compositions (LC elution time ≈ 9.7 min) observed for the glycopeptide GGH_6-333-TNLCPFGEVFNATR-346 obtained by Q-TOF MS analysis of trypsin digested SARS-CoV-2 RBD protein expressed in P. pastoris yeast cells. Main signals corresponding to monoisotopic glycopeptide masses have been included in the table and do not represent and exhaustive list. Some signals arise from in-source fragmentation.

Table S9 Di-phosphorylated N-glycan compositions of yRBD.

Mass observed (neutral)	Mass predicted	Ions observed	Glycan composition
3938.48 3938.50 3938.48	3938.52	985.62 (+4) 788.70 (5+) 657.41 (6+)	HexNAc2-Man5-Phos2
4100.64 4100.60 4100.52	4100.57	1026.16 (4+) 821.12 (5+) 684.42 (6+)	HexNAc2-Man6-Phos2
4262.60 4262.55	4262.62	1066.65 (4+) 853.51 (5+)	HexNAc2-Man7-Phos2
4424.60 4424.70	4424.68	1107.15 (4+) 885.94 (5+)	HexNAc2-Man8-Phos2
4586.76 4586.70	4586.73	1147.69 (4+) 918.35 (5+)	HexNAc2-Man9-Phos2
4748.72 4748.70	4748.78	1188.18 (4+) 950.74 (5+)	HexNAc2-Man10-Phos2
4910.80 4910.85	4910.84	1228.70 (4+) 983.17 (5+)	HexNAc2-Man11-Phos2
5072.88 5072.80	5072.89	1269.22 (4+) 1015.56 (5+)	HexNAc2-Man12-Phos2
5234.96 5234.85	5234.94	1309.74 (4+) 1047.97 (5+)	HexNAc2-Man13-Phos2
5396.92 5396.95	5396.99	1350.23 (4+) 1080.39 (5+)	HexNAc2-Man14-Phos2
5558.92 5559.15	5559.05	1390.73 (4+) 1112.83 (5+)	HexNAc2-Man15-Phos2
5721.00 5721.00	5721.10	1431.25 (4+) 1145.20 (5+)	HexNAc2-Man16-Phos2
5883.15	5883.15	1177.63 (5+)	HexNAc2-Man17-Phos2
6045.10	6045.21	1210.02 (5+)	HexNAc2-Man18-Phos2

Di-phosphorylated N-glycan compositions (LC elution time ≈ 10.8 min) observed for the glycopeptide GGH_6-333-TNLCPFGEVFNATR-346 obtained by Q-TOF MS analysis of trypsin digested SARS-CoV-2 RBD protein expressed in *P. pastoris* yeast cells. Main signals corresponding to monoisotopic glycopeptide masses have been included in the table and do not represent and exhaustive list. Some signals arise from in-source fragmentation.

Table S10. Normalized neutralization activity of pooled sera (PVNT).

Test Group	Serum Dilution		
	NT(50)	NT(70)	NT(90)
Qβ VLP only	-	-	-
yRBD	-	-	-
hRBD	41	-	-
hRBD:6AGDL	322	155	102
yRBD:VLP (NHS)	157	66	42
hRBD:VLP (NHS)	365	119	71
yRBD:VLP (6AGDL)	780	256	124
hRBD:VLP (6AGDL)	1302	431	170

Pooled serum samples from each immunized group were pre-incubated with SARS-CoV-2 spike pseudotyped MLV and used for infection of HEK293T-ACE2 cells. The percentage of infected cells (expressing GFP reporter) was quantified by flow cytometry and the normalized neutralization was calculated as a ratio from infection control (without serum). The neutralization percentage curve was analyzed using nonlinear regression function and the neutralization titres (NT50, 70, and 90) are represented as the serum dilutions resulting in 50, 70 and 90% neutralization.

Table S11. Viral load fold-reduction as measured by relative quantitation of RNA of viral against a positive control ($2^{-\Delta\Delta Ct}$).

Mouse	VLP only	yRBD only	hRBD only	hRBD:6AGDL	yRBD:VLP (NHS)	hRBD:VLP (NHS)	yRBD:VLP (6AGDL)	hRBD:VLP (6AGDL)
1	0.9	0.8	3428.4	19039.0	20.1	25649.9	26068.1	42939.0
2	1.0	1.0	8.7	42741.0	26.4	47643.8	47424.1	20834.2
3	2.1	0.7	44.6	9410.1	55749.3	1231.9	34317.7	3884.0
4	2.4	2.6	1.1	13969.6	13745.5	95067.6	13369.6	32466.6
5	1.2	0.6	44453.2	25296.8	21870.0	15716.6	32093.6	160624.6
6	1.1	2.8	2.3	6.4	52742.1	56397.1	80312.3	19083.0
7	0.8	3.1	9.3	8.6	62144.3	13216.0	14596.5	40716.7
8	0.7	0.8	370.5	46.1	28857.7	21819.5	12302.5	12707.0

Table S12. C_q data used to calculate values in Table S11.

	VLP only								yRBD only								hRBD only								hRBD:6AGDL							
Mouse	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Read 1	13.34	13.19	14.90	14.77	13.94	13.92	13.51	13.20	13.42	13.77	12.21	15.19	12.79	15.28	15.32	13.03	25.52	16.78	19.32	15.28		15.00	18.46	22.25	28.18	29.38	27.03	27.66	28.38	16.52	16.79	19.19
Read 2	13.86	14.11	14.75	15.25	14.13	13.89	13.73	13.46	13.24	13.96	13.94	15.33	12.71	15.62	15.42	14.13	25.29	17.02	19.16	13.35	29.11	14.97	16.22	22.31	27.41	29.05	26.96	27.46	28.35	14.33	16.92	19.51
Read 3	13.60	14.06	15.09	15.13	14.11	14.14	13.28	13.34	13.58	13.57	13.84	15.08	13.40	15.03	15.57	13.21	25.85	16.99	19.39	13.23	29.39	15.07	16.42	22.47	28.49	29.15	27.04	27.62	28.58	18.59	17.04	19.31
Median	13.60	14.06	14.90	15.13	14.11	13.92	13.51	13.34	13.42	13.77	13.84	15.19	12.79	15.28	15.42	13.21	25.52	16.99	19.32	13.35	29.25	15.00	16.42	22.31	28.18	29.15	27.03	27.62	28.38	16.52	16.92	19.31
Average	13.60	13.79	14.91	15.05	14.06	13.98	13.51	13.33	13.41	13.77	13.33	15.20	12.97	15.31	15.44	13.46	25.55	16.93	19.29	13.95	29.25	15.01	17.03	22.34	28.03	29.19	27.01	27.58	28.44	16.48	16.92	19.34
	yRBD:VLP (NHS)								hRBD:VLP (NHS)								yRBD:VLP (6AGDL)								hRBD:VLP (6AGDL)							
Mouse	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Read 1	18.03	18.51	29.49	26.63	28.26	30.67	29.66	27.06	29.14	29.28	24.70	30.25	27.80	29.75	26.71	27.89	27.73	28.78	29.12	27.42	28.36	29.85	27.45	28.01	29.04	27.53	25.80	28.57	30.59	27.49	28.83	30.26
Read 2	18.24	18.56	29.04	27.73	27.69	28.81	29.51	29.38	28.46	29.25	22.32	30.17	27.57	29.44	28.04	28.16	29.03	29.66	29.42	27.53	29.26	29.78	27.55	26.91	29.08	28.01	25.42	28.47	31.32	27.92	29.40	22.60
Read 3	18.15		30.20	28.31	28.73	29.01	30.03	29.44	27.77	29.52	25.21	30.62	27.88	29.59	27.75	28.62	28.68	29.59	28.09	27.60	28.72	30.68	27.93	27.27	29.48	28.93	25.98	29.35	31.40	28.68	29.14	29.47
Median	18.15	18.54	29.49	27.73	28.26	29.01	29.66	29.38	28.46	29.28	24.70	30.25	27.80	29.59	27.75	28.16	28.68	29.59	29.12	27.53	28.72	29.85	27.55	27.27	29.08	28.01	25.80	28.57	31.32	27.92	29.14	29.47
Average	18.14	18.54	29.58	27.56	28.23	29.50	29.73	28.63	28.46	29.35	24.08	30.35	27.75	29.59	27.50	28.22	28.48	29.34	28.88	27.52	28.78	30.10	27.64	27.40	29.20	28.16	25.73	28.80	31.10	28.03	29.12	27.44

The C_q for infected Vero E6 cells was 13.81.

Table S13 Overview of multimeric RBD-vaccine candidates.

RBD (range, SARS-COV-2 unless otherwise indicated)	**Antigen Purification**	**Multimerisation strategy**	**Carrier**	**Reference**
319-541	IMAC (C-terminal H_6-Tag)	SMPH (Succinimidyl 6-((beta-maleimidopropionamido)hexanoate))	$CuMV_{TT}$	[120]
319-591	C-tag (alpha-synuclein)	Plug-and-Display (SC-Ag) Plug-and-Display (Ag-SC)	SpyTag-AP205	[121]
spike	Protein A (single-chain Fc)	Plug-and-Display (SC-Ag)	LuS-N71-SpyTag Ferritin-N96-SpyTag	[122]
331–529	Spy&Go, size-exclusion	Plug-and-Display (ST003-Ag)	SpyCatcher003-mi3	[39]
330-532	CR3022 Ab, size exclusion	Plug-and-Display (Ag-ST), genetic fusion	SpyCatcher-Ferritin SpyCatcher-I3-01v9 SpyCatcher-I3-01v9-LD7-PADRE SApNPs	[123]
319–541	IMAC, size-exclusion	Plug-and-Display (Ag-ST)	$GSSH_6$-SpyCatcher-Ferritin $GSSH_6$-ΔN1-SpyCatcher-mi3 SpyCatcher-I53–50A1.1PT1-H_6 SpyCatcher-I53-50B.4PT1-H_6	[124]
328-531	IMAC, size exclusion	Genetic fusion	I53-50	[125]
331–529	Spy&Go, size-exclusion	Plug-and-Display (ST003-Ag)	SpyCatcher003-mi3	[126]
331-527	C-tag (alpha-synuclein) & Fc (EDC only)	Plug-and-Display (ST003-Ag), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC)	SpyCatcher-mi3 mcKLH (EDC)	[127]
Mosaic SARS-COV-2: 319-539 RaTG13-CoV: 319-541 SHC014-CoV: 307-524 Rs4081-CoV: 310-515 Pangolin17-CoV: 317-539 RmYN02-CoV: 298-503 Rf1-CoV:310-515 W1V1-CoV: 307-528 Yun11-CoV: 310-515 BM-4831-CoV: 310-530 BtkY72-CoV: 309-530	IMAC (C-terminal H_6-Tag)	Plug-and-Display (Ag-ST003)	$GSSH_6$-SpyCatcher003-mi3	[40]
331-526	IMAC, size-exclusion	6AGDL (SPAAC)	Qß	*This study*

Table S14 General comparison of N-terminal targeting strategies.

Method	**Pro**	**Con**
Reductive amination/alkylation	Selective for most N-termini N-terminal charge (+) is retained Availability of aldehyde reagents No need for protein engineering	Low specificity Moderate yields Long incubation times Needs reducing agent → possible disulphide shuffling
Oxidative coupling (Ser/Thr oxidation)	Selective for specific N-termini Minimal genetic engineering (1 amino acid) Demonstrated on mAbs	Need to use periodate (can modify other residues, glycans etc.)
Oxidative coupling (Pro + *o*-aminophenols)	Short reaction time	Low yields Requires cysteine protection/deprotection
Aromatic aldehydes (2PCA)	Highly selective for N-terminus Moderate to good yields Demonstrated on Abs	High temperature Need for organic solvent Long reaction time Protein denaturation sometimes observed Bond not stable

Aromatic aldehydes (2EBA)	Bond stability demonstrated over 12 h at 37 ºC, unclear long-term stability. Can yield high labelling for some substrates	High temperature (37 ºC) Long reaction time Multiple labelling Low yield (~30% after 16 h on protein) 10% organic solvent
Aromatic aldehydes (PLP) maybe group all aromatic aldhydes into one?		High temp Moderate yields
Direct azidation (imidazole-1-sulfonyl azide)		Long RT (O/N) Moderate yields
N-terminal cysteine approaches	~complete labelling N-terminus selective	Requires genetic engineering Cysteine reshuffling/dimer formation etc; Reducing agent may be needed to deprotect Cys.
Ketene	Quite N-terminus specific	Low labelling yield, High temp Long RT
Global AA replacement		Incorporation at multiple sites possible if multiple Met present in sequence. Competition with native amino acid(s) gives heterogeneity. Possible misincorporation at other residue positions.
Genetic code expansion	Full control over positioning across polypeptide chain.	Requires specialist expression host. Requires specialist amino acids.
6AGDL	Selective for specific N-termini, including commonly used HisTags Very good yield (>90%) Highly hydrophilic linker (no solvent required) Gentle labelling conditions (pH 7.5, 4-37 ºC) Scope shown: peptides, proteins, Abs & VLPs Monitoring 200 Da label possible with mP-AGE Primary purification mode (metal affinity) Secondary purification mode (borate affinity) Controlled reversibility	Genetic engineering is needed to install HisTag High excess (100mM) of label Reversibility may be undesirable

Supporting References

[80] F. Cramer, H. Otterbach, H. Springmann, *Chem. Ber.* **1959**, *92*, 384–391.

[81] S. Hanessian, *J. Org. Chem* **1969**, *34*, 675–681.

[82] S. Hanessian, D. Ducharme, R. Massé, M. L. Capmau, *Carbohydr. Res.* **1978**, *63*, 265–269.

[83] G. W. Fleet, N. G. Ramsden, D. R. Witty, *Tetrahedron* **1989**, *45*, 327–336.

[84] M. Juricek, P. Gruz, J. Veleminsky, J. Stanek, K. Kefurt, J. Moravcova, J. Jary, *Mutat. Res.* **1991**, *251*, 13–20.

[85] P. Wang, G. J. Shen, Y. F. Wang, Y. Ichikawa, C. H. Wong, *J. Org. Chem.* **1993**, *58*, 3985–3990.

[86] J. L. Jiménez Blanco, P. Bootello, C. Ortiz Mellet, R. Gutiérrez Gallego, J. M. García Fernández, *Chem. Commun.* **2004**, 92–93.

[87] L. Chaveriat, I. Stasik, G. Demailly, D. Beaupère, *Tetrahedron: Asymmetry* **2006**, *17*, 1349–1354.

[88] N. Darabedian, J. Gao, K. N. Chuh, C. M. Woo, M. R. Pratt, *J. Am. Chem. Soc.* **2018**, *140*, 7092–7100.

[89] S. K. Sinha, K. Brew, *Carbohydr. Res.* **1980**, *81*, 239–247.

[90] A. Liptak, I. Jodal, J. Harangi, P. Nanasi, *ACH - Models Chem.* **1983**, *113*, 415–422.

[91] W. A. Greenberg, E. S. Priestley, P. S. Sears, P. B. Alper, C. Rosenbohm, M. Hendrix, S.-C. Hung, C.-H. Wong, *J. Am. Chem. Soc.* **1999**, *121*, 6527–6541.

[92] P. R. Sleath, A. L. Handlon, N. J. Oppenheimer, *J. Org. Chem.* **1991**, *56*, 3608–3613.

[93] P. H. Seeberger, S. Roehrig, P. Schell, Y. Wang, W. J. Christ, *Carbohydr. Res.* **2000**, *328*, 61–69.

[94] S. Rajagopal, S. Vancheesan, J. Rajaram, J. C. Kuriacose, *J. Mol. Catal.* **1983**, *22*, 137–144.

[95] Youval. Shvo, Dorotha. Czarkie, Yocheved. Rahamim, D. F. Chodosh, *J. Am. Chem. Soc.* **1986**, *108*, 7400–7402.

[96] S. Hanessian, *Chem. Ind.* **1966**, 2126–2127.

[97] K. Kefurt, K. Čapek, Z. Kefurtová, J. Jarý, *Collect. Czech. Chem. Commun.* **1979**, *44*, 2526–2535.

[98] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.

[99] Y. Qu, F.-X. Sauvage, G. Clavier, F. Miomandre, P. Audebert, *Angew. Chem. Int. Ed.* **2018**, *57*, 12057–12061.

[100] K. A. Horner, N. M. Valette, M. E. Webb, *Chem. Eur. J.* **2015**, *21*, 14376–14381.

[101] C. A. DeForest, D. A. Tirrell, *Nat. Mater.* **2015**, *14*, 523–531.

[102] J. Dommerholt, S. Schmidt, R. Temming, L. J. A. Hendriks, F. P. J. T. Rutjes, J. C. M. van Hest, D. J. Lefeber, P. Friedl, F. L. van Delft, *Angew. Chem. Int. Ed.* **2010**, *49*, 9422–9425.

[103] M. Baalmann, L. Neises, S. Bitsch, H. Schneider, L. Deweid, P. Werther, N. Ilkenhans, M. Wolfring, M. J. Ziegler, J. Wilhelm, H. Kolmar, R. Wombacher, *Angew. Chem. Int. Ed.* **2020**, *59*, 12885–12893.

[104] D. Gibson, *Protocol Exchange* **2009**, DOI 10.1038/nprot.2009.77.

[105] F. W. Studier, *Protein Expression Purif.* **2005**, *41*, 207–234.

[106] A. Dell, H. R. Morris, *Science* **2001**, *291*, 2351–2356.

[107] M. Panico, L. Bouché, D. Binet, M.-J. O'Connor, D. Rahman, P.-C. Pang, K. Canis, S. J. North, R. C. Desrosiers, E. Chertova, B. F. Keele, J. W. Bess, J. D. Lifson, S. M. Haslam, A. Dell, H. R. Morris, *Sci. Rep.* **2016**, *6*, 32956.

[108] B. Bartosch, J. Dubuisson, F.-L. Cosset, *J. Exp. Med.* **2003**, *197*, 633–642.

[109] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, F. Li, *Nature* **2020**, *581*, 221–224.

[110] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, *Nature Methods* **2012**, *9*, 671–675.

[111] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, *Nat. Methods* **2012**, *9*, 676–682.

[112] J. O. Fierer, G. Veggiani, M. Howarth, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E1176–E1181.

[113] L. Li, J. O. Fierer, T. A. Rapoport, M. Howarth, *J. Mol. Biol.* **2013**, DOI 10.1016/j.jmb.2013.10.021.

[114] W. L. A. Brooks, B. S. Sumerlin, *Chem. Rev.* **2016**, *116*, 1375–1397.

[115] Argentinian AntiCovid Consortium, *Sci. Rep.* **2020**, *10*, 21779.

[116] S. Geissler, P. Boniforte, J. Plaschke, M. Weigandt, S. Jaekel, R. Kellner, T. Rysiok, D. Mueller-Pompalla, K. Hance, *Survivin-Directed Cancer Vaccine Therapy*, **2015**, WO2015090572A1.

[117] W. Awad, G. J. M. Ler, W. Xu, A. N. Keller, J. Y. W. Mak, X. Y. Lim, L. Liu, S. B. G. Eckle, J. Le Nours, J. McCluskey, A. J. Corbett, D. P. Fairlie, J. Rossjohn, *Nat. Immunol.* **2020**, *21*, 400–411.

[118] D. G. Smyth, W. H. Stein, S. Moore, *J. Biol. Chem.* **1962**, *237*, 1845–1850.

[119] Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, S. Gräslund, P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S.-H. Kim, Z. Rao, Y. Shi, T. C. Terwilliger, C.-Y. Kim, L.-W. Hung, G. S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, R. C. Stevens, S. A. Lesley, I. A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S. K. Burley, J. S. Emtage, J. M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, D. I. Stuart, R. J. Owens, S. Daenke, A. Schütz, U. Heinemann, S. Yokoyama, K. Büssow, K. C. Gunsalus, *Nat. Methods* **2008**, *5*, 135–146.

[120] L. Zha, X. Chang, H. Zhao, M. O. Mohsen, L. Hong, Y. Zhou, H. Chen, X. Liu, J. Zhang, D. Li, K. Wu, B. Martina, J. Wang, M. Vogel, M. F. Bachmann, *Vaccines* **2021**, *9*, 395.

[121] C. Fougeroux, L. Goksøyr, M. Idorn, V. Soroka, S. K. Myeni, R. Dagil, C. M. Janitzek, M. Søgaard, K.-L. Aves, E. W. Horsted, S. M. Erdoğan, T. Gustavsson, J. Dorosz, S. Clemmensen, L. Fredsgaard, S. Thrane, E. E. Vidal-Calvo, P. Khalifé, T. M. Hulen, S. Choudhary, M. Theisen, S. K. Singh, A. Garcia-Senosiain, L. Van Oosten, G. Pijlman, B. Hierzberger, T. Domeyer, B. W. Nalewajek, A. Strøbæk, M. Skrzypczak, L. F. Andersson, S. Buus, A. S. Buus, J. P. Christensen, T. J. Dalebout, K. Iversen, L. H. Harritshøj, B. Mordmüller, H. Ullum, L. S. Reinert, W. A. de Jongh, M. Kikkert, S. R. Paludan, T. G. Theander, M. A. Nielsen, A. Salanti, A. F. Sander, *Nat Commun* **2021**, *12*, 324.

[122] B. Zhang, C. W. Chao, Y. Tsybovsky, O. M. Abiona, G. B. Hutchinson, J. I. Moliva, A. S. Olia, A. Pegu, E. Phung, G. B. E. Stewart-Jones, R. Verardi, L. Wang, S. Wang, A. Werner, E. S. Yang, C. Yap, T. Zhou, J. R. Mascola, N. J. Sullivan, B. S. Graham, K. S. Corbett, P. D. Kwong, *Sci Rep* **2020**, *10*, 18149.

[123] L. He, X. Lin, Y. Wang, C. Abraham, C. Sou, T. Ngo, Y. Zhang, I. A. Wilson, J. Zhu, *Sci. Adv.* **2021**, *7*, eabf1591.

[124] Y.-F. Kang, C. Sun, Z. Zhuang, R.-Y. Yuan, Q. Zheng, J.-P. Li, P.-P. Zhou, X.-C. Chen, Z. Liu, X. Zhang, X.-H. Yu, X.-W. Kong, Q.-Y. Zhu, Q. Zhong, M. Xu, N.-S. Zhong, Y.-X. Zeng, G.-K. Feng, C. Ke, J.-C. Zhao, M.-S. Zeng, *ACS Nano* **2021**, *15*, 2738–2752.

[125] A. C. Walls, B. Fiala, A. Schäfer, S. Wrenn, M. N. Pham, M. Murphy, L. V. Tse, L. Shehata, M. A. O'Connor, C. Chen, M. J. Navarro, M. C. Miranda, D. Pettie, R. Ravichandran, J. C. Kraft, C. Ogohara, A. Palser, S. Chalk, E.-C. Lee, K. Guerriero, E. Kepl, C. M. Chow, C. Sydeman, E. A. Hodge, B. Brown, J. T. Fuller, K. H. Dinnon, L. E. Gralinski, S. R. Leist, K. L. Gully, T. B. Lewis, M. Guttman, H. Y. Chu, K. K. Lee, D. H. Fuller, R. S. Baric, P. Kellam, L. Carter, M. Pepper, T. P. Sheahan, D. Veesler, N. P. King, *Cell* **2020**, *183*, 1367-1382.e17.

[126] T. K. Tan, P. Rijal, R. Rahikainen, A. H. Keeble, L. Schimanski, S. Hussain, R. Harvey, J. W. P. Hayes, J. C. Edwards, R. K. McLean, V. Martini, M. Pedrera, N. Thakur, C. Conceicao, I. Dietrich, H. Shelton, A. Ludi, G. Wilsden, C. Browning, A. K. Zagrajek, D. Bialy, S. Bhat, P. Stevenson-Leggett, P. Hollinghurst, M. Tully, K. Moffat, C. Chiu, R. Waters, A. Gray, M. Azhar, V. Mioulet, J. Newman, A. S. Asfor, A. Burman, S. Crossley, J. A. Hammond, E. Tchilian, B. Charleston, D. Bailey, T. J. Tuthill, S. P. Graham, H. M. E. Duyvesteyn, T. Malinauskas, J. Huo, J. A. Tree, K. R. Buttigieg, R. J. Owens, M. W. Carroll, R. S. Daniels, J. W. McCauley, D. I. Stuart, K.-Y. A. Huang, M. Howarth, A. R. Townsend, *Nat. Commun.* **2021**, *12*, 542.

[127] B. D. Quinlan, W. He, H. Mou, L. Zhang, Y. Guo, J. Chang, S. Peng, A. Ojha, R. Tavora, M. S. Parcells, G. Luo, W. Li, G. Zhong, H. Choe, M. Farzan, *An Engineered Receptor-Binding Domain Improves the Immunogenicity of Multivalent SARS-CoV-2 Vaccines*, **2020**.