Supplementary Material (ESI) for Chemical Communications

Analysis of the dispersity in carbohydrate loading of synthetic glycoproteins using MALDI-TOF mass spectrometry

Mitul K. Patel, Balakumar Vijayakrishnan, Julia R. Koeppe, Justin M. Chalker, Katie J. Doores and Benjamin G. Davis*

Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford, OX1 3TA, UK.

ben.davis@chem.ox.ac.uk

Supplementary Information

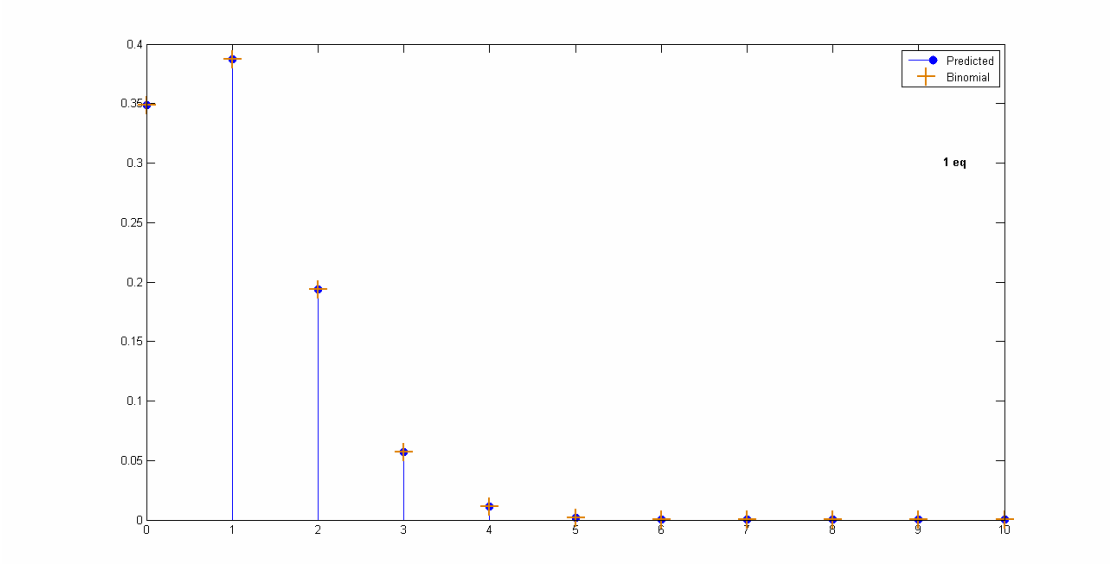
Contents

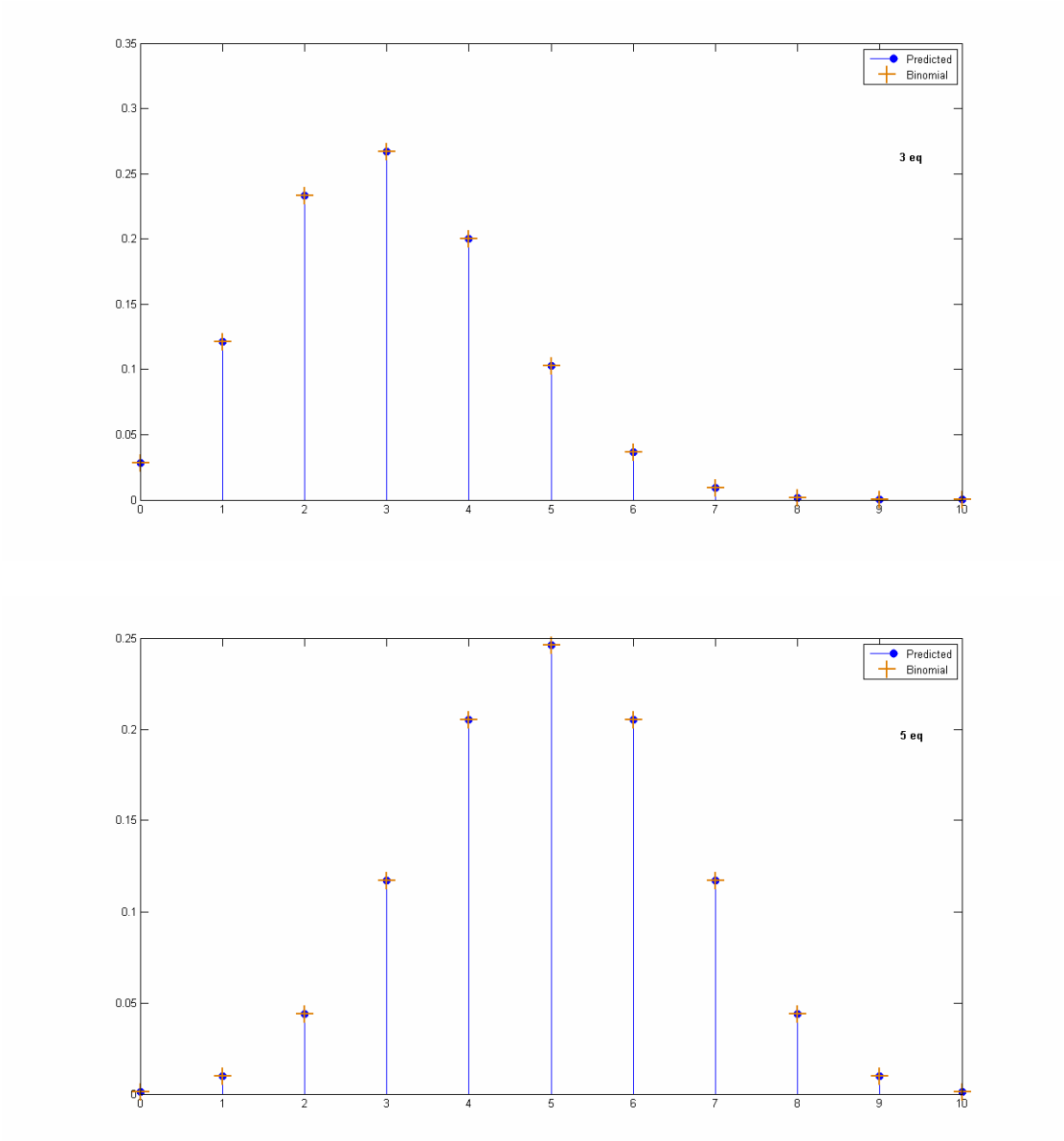
1. Discussion on the Gaussian approximation

A protein modification reaction can be modeled as a series of competitive consecutive second order reactions of the type:

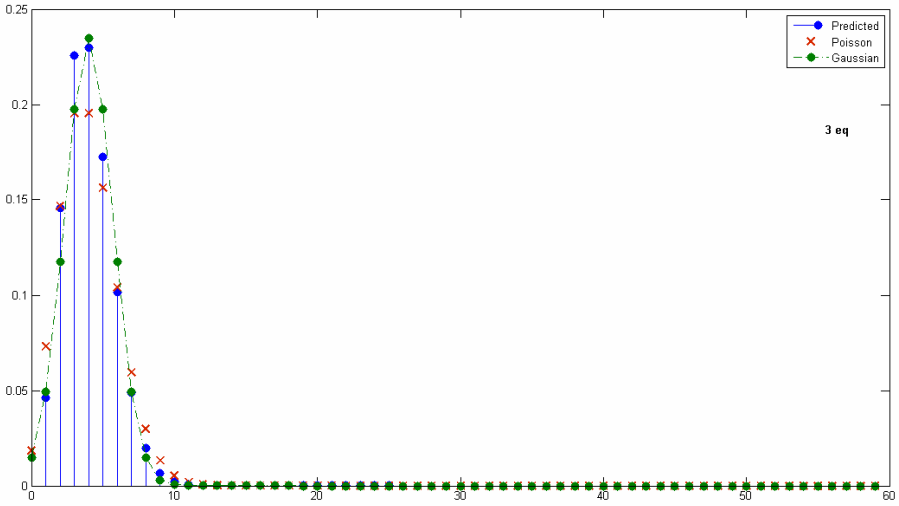
$$A + R \xrightarrow{k_1} B$$
$$B + R \xrightarrow{k_2} C$$
$$C + R \xrightarrow{k_3} D$$
$$etc...$$

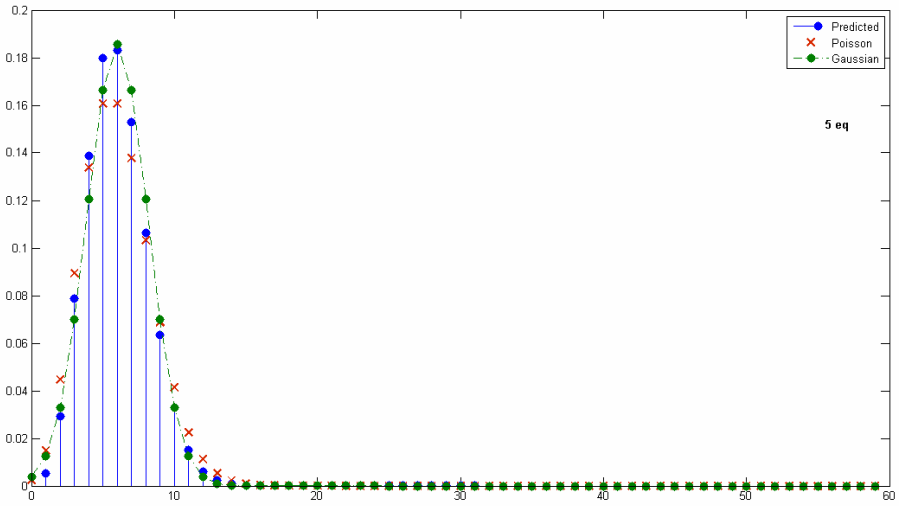
The analytical solution for the product distribution of such a reaction sequence has been obtained previously.[1] For a protein modification reaction, the consecutive reaction rate constants fall in proportion with the number of available reactive sites. In this particular case, the product distribution is *exactly* given by a binomial probability distribution. To illustrate this, the theoretical product distribution (calculated by computationally solving the rate equations for the above reactions) for a protein with 10 reactive sites treated with various equivalents of a reagent is shown below, and compared with the binomial distribution function.

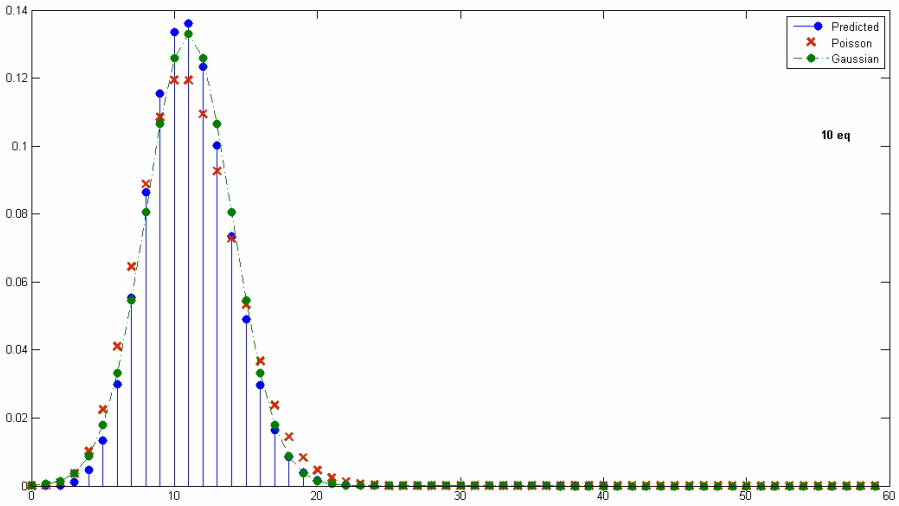


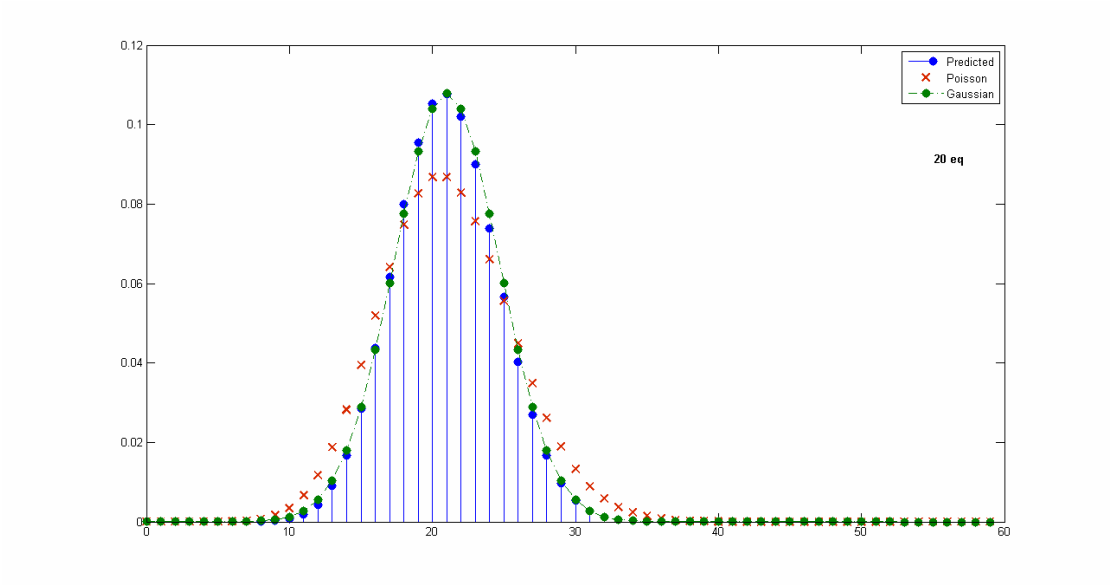


This theoretical product distribution can be approximated by a Gaussian distribution (after applying a continuity correction) or a Poisson distribution, as demonstrated below for a protein with 60 reactive sites (BSA- 59 lysines and the *N*-terminus).


Predicted
Poisson
Gaussian
3 eq


Predicted
Poisson
Gaussian
5 eq

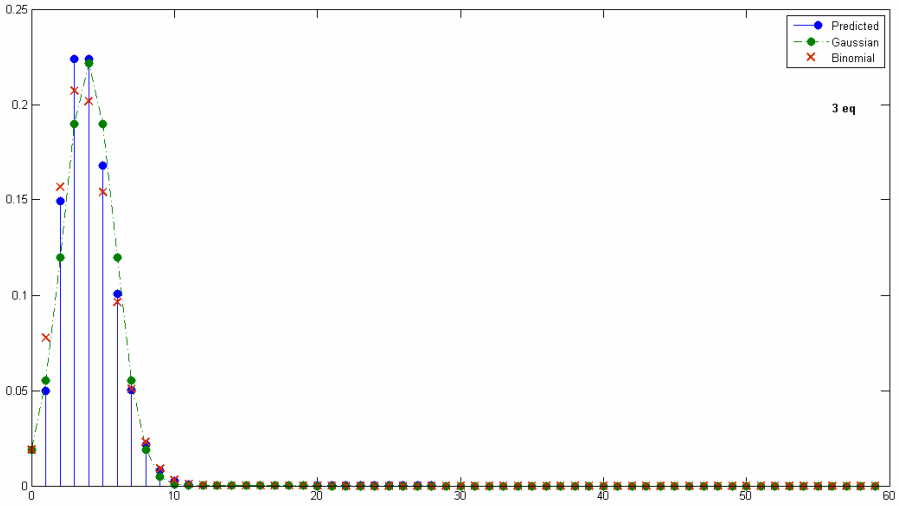

Predicted
Poisson
Gaussian
10 eq

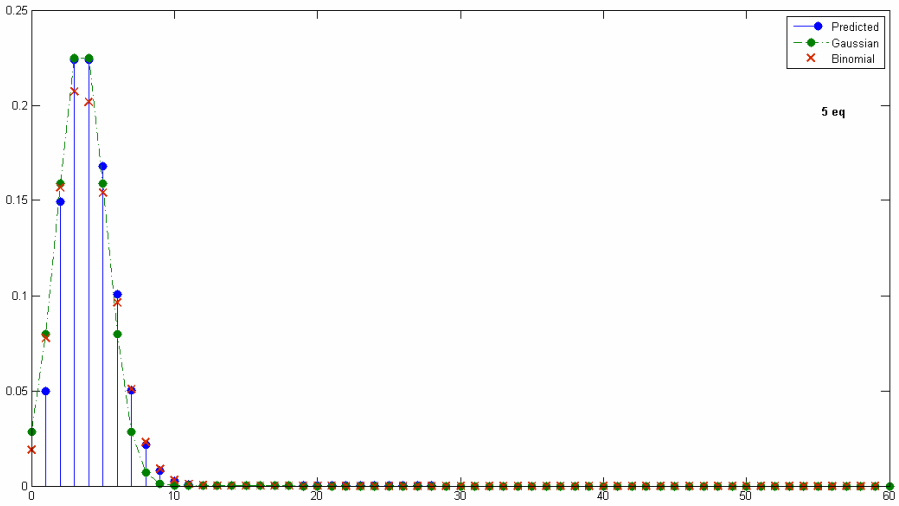


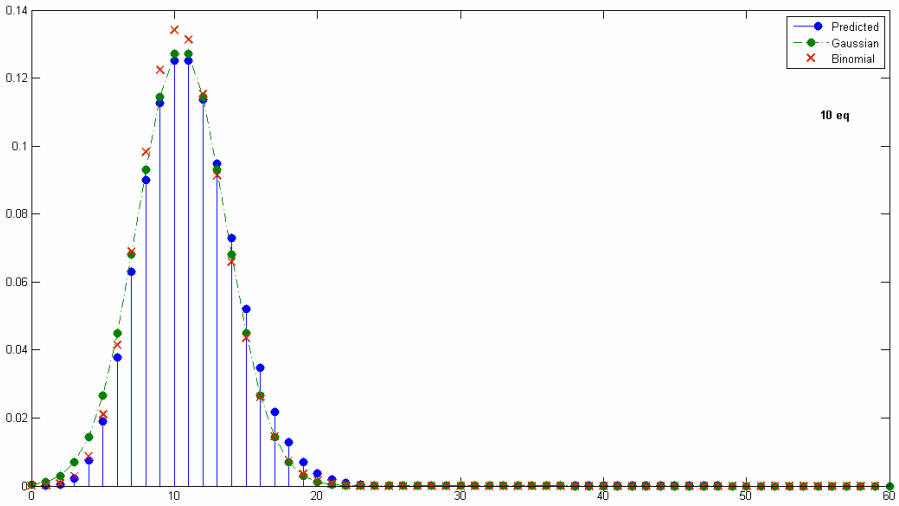
As can be seen, the Gaussian distribution offers a very good approximation to the theoretical product distribution. Crucially, the dispersity is well represented. Even at very low modification levels (3 equivalents, well below those found experimentally) there is very little asymmetry, therefore the Gaussian distribution is still valid.

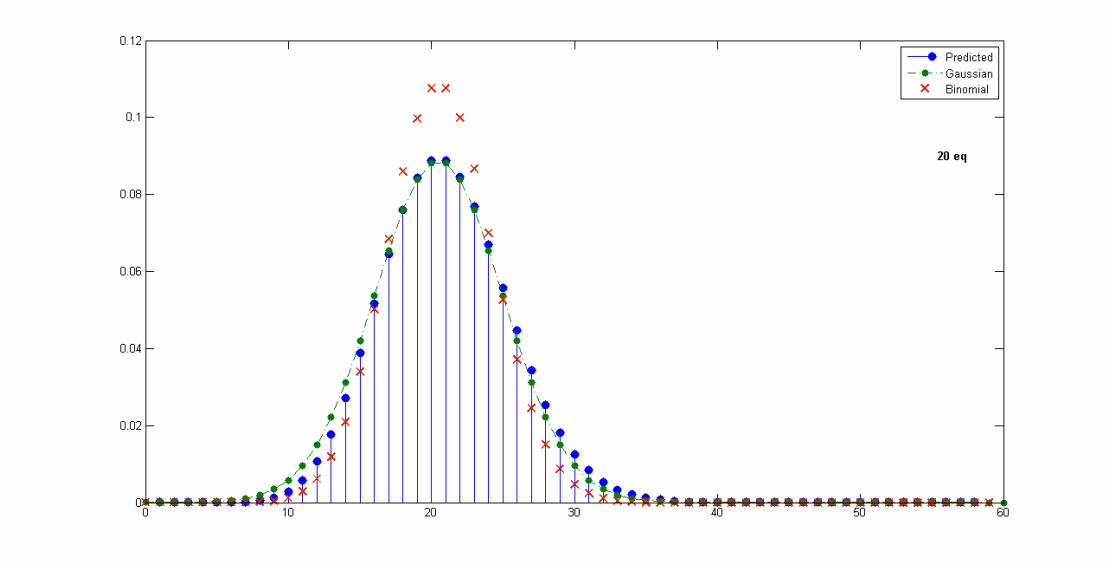
As expected, with a low concentration of reagent (3 equivalents, i.e. when a modification event can be considered rare), a Poisson distribution also offers a good approximation. However, at higher modifications (20 equivalents) the Poisson distribution shows significant deviations and no longer offers a good approximation.

Real proteins, however, do not necessarily follow this ideal model which assumes the rate constants are solely dependent on the number of free lysines and neglects structural effects. Whereas for the ideal case the binomial distribution offered an exact analytical solution, this is not the case when the rate constants are not purely defined by statistics. To demonstrate this point, consider a reaction on BSA where the consecutive rate constants are all identical, giving a theoretical hypergeometrical distribution of protein products.


Predicted
Gaussian
Binomial
3 eq


Predicted
Gaussian
Binomial
5 eq


Predicted
Gaussian
Binomial
10 eq



As before, the Gaussian distribution offers a very good approximation to the actual product distribution and faithfully reproduces the dispersity. Even at low modification levels, asymmetry is not observed.

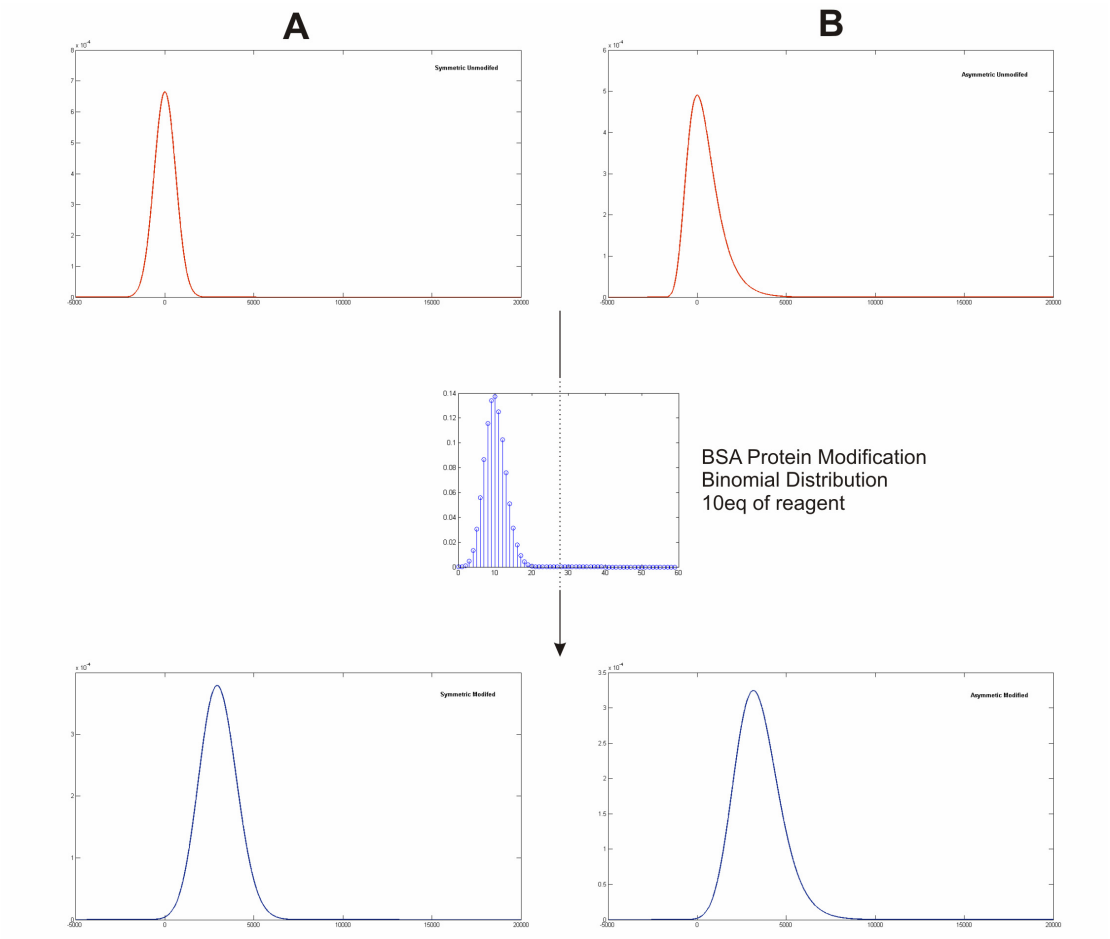
The binomial distribution work wells at low modification levels. However, at higher concentrations (20 equivalents), the binomial distribution is completely inadequate.

The Gaussian distribution offers a very good approximation to the product distribution of a series of competitive consecutive second order events, under a range of criteria for the rate constants. This arises because of the mathematical flexibility offered by the Gaussian function- both the mean and peak width (variance) can be independently set through μ and σ, and the function can be easily modified to represent the underlying distribution. This is not the case for the Poisson or binomial distributions- the mean and variance are intrinsically linked to each other and so these distributions are not always suitable. Additionally, we have demonstrated that even at low modification levels, the asymmetry of the product distribution is insignificant and the Gaussian approximation is still valid. The broad applicability and the intuitive

description of the peak-width through the standard deviation (σ) make the Gaussian distribution an ideal choice for the dispersity analysis.

2. Discussion on peak asymmetry

The MALDI mass spectrum of the unmodified protein can often exhibit some high mass peak tailing caused by in source decay-desolvation of protein-matrix cluster ions, as well as the formation of cation and matrix photochemical adducts.[2] To determine the effect of this peak tailing on the dispersity analysis, we simulated the protein modification reaction using both symmetric and skewed peaklets.



In case A, the unmodified protein peak is a symmetric Gaussian peak and represents the theoretical ideal on which the dispersity analysis is based. In case B, a Gumbel distribution is used to simulate high mass peak tailing, representing a more realistic

experimental outcome. In both cases, the protein modification reaction is simulated to produce a binomial distribution of products. Hence, we can say the true dispersity of the products is 5.77 (from $\sigma^2 = NP(1-P)$), which can be compared to the calculated dispersity from the broadened MALDI peaks of the modified protein.

For case A, the analysis gives a dispersity of 5.74, thus replicating the true dispersity very well and provides validation for our core model.

For case B, the peak tailing for the unmodified protein also leads to some asymmetry in the modified protein peak. If we approximate both as Gaussian functions, the analysis gives a dispersity of 5.86. Hence, even with a significant amount of peak tailing, the dispersity analysis only produces a small deviation from the true value. This is not too surprising since the peak tailing serves to increase the FWHM of *both* the unmodified and the modified protein peaks, and is therefore, to an extent, self compensating. Hence, while MALDI conditions should be screened to minimize peak asymmetry, a small amount of tailing can be tolerated.

3. General experimental considerations

Proton nuclear magnetic resonance spectra (^{1}H NMR) were recorded on a Bruker DPX400 (400 MHz), DQX400 (400 MHz) or AVC500 (500 MHz) spectrometer. Carbon nuclear magnetic resonance spectra (^{13}C NMR) were recorded on a Bruker DQX400 (101 MHz), AVC500 (126 MHz) and are proton decoupled. Spectra were assigned using COSY, DEPT-135, HMQC, HSQC, and HMBC if required. All chemical shifts are quoted on the δ scale in ppm using residual solvent as an internal standard.

Low resolution mass spectra were recorded on a Micromass Platform 1 spectrometer using electrospray ionization (ES), or on a Bruker Daltronic MicroTOF spectrometer.

High resolution mass spectra were recorded on a Bruker Daltronic MicroTOF spectrometer. *m*/*z* values are reported in Daltons. Infrared spectra (FT-IR) were recorded on a Bruker Tensor 27 Fourier Transform spectrophotometer as a thin film on NaCl plates. Absorption maxima are reported in wavenumbers (cm^{-1}). Only signals representing functional groups are reported; the fingerprint region is not listed. Optical rotations were measured on a Perkin-Elmer 241 polarimeter at 589nm (Na D-line) with a path length of 1.0 dm and are reported in units of deg $dm^{-1}cm^3g^{-1}$. Concentrations are given in g/100 mL.

Thin Layer Chromatography (TLC) was carried out using Merck aluminium backed sheets coated with Kieselgel $60F_{254}$ silica gel. Visualization of the sheets was achieved using a UV lamp (λ_{max} = 254 or 365 nm) and/or ammonium molybdate (5% in 2M H_2SO_4), or sulfuric acid (0.2M in 1 MeOH : 1 H_2O). Silica gel chromatography was carried out using Fluka Kieselgel 60 220-240 mesh silica.

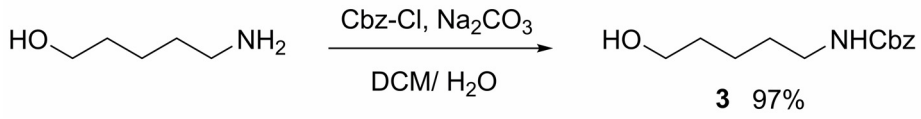
Dowex ion exchange resin 50WX8 was conditioned by washing with methanol, water, 1M HCl, and water until the filtrate was neutral.

Anhydrous THF and DCM were dried under pressure through a column of alumina. Other anhydrous solvents were purchased from Fluka and stored under Argon over molecular sieves. All other solvents were used as supplied (analytical or HPLC grade). “Petrol” refers to the fraction of light petroleum ether boiling in the range 40-60°C. “Brine” refers to a saturated aqueous solution of sodium chloride.

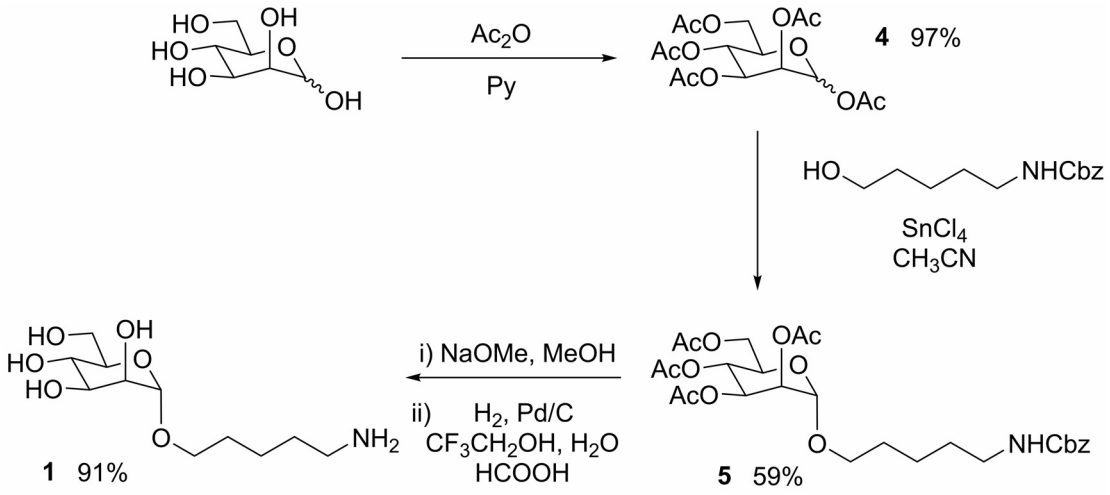
Bovine serum albumin (BSA) was purchased from Sigma Aldrich.

All other reagents were purchased from Fisher Scientific or Sigma Aldrich.

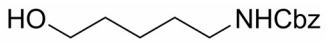
4. Reaction protocols and characterization data



Scheme S1



Scheme S2

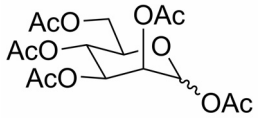


5-Aminobenzyloxycarbonyl pentanol (3)

To a solution of 5-aminopentanol (3.10 g, 30.0 mmol, 1 eq) in DCM/H_2O (100 mL, 1:1) was added benzyloxycarbonyl chloride (6.40 g, 37.5 mmol, 1.25 eq) and sodium carbonate (7.95 g, 75 mmol, 2.5 eq). After stirring at RT for 4 hours, the reaction mixture was diluted with DCM and H_2O and the organic layer was separated and washed with brine, dried over Na_2SO_4, filtered and concentrated. The residue was purified by flash column chromatography (40 – 60 % EtOAc/Petrol) to yield the titled compound (6.93g, 97 %) as a white solid.

R_f 0.29 (50% EtOAc/Petrol); ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.34 – 7.27 (m, 5H, Ph), 5.10 (s, 2H, CH_2Ph of Cbz), 4.84 (s, 1H, NH), 3.64 (br. s, 2H, OCH_2), 3.21 (dd, *J* = 13.1, 6.7 Hz, 2H, C*H*$_2$-NHCbz), 1.65 - 1.48 (m, 4H, OCH_2-C*H*$_2$, C*H*$_2$-CH_2-

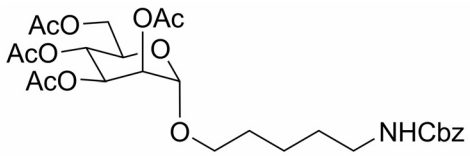
NHCbz), 1.45 - 1.35 (m, 2H, CH_2); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm 156.4 (NH*C*O), 136.5, 128.4, 128.0 (C-Ph), 66.5 (CH_2Ph), 62.5 (OCH_2), 40.8 (CH_2-NHCbz), 32.1 (OCH_2-*C*H_2), 29.7 (*C*H_2-CH_2-NHCbz), 22.8 (CH_2); HRMS (ES^+) *m/z* 260.1258 $[M + Na]^+$ (required 260.1257).



1,2,3,4,6-Penta-*O*-acetyl-α-D-mannopyranoside (4)

Acetic anhydride (524 mL, 5.55 mol, 20 eq) and DMAP (1.70 g, 0.01 mol, 0.04 eq) were added to a solution of D-mannose (50.0 g, 0.277 mol, 1 eq) in pyridine (560 mL) at 0 °C. The resultant solution was stirred overnight and the reaction mixture was allowed to warm to RT, after which cold EtOH (100 mL) was added and the reaction mixture stirred for additional one hour. The solvent was removed by co-evaporation with toluene (2 x 200 mL). The resultant residue was re-dissolved in DCM (500 mL) and washed with water (2 x 500 mL), 1M HCl (2 x 500 mL), saturated $NaHCO_3$ (2 x 500 mL), brine (1 x 500 mL) and dried over $MgSO_4$, filtered and concentrated to yield the product as a viscous oil (105.4 g, 97%) as a mixture of anomers (α:β 1:0.35).

R_f = 0.4 (50% EtOAc/Petrol); data for α anomer: ^1H NMR (400 MHz, $CDCl_3$) δ ppm 6.10 (s, 1H, H-1), 5.36 (d, *J* = 6.0 Hz, 1H, H-2), 5.31 (m, 1H, H-3), 5.27 (m, 1H, H-4), 4.29 (dd, *J* = 4.8, 12.0 Hz, 1H, H-6a), 4.16 – 4.05 (m, 2H, H-6b, H-5), 2.19 (COC*H*$_3$), 2.18 (COC*H*$_3$), 2.11 (COC*H*$_3$), 2.06 (COC*H*$_3$), 2.02 (COC*H*$_3$); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 170.7 (*C*OCH_3), 170.0 (*C*OCH_3), 169.8 (*C*OCH_3), 169.5 (*C*OCH_3), 168.1 (*C*OCH_3), 90.6 (C-1), 70.6 (C-5), 68.7 (C-2), 68.3 (C-3), 65.5 (C-4), 62.1 (C-6), 20.9 (CO*C*H_3), 20.8 (CO*C*H_3), 20.7 (CO*C*H_3), 20.6 (CO*C*H_3), 20.5 (CO*C*H_3); LRMS *m/z* (ES^+) 413.1 $[M + Na]^+$, 803.2 $[2M + Na]^+$.

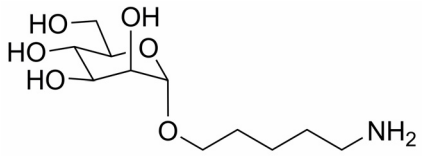


5-(Benzyloxycarbonylamino)pentyl-2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (5)

To a solution of **4** (0.80 g, 2.1 mmol, 1 eq) in CH_3CN (15 mL) was added $SnCl_4$ (0.69 g, 2.67 mmol, 1.3 eq) and the mixture stirred for 20 minutes. To the reaction mixture, **3** (0.535 g, 2.25 mmol, 1.1 eq) dissolved in dry CH_3CN (5 mL) was added dropwise and the resultant mixture was stirred for an additional 19 hours at RT. The reaction mixture was quenched by adding saturated $NaHCO_3$ solution and then filtered over a pad of Celite® and washed with hot $CHCl_3$. The filtrate was concentrated and the residue re-dissolved in DCM and then washed with brine, dried over Na_2SO_4, filtered and concentrated. The residue was purified by flash column chromatography (40 – 60 % EtOAc/Petrol) to yield the titled compound (0.681g, 59 % yield) as a colourless oil.

R_f 0.43 (50% EtOAc/Petrol); $[\alpha]_D^{25}$ +32.2 (c = 1.03, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.34 – 7.26 (m, 5H, Ph), 5.34 (dd, *J* = 10.0, 3.3 Hz, 1H, H-3), 5.27 (t, *J* 9.9 Hz, 1H, H-4), 5.23 (dd, *J* = 3.3, 1.7 Hz, 1H, H-2), 5.10 (s, 2H, C*H*$_2$-NHCbz), 4.86 (s, 1H, NH), 4.80 (d, *J* = 1.3 Hz, 1H, H-1), 4.29 (dd, *J* = 12.2, 5.34 Hz, 1H, H-6a), 4.10 (dd, *J* = 10.2, 1.9 Hz, 1H, H-6b), 3.98 (ddd, *J* = 9.4, 5.2, 2.3 Hz, 1H, H-5), 3.69 (td, *J* = 9.2, 6.5 Hz, 1H, OC*H*HCH$_2$), 3.45 (td, *J* = 9.4, 6.3 Hz, 1H, OCH*H*CH$_2$), 3.21 (dd, *J* = 13.2, 6.7 Hz, 2H, CH$_2$-NHCbz), 2.16 (s, 3H, $COCH_3$), 2.10 (s, 3H, $COCH_3$), 2.04 (s, 3H, $COCH_3$), 1.99 (s, 3H, $COCH_3$), 1.71 - 1.47 (m, 4H, OCH$_2$-C*H*$_2$, C*H*$_2$-CH$_2$-NHCbz), 1.45 - 1.33 (m, 2H, CH$_2$); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm 170.6 (*C*OCH_3), 170.0 (*C*OCH_3), 169.9 (*C*OCH_3), 169.7 (*C*OCH_3), 156.3 (NHCO), 136.5, 128.4, 128.0 (C-Ph), 97.4 (C-1), 69.6 (C-2), 69.0 (C-3), 68.3 (C-5), 68.2 (O-CH$_2$), 66.5 (CH$_2$Ph of NHCbz), 66.2 (C-4), 62.5 (C-6), 40.8 (*C*H$_2$-NHCbz), 29.7

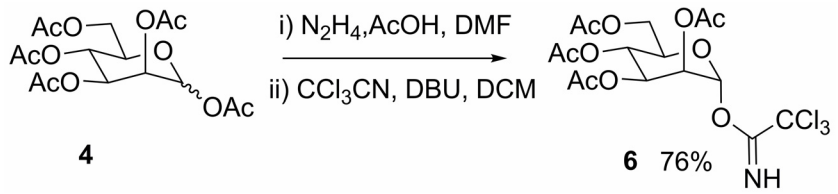
(OCH_2-CH_2), 28.8 (CH_2-CH_2-NHCbz), 23.3 (CH_2), 20.9 ($COCH_3$), 20.8 ($COCH_3$), 20.7 (2 x $COCH_3$); HRMS (ES^+) *m/z* 590.2208 $[M + Na]^+$ (required 590.2197).



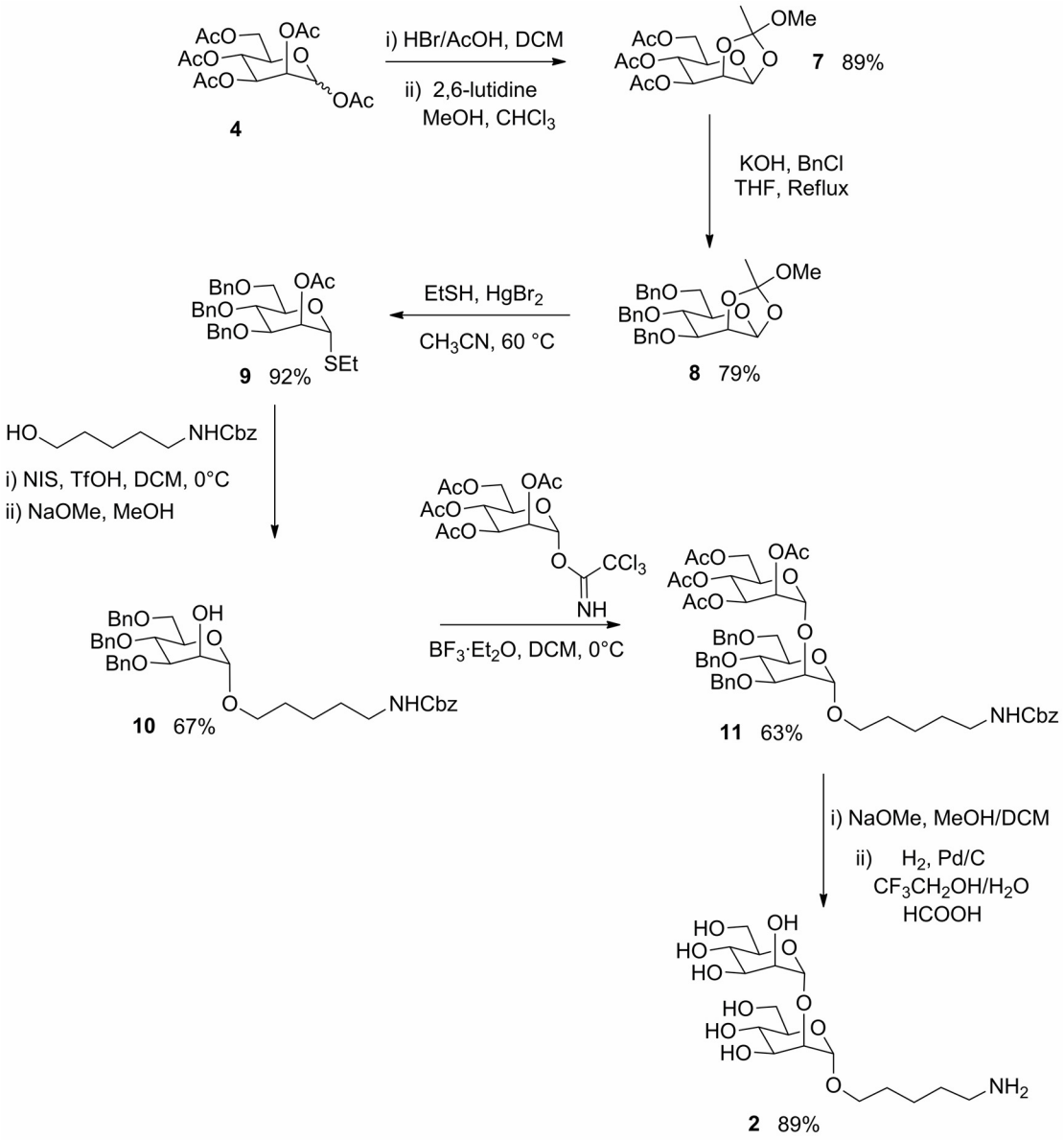
5-(Amino)pentyl α-D-mannopyranoside (1)

To a solution of **5** (0.30 g, 0.53 mmol, 1.0 eq) in dry MeOH/DCM (8 ml, 1:1) was added sodium methoxide (25% wt. solution, 63 µL, 0.27 mmol, 0.5 eq) and stirred at RT for 3 hours. The reaction mixture was neutralized with Dowex 50WX8 (H^+ form), filtered and the solvent removed *in vacuo*. The residue was dissolved in CF_3CH_2OH/H_2O + 1% HCOOH (9:1, 3 mL) was added Pd/C (10%, 0.11 g) and stirred continuously under a hydrogen atmosphere for 24 hours. The reaction mixture filtered over Celite® and washed with $MeOH/H_2O$ (4:1), concentrated, re-dissolved in H_2O and lyophilized to yield the titled compound as an amorphous solid (0.128 g, 91 %).

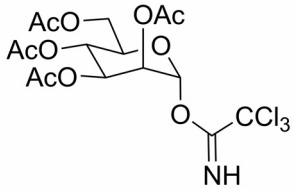
R_f 0.5 (5 ethanol : 3 NH_4OH : 1 water); $[\alpha]_D^{25}$ +44.9 (c = 1.0, MeOH); ^1H NMR (500 MHz, MeOD) δ 4.76 (s, 1H, H-1), 3.85 (dd, *J* = 11.8, 2.0 Hz, 1H, H-6a); 3.80 (dd, *J* = 3.2, 1.6 Hz, 1H, H-2), 3.78 (dt, *J* = 9.8, 6.6 Hz, 1H, O-CH*H*CH_2), 3.72 (dd, *J* = 11.2, 5.2 Hz, 1H, H6-b), 3.70 (dd, *J* = 9.6, 3.9 Hz, 1H, H-3), 3.61 (t, *J* = 9.5 Hz, 1H, H-4), 3.54 (ddd, *J* = 9.7, 6.0, 1.9 Hz, 1H H-5), 3.45 (dt, *J* = 9.8, 6.1 Hz, 1H O-C*H*HCH_2), 2.70 (t, *J* = 7.3 Hz, 2H, C*H*$_2NH_2$), 1.63 (dq, *J* = 13.5, 6.6, 6.6 Hz, 2H, OCH_2CH_2), 1.55 (quin, *J* = 7.1Hz, 2H, C*H*$_2CH_2NH_2$), 1.45 (2H, quin, *J* = 6.9 Hz, CH_2); ^{13}C NMR (126 MHz, MeOD) δ ppm 101.4 (C-1), 74.7 (C-5), 72.7 (C-3), 72.3 (C-2), 68.7 (C-4), 68.4 (OCH_2), 63.0 (C-6), 42.3 (CH_2NH_2), 32.9 ($CH_2CH_2NH_2$), 30.4 (OCH_2CH_2), 24.7 (CH_2), FT-IR (thin film) υ 1595 (NH_2 scissoring), 2968, 2931 (C-H), 3355 br (OH); HRMS *m/z* (ES^+) 266.1602 $[M + H]^+$ required 266.1598.



Scheme S3



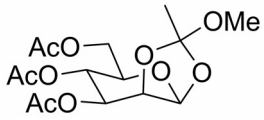
Scheme S4



2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl trichloroacetimidate (6)

To a solution of **4** (4.84 g, 12.4 mmol, 1 eq) in dry DMF (40 mL) was added hydrazine acetate (1.26 g, 13.6 mmol, 1.1 eq). The reaction mixture was stirred for one hour, after which DCM (250 mL) was added. The organic layer was washed with cold saturated $NaHCO_3$ (4 x 250 mL), dried over $MgSO_4$, filtered and concentrated. The crude product was re-dissolved in DCM (100 mL), and CCl_3CN (12.3 mL, 123.5 mmol, 10 eq), DBU (0.47 mL, 3.08 mmol, 0.25 eq) were added and the reaction mixture stirred overnight at RT. Solvents were removed *in vacuo* and the residue purified using flash column chromatography using 30 % EtOAc/Petrol to yield the product as yellow solid (4.60 g, 76 %).

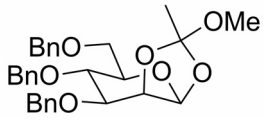
R_f 0.42 (40% EtOAc/Petrol); $[\alpha]_D^{25}$ +49.4 (c = 1.18, $CHCl_3$), lit.[3] $[\alpha]_D^{25}$ +53 (c = 1.01, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) ppm 8.79 (s, 1H, NH), 6.27 (s, 1H, H-1), 5.46 (s, 1H, H-2), 5.39 (m, 2H, H-3, H-4), 4.27 (dd, *J* = 11.9, 4.6 Hz, 1H, H-6a), 4.20 – 4.10 (m, 2H, H-6b, H-5), 2.19 (s, 3H, COC*H*$_3$), 2.08 (s, 3H, COC*H*$_3$), 2.06 (s, 3H, COC*H*$_3$), 2.00 (s, 3H, COC*H*$_3$); ^{13}C NMR (100 MHz, $CDCl_3$) δ ppm 170.6, 169.8, 169.7, 169.2 (*C*OCH$_3$), 159.7 (*C*=NH), 94.5 (C-1), 90.5 (CCl_3), 71.2 (C-5), 68.8 (C-3), 67.8 (C-2), 65.3 (C-4), 62.0 (C-6), 20.8 (CO*C*H$_3$), 20.7 (CO*C*H$_3$), 20.6 (CO*C*H$_3$); LRMS *m/z* (ES$^+$) 550.1 $[M + CH_3CN + NH_4]^+$.



3,4,6-tri-*O*-Acetyl-β-D-mannose 1,2-(methyl orthoacetate) (7)

To a solution of **4** (25.5 g, 65.3 mmol, 1 eq) in dry DCM (200 ml) was added 33% HBr in acetic acid (68.7 g, 850 mmol, 13 eq) at RT and stirred for 2 hours. Cold H_2O (200 mL) was added and the organic layer separated. The aqueous layer was extracted with DCM (2 x 100 mL) and the combined organic extracts were washed with saturated $NaHCO_3$, brine, dried over Na_2SO_4 and concentrated to yield the anomeric bromide which was used without further purification. To the bromo derivative (27.0 g, 65.7 mmol, 1 eq) in dry MeOH and dry $CHCl_3$ (200 mL, 1:1) was added 2,6-lutidine (27.2 g, 253.5 mmol, 3.9 eq) and stirred overnight at RT. Solvents were removed under reduced pressure and the residue co-evaporated with toluene (2 x 100 mL). The crude product was re-dissolved in $CHCl_3$, washed with H_2O, brine, dried over Na_2SO_4, filtered and concentrated. The residue was purified by flash column chromatography using 20% EtOAc/Petrol, to yield the product as white solid (21.3 g, 89%).

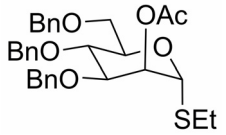
R_f 0.43 (50% EtOAc/Petrol); $[\alpha]_D^{25}$ -22.3 (c = 1.04, $CHCl_3$), lit.[4] $[\alpha]_D^{25}$ -22.2 (c = 2.5, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 5.50 (d, *J* = 2.5 Hz, 1H, H-1), 5.31 (t, *J* = 9.7 Hz, 1H, H-4), 5.15 (dd, *J* = 9.9, 4.0 Hz, 1H, H-3), 4.62 (dd, *J* = 3.9, 2.6 Hz, 1H, H-2), 4.24 (dd, *J* = 12.1, 4.9 Hz, 1H, H-6a), 4.15 (dd, *J* = 12.1, 2.6 Hz, 1H, H-6b), 3.69 (ddd, *J* = 9.5, 4.8, 2.6 Hz, 1H, H-5), 3.28 (s, 3H, OMe), 2.13 (s, 3H, $COCH_3$), 2.08 $COCH_3$), 2.06 $COCH_3$), 1.75 (s, 3H, CH_3); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm 170.6 (*C*OCH_3), 170.3 (*C*OCH_3), 169.4 (*C*OCH_3), 124.5 (C(OMe)CH_3), 97.3 (C-1), 76.5 (C-2), 71.2 (C-5), 70.6 (C-3), 65.4 (C-4), 62.2 (C-6), 49.9 (OMe), 24.3 (CH_3), 20.7 (CO*C*H_3), 20.7 (CO*C*H_3), 20.6 (CO*C*H_3); HRMS *m/z* (ES^+) 385.1106 $[M + Na]^+$ (required 385.1105).



3,4,6-tri-*O*-Benzyl-β-D-mannose 1,2-(methyl orthoacetate) (8)

To a solution of **7** (1.00 g, 2.76 mmol, 1 eq) in dry THF (10 mL) was added benzyl chloride (5.94 g, 46.9 mmol, 17 eq) and the reaction heated to 80 °C. After 15 minutes, the reaction mixture was cooled and powdered KOH (2.32 g, 41.4 mmol, 15 eq) was added and heating continued overnight, after which H_2O was added and the reaction mixture diluted with DCM. The organic layer was separated, washed with saturated $NaHCO_3$, brine, dried over Na_2SO_4, filtered and concentrated. The residue was purified by flash column chromatography using 20% EtOAc/Petrol, to yield the product as a white solid (1.11 g, 79%).

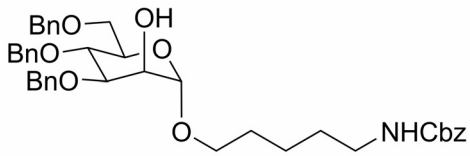
R_f 0.49 (30% EtOAc/Petrol); $[\alpha]_D^{25}$ +29.8 (c = 1.04, $CHCl_3$), lit.[5] $[\alpha]_D^{24}$ +26 (c = 1.1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.42 – 7.24 (m, 15H, Ph), 5.36 (d, *J* = 2.50 Hz, 1H, H-1), 4.91 (d, *J* = 10.7 Hz, 1H, CH*H*Ph), 4.80 (AB Quartet, *J* = 12.0 Hz, 2H, CH_2Ph), 4.65 -4.53 (m, 3H, CH_2Ph; C*H*HPh), 4.41 (dd, *J* = 3.9, 2.6 Hz, 1H, H-2), 3.93 (t, *J* = 9.3 Hz, 1H, H-4), 3.80 - 3.68 (m, 3H, H-3, H-6a,b), 3.43 (ddd, *J* = 9.4, 4.3, 2.3 Hz, 1H, H-5), 3.30 (s, 1H, OMe), 1.75 (s, 3H, CH_3); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm 138.2, 137.8, 128.5, 128.4, 128.3, 128.0, 128.0, 127.7, 127.5 (C-Ph), 123.9 (*C*(OMe)CH_3), 97.5 (C-1), 79.0 (C-2), 77.1 (C-3), 75.2 (CH_2Ph), 74.1 (C-5), 74.1 (C-4), 73.3 (CH_2Ph), 72.3 (CH_2Ph), 68.9 (C-6), 49.7 (OMe), 24.4 (CH_3); HRMS *m/z* (ES^+) 529.2196 $[M + Na]^+$ (required 529.2197).



Ethyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-1-thio-α-D-mannopyranoside (9)

To a solution of **8** (2.36 g, 4.66 mmol, 1 eq) and freshly activated 4 Å molecular sieves in dry CH_3CN (30 mL) was added ethane thiol (1.45 g, 23.3 mmol, 5 eq) and the resultant mixture stirred for one hour. $Hg(Br)_2$ (0.336 g, 0.932 mmol) was added and the reaction heated to 60 °C overnight. The reaction mixture was filtered over a pad of Celite®, washed with EtOAc and the filtrate concentrated. The residue was purified by flash column chromatography using 10% EtOAc/Petrol, to yield the product as a white solid (2.30 g, 92%).

R_f 0.6 (30% EtOAc/Petrol); $[\alpha]_D^{25}$ +74.5 (c = 1.10, $CHCl_3$), lit.[6] $[\alpha]_D^{20}$ +84 (c = 1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.53 - 7.09 (m, 15H, Ph), 5.45 (dd, *J* = 2.7, 1.7 Hz, 1H, H-2), 5.34 (d, *J* = 1.4 Hz, 1H, H-1), 4.87 (d, *J* = 10.8 Hz, 1H, C*H*HPh), 4.70 (AB Quarter, *J* = 11.6 Hz, 2H, CH_2Ph), 4.57 - 4.47 (m, 3H, CH_2Ph, CH*H*Ph), 4.12 (ddd, *J* = 8.1, 4.2, 2.1 Hz, 1H, H-5), 3.98 – 3.90 (m, 2H, H-3, H-4), 3.85 (dd, *J* = 10.8, 4.2 Hz, 1H, H-6a), 3.70 (dd, *J* = 10.8, 1.8 Hz, 1H, H-6b), 2.75 - 2.53 (m, 2H, S-CH_2), 2.18 (s, 1H, $COCH_3$), 1.29 (t, *J* = 7.4 Hz, 3H, SCH_2CH_3); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm 170.4 ($COCH_3$), 138.3, 138.1, 137.6, 128.4, 128.3, 128.1, 127.8, 127.7, 127.6, 127.6 (C-Ph), 82.4 (C-1), 78.5 (C-4), 75.1 (C-3), 74.5 (CH_2Ph), 73.3 (C-5), 71.8 (CH_2Ph), 71.7 (CH_2Ph), 70.5 (C-2), 68.7 (C-6), 25.4 (S-CH_2), 21.1 ($COCH_3$), 14.8 (CH_3); HRMS *m/z* (ES^+) 559.2125 $[M + Na]^+$ (required 529.2125).

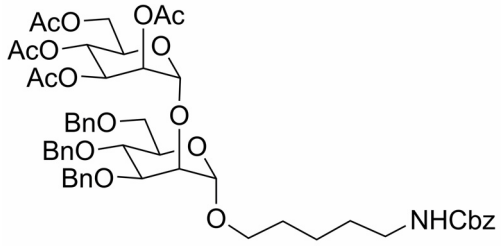


5-(Benzyloxycarbonylamino)pentyl 3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (10)

To a stirred solution of **9** (1.00 g, 1.86 mmol, 1.1 eq) and **3** (0.40 g, 1.69 mmol, 1 eq) and 4 Å molecular sieves in dry DCM (6 mL) were successively added, at 0 °C and under argon, NIS (0.762 g, 3.39 mmol, 2 eq) and TfOH (0.051 g, 0.339 mmol, 0.2eq). After stirring for 2 hours, the reaction mixture was diluted with DCM, quenched with Et_3N and filtered through a pad of Celite®. The organic layer was washed with 5 % $Na_2S_2O_3$, brine, dried over Na_2SO_4, filtered, concentrated and the crude product used without further purification. The linker attached mannose derivative (1.21 g, 1.69 mmol, 1 eq) was dissolved in dry MeOH (25 mL) and then treated with 25 wt% sodium methoxide (0.114 g, 2.12 mmol, 1.25 eq). The resultant reaction mixture was stirred at RT for 4 hours, after which the reaction mixture was neutralized with Dowex 50WX8 (H^+) resin, filtered, washed with methanol and concentrated. The residue was purified by flash chromatography (30 – 40 % EtOAc/Petrol) to afford the titled compound (0.76 g, 67 % yield) as colorless oil.

R_f 0.18 (40 % EtOAc/Petrol); $[\alpha]_D^{25}$ +31.8 (c = 1.35, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.18 – 7.38 (m, 20H, ArH), 5.09 (s, 2H, C*H*$_2$Ph of Cbz), 4.86 (s, 1H, H-1), 4.83 (d, *J* = 10.8 Hz, 1H, CH*H*Ph), 4.71 (AB Quartet, *J* = 11.2 Hz, 2H, C*H*$_2$Ph), 4.65 (d, *J* = 12.4 Hz, 1H, C*H*HPh), 4.54 (d, *J* = 12.1 Hz, 1H, C*H*$_2$Ph), 4.51 (d, *J* = 10.1 Hz, 1H, C*H*$_2$Ph), 4.04 (s, 1H, H-2), 3.82 – 3.90 (m, 2H, H-3, H-4), 3.71 – 3.78 (m, 3H, H-5, H-6a,b), 3.66 – 3.70 (m, 1H, OCH*H*CH$_2$), 3.39 – 3.44 (m, 1H, OC*H*HCH$_2$), 3.18 (dd, *J* = 6.4, 12.0 Hz, 2H, C*H*$_2$NHCbz), 2.62 (br. s, 1H, OH), 1.84 (br. S, 1H, NH), 1.46 – 1.59 (m, 4H, OCH$_2$C*H*$_2$, C*H*$_2$CH$_2$NHCbz), 1.33 – 1.39 (m, 2H, C*H*$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 156.4 (NH*C*O), 99.1 (C-1), 80.2 (C-

3), 75.1 (*C*H_2Ph), 74.3 (C-5), 73.4 (*C*H_2Ph), 71.9 (*C*H_2Ph), 71.0 (C-4), 68.9 (C-2), 68.3 (C-6), 67.4 (O*C*H_2), 40.9 (*C*H_2NHCbz), 29.7 (OCH_2*C*H_2), 29.0 (*C*H_2CH_2NHCbz), 23.2 (*C*H_2); HRMS m/z (ES$^+$) 692.3194 $[M + Na]^+$ (required 692.3187).

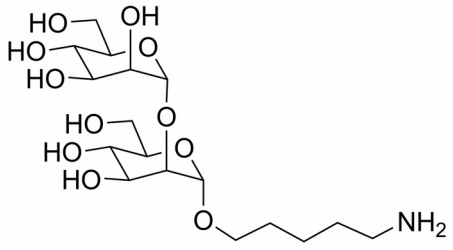


5-(Benzyloxycarbonylamino)pentyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl (1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (11)

To a stirred solution of **10** (0.31 g, 0.46 mmol, 1 eq), **6** (0.29 g, 0.60 mmol, 1.3 eq) and 4 Å molecular sieves in dry DCM (6 ml) at 0 °C under argon was added $BF_3{\cdot}Et_2O$ (0.033 g, 0.114 mmol, 0.25 eq). After stirring for 2 hours, the reaction mixture was quenched with Et_3N and concentrated. The residue was purified using 35-40 % EtOAc/petrol to yield the compound as a colourless oil (0.29 g, 63%).

R_f 0.24 (40 % EtOAc/Petrol); $[\alpha]_D^{25}$ +24.1 (c = 1.92, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.25 – 7.38 (m, 18H, ArH), 7.15 – 7.17 (m, 2H, ArH), 5.47 (dd, *J* = 4.0, 2.0 Hz, 1H, H-2'), 5.41 (dd, *J* = 10.0, 3.2 Hz, 1H, H-3'), 5.26 (t, *J* = 10.0 Hz, 1H, H-4'), 5.08 (s, 2H, C*H*$_2$Ph of Cbz), 4.98 (d, *J* = 1.2 Hz, 1H, H-1'), 4.86 (d, *J* = 1.2 Hz, 1H, H-1), 4.83 (d, *J* = 10.9 Hz, 1H, C*H*HPh), 4.73 (d, *J* = 11.9 Hz, 1H, C*H*HPh), 4.64 (d, *J* = 11.9 Hz, 1H, CH*H*Ph), 4.63 (d, *J* = 12.5 Hz, 1H, C*H*HPh), 4.58 (d, *J* = 12.3 Hz, 1H, CH*H*Ph), 4.50 (d, *J* = 10.8 Hz, CH*H*Ph), 4.25 (dd, *J* = 11.9, 5.1 Hz, 1H, H-6'a), 4.19 (ddd, *J* = 9.7, 5.1, 1.9 Hz, 1H, H-5'), 4.10 (dd, *J* = 11.9, 2.0 Hz, 1H, H-6'b), 3.93 (dd, *J* = 4.3, 2.2 Hz, 1H, H-2), 3.90 (dd, *J* = 9.2, 2.4 Hz, 1H, H-3), 3.86 (t, *J* = 8.9 Hz, 1H, H-4), 3.72-3.78 (m, 3H, H-5, H-6a, H-6b), 3.68 (dt, *J* = 9.6, 6.5, 6.5 Hz, 1H, OC*H*HCH$_2$), 3.38 (dt, *J* = 9.4, 6.3, 6.3 Hz, 1H, OCH*H*CH$_2$), 3.17 (dd, *J* = 13.4, 6.7

Hz, 1H, C*H*$_2$NHCbz), 2.11 (s, 3H, COC*H*$_3$), 2.09 (s, 3H, COC*H*$_3$), 2.01 (s, 3H, COC*H*$_3$), 2.00 (s, 3H, COC*H*$_3$), 1.67 (br. s, 1H, NH), 1.57 (quin, *J* = 7.0 Hz, 2H, OCH$_2$C*H*$_2$), 1.50 (quin, *J* = 7.3 Hz, 2H, CH$_2$C*H*$_2$NHCbz), 1.33 (quin, *J* = 8.2 Hz, 2H, C*H*$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 170.6 (*C*OCH$_3$), 169.8 (*C*OCH$_3$), 169.7 (*C*OCH$_3$), 169.6 (*C*OCH$_3$), 156.4 (NH*C*O), 138.3, 136.6, 128.5, 128.3, 128.1, 128.0, 127.6, 127.5, 127.4, 127.3 (Ph), 99.3 (C-1'), 98.5 (C-1), 79.5 (C-3), 76.3 (C-2), 75.3 (*C*H$_2$Ph), 74.8 (C-4), 73.1 (*C*H$_2$Ph), 72.3 (*C*H$_2$Ph), 71.7 (C-5), 69.4 (C-2'), 69.1 (C-6), 69.0 (C-3'), 68.7 (C-5'), 67.5 (OCH$_2$), 66.5 (*C*H$_2$Ph of NHCbz), 66.2 (C-4'), 62.5 (C-6'), 40.9 (*C*H$_2$NHCbz), 29.7 (OCH$_2$*C*H$_2$), 29.0 (CH$_2$*C*H$_2$NHCbz), 23.3 (*C*H$_2$), 20.9 (CO*C*H$_3$), 20.7 (CO*C*H$_3$), 20.7 (CO*C*H$_3$), 20.6 (CO*C*H$_3$); HRMS m/z (ES$^+$) 1022.4148 [M + Na]$^+$ (required 1022.4150).



5-(Amino)pentyl α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (2)

To a solution of **11** (0.29 g, 0.29 mmol, 1 eq) in dry MeOH/DCM (8 mL, 1:1) was added sodium methoxide (25% wt. solution, 104 μL, 0.44 mmol 1.5 eq) and stirred at RT for 2 hours. The reaction mixture was neutralized with Dowex 50WX8 (H^+ form), filtered and concentrated *in vacuo*. The crude product was dissolved in CF_3CH_2OH/H_2O + 1% HCOOH (9:1, 6 mL) with Pd/C (10%, 0.25 g) and stirred continuously under a H_2 atmosphere for 24 hours. The reaction mixture was filtered over Celite® and washed with MeOH/H_2O (4:1), concentrated, re-dissolved in H_2O and lyophilized to yield the desired compound as an amorphous white solid (0.109 g, 89 %).

R_f 0.4 (1 water : 4 isopropanol : 5 ethyl acetate), $[\alpha]_D^{25}$ + 48.3 (c = 1.0, H_2O); ^{1}H NMR (400 MHz, D_2O) δ ppm 4.97 (d, *J* = 1.2 Hz, 1H, H-1), 4.88 (d, *J* = 1.2 Hz, 1H,

H-1'), 3.94 (dd, *J* = 1.2, 3.2 Hz, 1H, H-2'), 3.81 (dd, *J* = 1.2, 3.2 Hz, 1H, H-2), 3.76 – 3.78 (m, 1H, H-6'a), 3.74 – 3.76 (m, 2H, H-6'b, H-3), 3.71 (dd, *J* = 3.2, 9.6 Hz, 1H, H-3'), 3.65 (br. s, 1H, H-5'), 3.60 – 3.63 (m, 2H, H-6a, OC*H*HCH$_2$), 3.56 (d, *J* = 9.6 Hz, 1H, H-6b), 3.45 – 3.52 (m, 3H, H-4, H-4', H-5), 3.41 (dt, *J* = 6.0, 10.0 Hz, 1H, OCH*H*CH$_2$), 2.86 (t, *J* = 8.0 Hz, 2H, C*H*$_2$-NH$_2$), 1.49 – 1.59 (m, 4H, OCH$_2$C*H*$_2$, C*H*$_2$CH$_2$NH$_2$), 1.27 – 1.37 (m, 2H, C*H*$_2$); ^{13}C NMR (101 MHz, CDCl$_3$) δ ppm 102.7 (C-1'), 98.4 (C-1), 79.1 (C-2), 73.7 (C-5'), 73.1 (C-5), 70.7 (C-3), 70.6 (C-3'), 70.3 (C-2'), 67.9 (O*C*H$_2$), 67.3 (C-4'), 67.2 (C-4), 61.5 (C-6'), 61.3 (C-6), 39.8 (*C*H$_2$NH$_2$), 28.3 (OCH$_2$*C*H$_2$), 27.1 (*C*H$_2$CH$_2$NH$_2$), 22.8 (*C*H$_2$); HRMS m/z (ES$^+$) 428.2126 [M + Na]$^+$ (required 428.2128).

5. Preparation of CRM-197

Expression of CRM

A 10 mL starter culture of LB containing 0.1 mg/mL ampicillin was inoculated with a single colony from a fresh transformation of CRM in *E. coli* BL21(DE3) cells and grown at 37 ºC overnight with shaking.

One litre of LB containing 0.1 mg/mL ampicillin was inoculated with the 10 mL overnight culture. This was grown at 37 ºC with shaking to an OD_{600} of 0.6. Protein production was induced with the addition of 1 mM IPTG. The culture was allowed to continue shaking at 25 ºC for 16 hours.

The cells were harvested by spinning for 20 minutes at 13000g. The cell pellets were collected and stored at -80 ºC until next use.

Purification of CRM

Cell lysis: The cell pellets collected from 2 L of growth media were lysed in 40 mL of a buffer containing 50 mM TEA at pH 7.8, 300 mM NaCl, 15 mM imidazole, 1

mM BME, 1 mg/mL lysozyme and 0.1 mg/mL DNAse. The pellet and buffer were allowed to incubate on ice for 30 minutes.

The cell pellets were then sonicated at 150 W in bursts of 30 seconds with a wait time of 1 minute between each burst. A total of 4 bursts were used to break the cells. The sonication of the cells was carried out in a plastic beaker placed on ice.

The cell debris was removed from the lysis mixture by centrifugation at 13000g for 30 minutes. Samples of the pellet and soluble fractions were saved for SDS-PAGE evaluation.

NiNTA affinity purification: The soluble portion of the lysed cells was loaded onto a 5 mL pre-packed NiNTA column (GE Healthcare) which had been equilibrated in buffer containing 50 mM TEA at pH 7.8, 300 mM NaCl, 1 mM BME and 15 mM imidazole. CRM was eluted using a linear gradient to the same buffer containing 300 mM imidazole. Protein-containing fractions (as determined by UV absorbance) were analyzed by SDS-PAGE.

Anion exchange purification: Protein-containing fractions from the NiNTA column were pooled and then loaded onto a 1 mL MonoQ 5/50 GL anion exchange column (GE healthcare) which had been equilibrated in 20 mM Tris, pH 7.8, 150 mM NaCl. The protein was eluted using a linear gradient to the same buffer containing 1 M NaCl. Protein-containing fractions were then analyzed by SDS-PAGE.

Size exclusion chromatography: Protein-containing fractions from the MonoQ column were concentrated to a volume < 10 mL. The protein was then loaded onto a S75 size exclusion column (HiLoad 16/60 Superdex 75, GE Healthcare) which had been equilibrated in 20 mM Tris, pH 7.8, 150 mM NaCl. Only 2mL of protein solution was added at any one time. Protein-containing fractions (as determined by

UV absorbance) were analyzed by SDS-PAGE and combined and concentrated and stored at -20 ºC until further use.

For glycoconjugation experiments, the protein was exchanged into 300 mM $NaHCO_3$ buffer at pH 9.0 using a PD10 column (GE Healthcare). The protein concentration was checked by BCA assay (Pierce), and the protein was diluted to 3.0 mg/mL with the same $NaHCO_3$ buffer prior to the experiment.

6. Protein sequences

BSA

```
dthkseiahr fkdlgeehfk glvliafsqy lqqcpfdehv klvneltefa ktcvadesha  60
gcekslhtlf gdelckvasl retygdmadc cekqeperne cflshkddsp dlpklkpdpn  120
tlcdefkade kkfwgkylye iarrhpyfya pellyyanky ngvfqeccqa edkgacllpk  180
ietmrekvlt ssarqrlrca siqkfgeral kawsvarlsq kfpkaefvev tklvtdltkv  240
hkecchgdll ecaddradla kyicdnqdti ssklkeccdk plleкshcia evekdaipen  300
lppltadfae dkdvcknyqe akdaflgsfl yeysrrhpey avsvllrlak eyeatleecc  360
akddphacys tvfdklkhlv depqnlikqn cdqfeklgey gfqnalivry trkvpqvstp  420
tlvevsrslg kvgtrccтkp esermpcted ylslilnrlc vlhektpvse kvtkcctesl  480
vnrrpcfsal tpdetyvpka fdeklftfha dictlpdtek qikkqtalve llkhkpkate  540
eqlktvmenf vafvdkccaa ddkeacfave gpklvvstqt ala                    583
```

Molecular weight: 66463

CRM

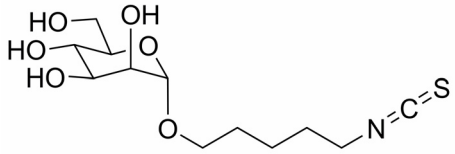
```
GADDVVDSSK SFVMENFSSY HGTKPGYVDS IQKGIQKPKS GTQGNYDDDW KGFYSTDNKY  60
DAAGYSVDNE NPLSGKAGGV VKVTYPGLTK VLALKVDNAE TIKKELGLSL TEPLMEQVGT  120
EEFIKRFGDG ASRVVLSLPF AEGSSSVEYI NNWEQAKALS VELEINFETR GKRGQDAMYE  180
YMAQACAGNR VRRSVGSSLS CINLDWDVIR DKTKTKIESL KEHGPIKNKM SESPNKTVSE  240
EKAKQYLEEF HQTALEHPEL SELKTVTGTN PVFAGANYAA WAVNVAQVID SETADNLEKT  300
TAALSILPGI GSVMGIADGA VHHNTEEIVA QSIALSSLMV AQAIPLVGEL VDIGFAAYNF  360
VESIINLFQV VHNSYNRPAY SPGHKTQPFL HDGYAVSWNT VEDSIIRTGF QGESGHDIKI  420
TAENTPLPIA GVLLPTIPGK LDVNKSKTHI SVNGRKIRMR CRAIDGDVTF CRPKSPVYVG  480
NGVHANLHVA FHRSSSEKIH SNEISSDSIG VLGYQKTVDH TKVNSKLSLF FEIKSLEHHH  540
HHH                                                                543
```

Molecular weight: 59406

7. Protein modification

General protocol for preparation of isothiocyanates

To a solution of **1** or **2** (1 eq) in 0.3 M $NaHCO_3$ (6 eq) and 0.3 M Na_2CO_3 (3 eq) at pH 9 (pH paper) was added a solution of thiophosgene (6 eq) dissolved in chloroform. The biphasic mixture was vigorously stirred at RT and monitored by TLC (1 water : 4 isopropanol : 4 ethyl acetate) and mass spectrometry (ES^+). After ~ 3 hours, complete consumption of the starting sugar and the formation of a single product was detected. Solvents and excess thiophosgene were removed under reduced pressure. The isothiocyanate of **1** was desalted for characterization purposes by silica gel chromatography (9 ethyl acetate : 1 methanol) to afford the desired compound **12** as a pale yellow oil (24 mg, 84 %). The isothiocyanate of **2** was used without further purification.



1-Isothiocyano-5-pentyl-O-α-D-mannoopyranoside (12)

R_f 0.23 (9 ethyl acetate : 1 methanol), $[\alpha]_D^{18}$ +33.2 (c = 1.2, MeOH); ^{1}H NMR (500 MHz, MeOD) δ ppm 4.77 (1d, *J* = 1.6 Hz, 1H, H-1), 3.83 (m, 1H, H-6a), 3.83 (t, *J* = 2.5 Hz, 1H, H-2), 3.76 (ad, *J* 5.7 Hz, 1H, OC*H*HCH$_2$), 3.75 (m, 1H, H-6b), 3.73 (dd, *J* = 10.1, 2.5 Hz, 1H, H-3), 3.65 (t, *J* = 9.6 Hz, 1H, H-4), 3.60 (t, *J* = 6.6 Hz, 1H, C*H*HNCS), 3.54 (ddd, *J* = 9.5, 5.5, 2.4 Hz, 1H, H-5), 3.48 (t, *J* = 6.3 Hz, 1H, CH*H*NCS), 3.47 (ddd, *J* = 14.3, 8.8, 2.2 Hz, 1H, OCH*H*CH$_2$), 1.76 (quin, *J* = 6.9 Hz, 1H, OCH$_2$C*H*HCH$_2$), 1.60-1.70 (m, 3H, OCH$_2$CH*H*, C*H*$_2$CH$_2$NCS), 1.55 (m, 1H, C*H*HCH$_2$), 1.45 (m, 1H, CH*H*CH$_2$); ^{13}C NMR (126 MHz, MeOD) δ ppm 101.6 (C-1), 74.6 (C-5), 72.6 (C-3), 72.2 (C-2), 68.6 (C-4), 68.3 (OCH$_2$), 62.7 (C-6), 45.9 (*C*H$_2$NCS), 30.9 (OCH$_2$*C*H$_2$), 29.8

(*C*H_2CH_2NCS), 24.5 (CH_2), low intensity NCS peak not observed; FT-IR (thin film) υ 2102 (N=C=S), 2180 (N=C=S), 2936 (C-H), 3354 br (OH); HRMS m/z (ES^+) 330.0970 $[M + Na]^+$ (required 330.0982).

General protocol for reactions

Reactions were performed using between 0.1 – 1.0 mg of protein at a concentration of 1.5mg/ml. The reactions were conducted in 0.3 M $NaHCO_3$ and 0.3 M Na_2CO_3. A solution of the isothiocyanate was added and the reaction mixture was gently shaken at RT for 24 hours. Increasing ratios of isothiocyanate per lysine residue were used per reaction set.

8. MALDI mass spectrometric analysis

5 μL of the reaction mixture was diluted to 250 μL with 0.1 % TFA in water. This solution was mixed in a 1 : 1 ratio (v/v) with a solution of mass spectrometry grade sinapinic acid (10 mg/mL in 1 : 1 water : acetonitrile with 0.1% TFA). From this combined matrix/ sample solution, 2 μL was spotted onto a steel target and allowed to co-crystallize at room temperature over 3 hours. MALDI analysis was conducted on a Water MALDI Micro MX spectrometer with TOF detection, in positive linear mode. Unmodified protein was used as a lock mass calibrant. For each comparative set of experiments, laser energy and pulse width were optimised for a modified protein sample, and kept identical for the remainder of the sample set. In general, multiple laser shots were combined to produce the final spectrum. Data were further processed using Mass Lynx 4.1. Mass spectra were smoothed with Savitzky-Golay smoothing prior to dispersity analysis.

9. MATLAB M-file scripts

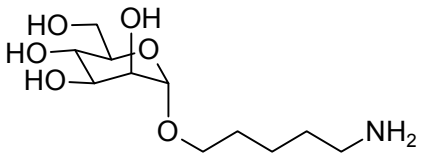
Example M-file code for BSA (n = 59) modified with **1** (m = 307), where the measured FWHM for unmodified BSA = 1298:

```
function [estimates, model] = Dispersity(xdata, ydata)
start_point = [0 500];
model = @MALDIfit;
options = optimset('Display','iter');
estimates = fminsearch(model, start_point, options);
    function [sse, FittedCurve] = MALDIfit(params)
        mu = params(1);
        sigma = params(2);
        m = 307;
        % m = the mass change per modification
        delta = 1298/2.35;
        % delta = standard deviation of MALDI peak for unmodified protein
        % 1298 is the measured FWHM
        x = (-10000:20000);
        FittedCurve = 0;
            for n=0:60
            % n = the number of lysines + N-terminus
            FittedCurve = FittedCurve + (normcdf((n+0.5)*m,mu,sigma)-
normcdf((n-0.5)*m,mu,sigma))*normpdf(x, n*m, delta);
            end
        ErrorVector = FittedCurve - ydata;
        sse = sum(ErrorVector .^ 2);
    end
end
```

Command line entry for a MALDI peak of modified protein of average mass relative to unmodified BSA = 3247, and measured FWHM = 3277:

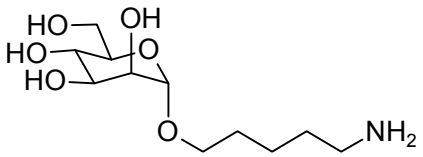
```
xdata = (-10000:20000);
ydata = normpdf(xdata, 3247, 3277/2.35);
[estimates, model] = Dispersity(xdata,ydata)
```

10. NMR spectra



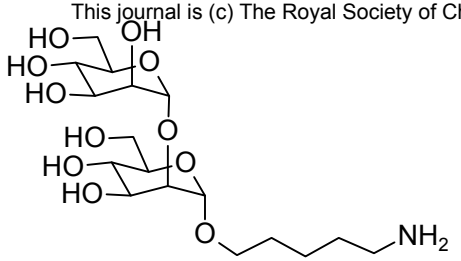
1

Frequency (MHz)	500.30
Nucleus	1H
Number of Transients	16
Origin	avc500
Pulse Sequence	zg30
Receiver Gain	4.00
SW(cyclical) (Hz)	10330.58
Solvent	MeOD
Spectrum Offset (Hz)	3089.5557
Sweep Width (Hz)	10330.26
Temperature (degree C)	27.000



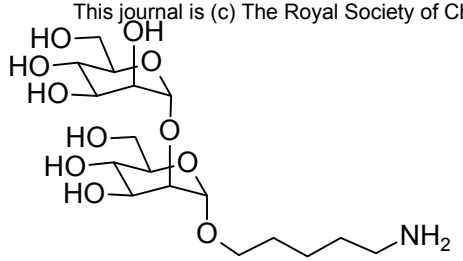
1

Frequency (MHz)	125.80
Nucleus	13C
Number of Transients	512
Origin	avc500
Pulse Sequence	zgpg30
Receiver Gain	1820.00
SW(cyclical) (Hz)	31250.00
Solvent	MeOD
Spectrum Offset (Hz)	12758.7549
Sweep Width (Hz)	31249.05
Temperature (degree C)	27.000



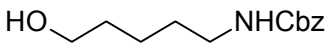
2

Frequency (MHz)	400.13
Nucleus	1H
Number of Transients	16
Origin	dpx400
Pulse Sequence	zg60
Receiver Gain	143.70
SW(cyclical) (Hz)	5592.84
Solvent	DEUTERIUM OXIDE
Spectrum Offset (Hz)	2000.6499
Sweep Width (Hz)	5592.67
Temperature (degree C)	27.000



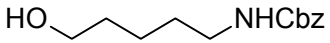
2

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	DEUTERIUM OXIDE
Spectrum Offset (Hz)	10063.0566
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



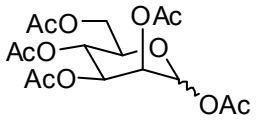
3

Frequency (MHz)	400.20
Nucleus	1H
Number of Transients	16
Origin	av400
Pulse Sequence	zg60
Receiver Gain	90.50
SW(cyclical) (Hz)	8278.15
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	2468.5974
Sweep Width (Hz)	8277.89
Temperature (degree C)	27.000



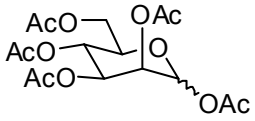
3

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



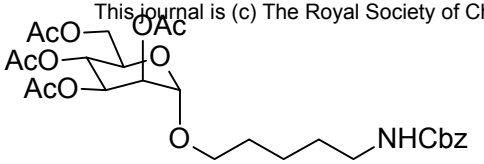
4

Frequency (MHz)	400.13
Nucleus	1H
Number of Transients	16
Origin	dpx400
Pulse Sequence	zg60
Receiver Gain	90.50
SW(cyclical) (Hz)	5592.84
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	1982.4041
Sweep Width (Hz)	5592.67
Temperature (degree C)	27.000



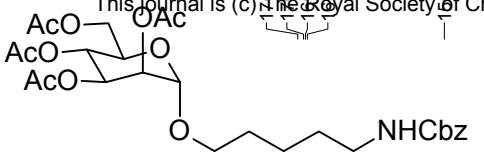
4

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



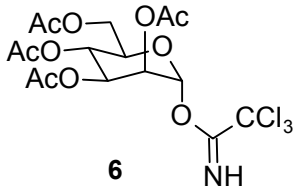
5

Frequency (MHz)	400.20
Nucleus	1H
Number of Transients	16
Origin	av400
Pulse Sequence	zg60
Receiver Gain	128.00
SW(cyclical) (Hz)	8278.15
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	2468.5974
Sweep Width (Hz)	8277.89
Temperature (degree C)	27.000



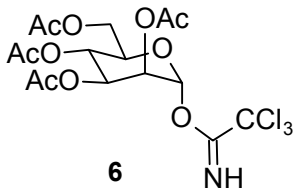
5

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



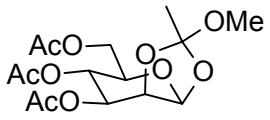
6

Frequency (MHz)	400.20
Nucleus	1H
Number of Transients	16
Origin	av400
Pulse Sequence	zg60
Receiver Gain	90.50
SW(cyclical) (Hz)	8278.15
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	2468.5974
Sweep Width (Hz)	8277.89
Temperature (degree C)	21.300



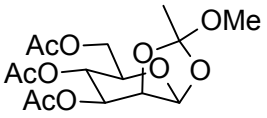
6

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	21.600



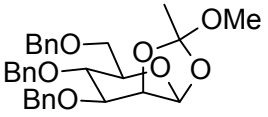
7

Frequency (MHz)	400.20
Nucleus	1H
Number of Transients	16
Origin	av400
Pulse Sequence	zg60
Receiver Gain	228.10
SW(cyclical) (Hz)	8278.15
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	2468.5974
Sweep Width (Hz)	8277.89
Temperature (degree C)	27.000



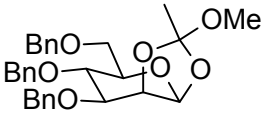
7

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



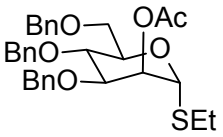
8

Frequency (MHz)	400.20
Nucleus	1H
Number of Transients	16
Origin	av400
Pulse Sequence	zg60
Receiver Gain	228.10
SW(cyclical) (Hz)	8278.15
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	2468.5974
Sweep Width (Hz)	8277.89
Temperature (degree C)	27.000



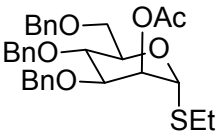
8

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



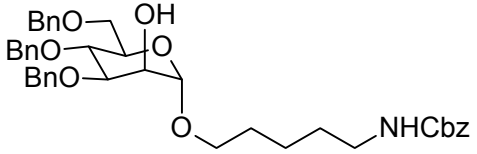
9

Frequency (MHz)	400.20
Nucleus	1H
Number of Transients	16
Origin	av400
Pulse Sequence	zg60
Receiver Gain	90.50
SW(cyclical) (Hz)	8278.15
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	2468.5974
Sweep Width (Hz)	8277.89
Temperature (degree C)	27.000



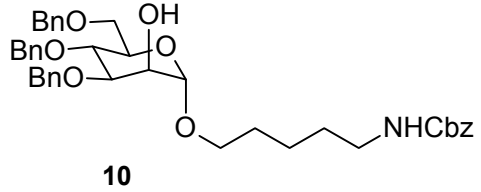
9

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



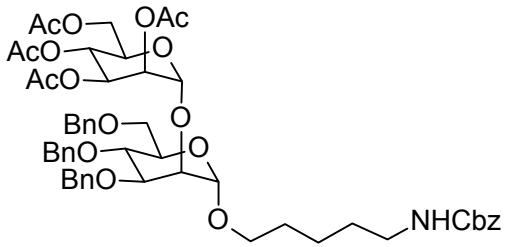
10

Frequency (MHz)	400.20
Nucleus	1H
Number of Transients	16
Origin	av400
Pulse Sequence	zg60
Receiver Gain	90.50
SW(cyclical) (Hz)	8278.15
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	2468.5974
Sweep Width (Hz)	8277.89
Temperature (degree C)	27.000



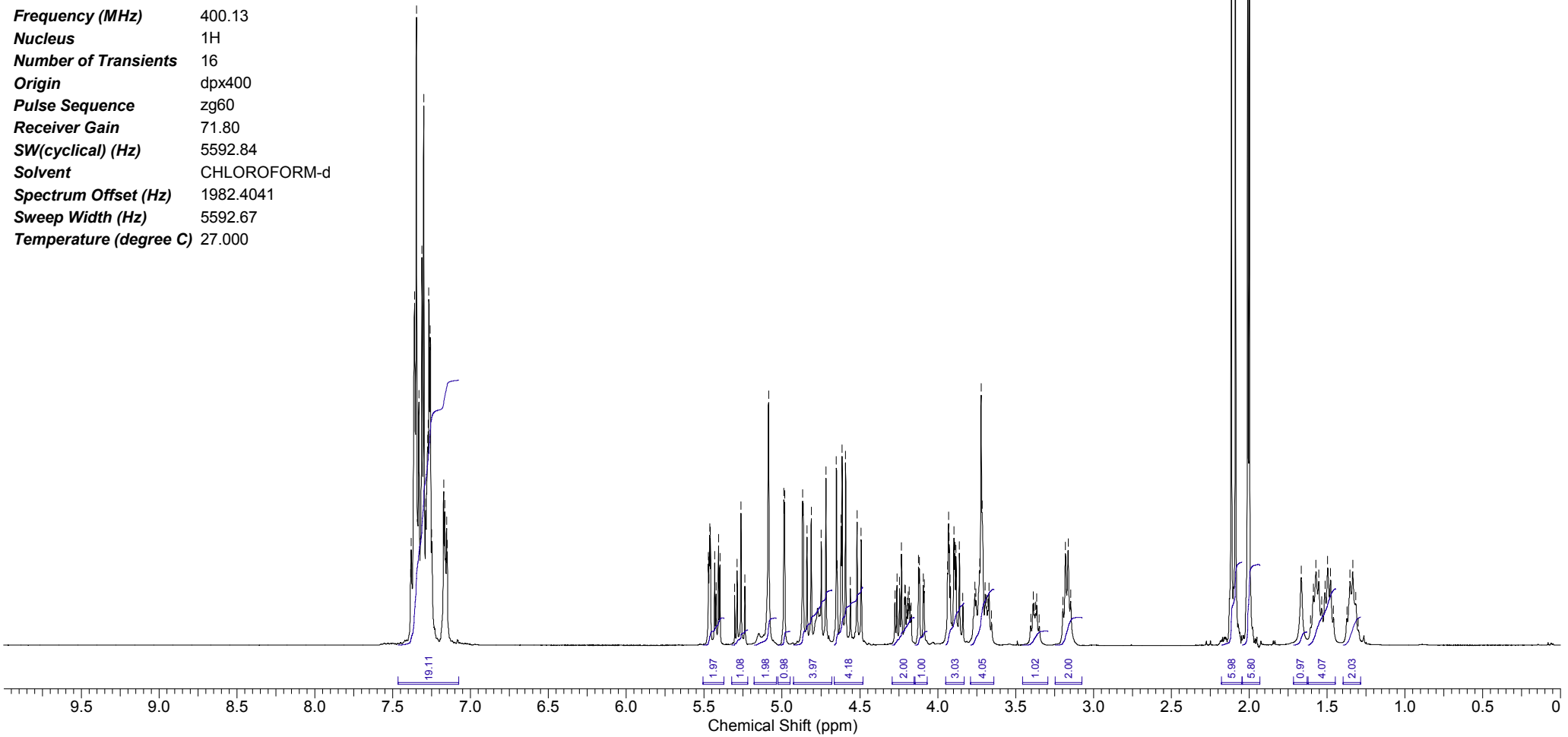
10

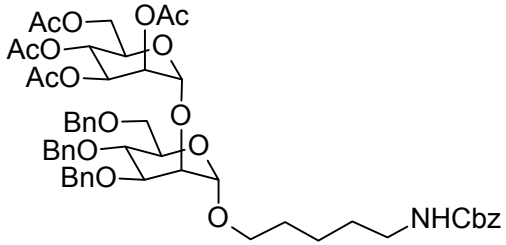
Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



11

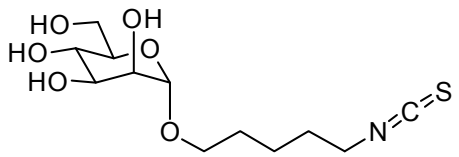
Frequency (MHz)	400.13
Nucleus	1H
Number of Transients	16
Origin	dpx400
Pulse Sequence	zg60
Receiver Gain	71.80
SW(cyclical) (Hz)	5592.84
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	1982.4041
Sweep Width (Hz)	5592.67
Temperature (degree C)	27.000





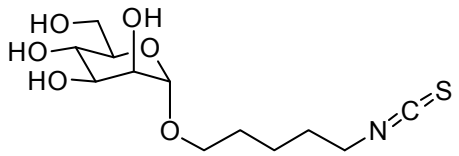
11

Frequency (MHz)	100.63
Nucleus	13C
Number of Transients	256
Origin	av400
Pulse Sequence	zgpg30
Receiver Gain	32768.00
SW(cyclical) (Hz)	26178.01
Solvent	CHLOROFORM-d
Spectrum Offset (Hz)	10021.2939
Sweep Width (Hz)	26177.21
Temperature (degree C)	27.000



12

Frequency (MHz)	500.30
Nucleus	1H
Number of Transients	16
Origin	avc500
Pulse Sequence	zg30
Receiver Gain	4.00
SW(cyclical) (Hz)	10330.58
Solvent	MeOD
Spectrum Offset (Hz)	3089.5557
Sweep Width (Hz)	10330.26
Temperature (degree C)	27.000



12

Frequency (MHz)	125.80
Nucleus	13C
Number of Transients	1024
Origin	avc500
Pulse Sequence	zgpg30
Receiver Gain	1820.00
SW(cyclical) (Hz)	31250.00
Solvent	MeOD
Spectrum Offset (Hz)	12758.7549
Sweep Width (Hz)	31249.05
Temperature (degree C)	24.970

11. References

[1] H. J. R. Maget, *J. Polym. Sci., Part A: Polym. Chem.*, 1964, **2**, 1281.
[2] E. Sachon, G. Clodic, T. Blasco and G. Bolbach, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 1880.
[3] A. Matsuo, M. Isomura, T. Miyazaki, T. Sakakibara and K. Ajisaka, *Carbohydr. Res.*, 1997, **305**, 401.
[4] P. Dais, T. K. M. Shing and A. S. Perlin, *Carbohydr. Res.*, 1983, **122**, 305.
[5] H. Ammann and G. Dupuis, *Can. J. Chem.*, 1988, **66**, 1651.
[6] Y. Zhang, J. Mallet and P. Sinaÿ, *Carbohydr. Res.*, 1991, **236**, 73.