S. Afr. J. Chem., 2014, **67**, 180–183,
<http://journals.sabinet.co.za/sajchem/>.

A Serendipitous Formation of a Cysteine-bridged Disaccharide

Mbulelo G. Nokwequ[a,*], Comfort M. Nkambule[a] and David W. Gammon[b]

[a]*Department of Chemistry, Tshwane University of Technology, Private Bag X680, Pretoria, 0001, South Africa.*

[b]*Department of Chemistry, University of Cape Town, Rondebosch, 7700, South Africa.*

Received 4 July 2014, revised 25 September, accepted 26 September.

ABSTRACT

***N*-acetyl-L-cysteine bearing free carboxylic acid and sulfhydryl groups was glycosylated with 1,2,3,4,6-Penta-*O*-acetyl-β-D-glucopyranoside in the presence of $SnCl_4$ as a promoter to give the *S*-glycosylated cysteine in 64 % yield. However, when excess donor was used, a previously unreported cysteine-bridged disaccharide was isolated in 54 % yield. The acetamido group on cysteine, which lowers the pK_a of the carboxylic acid group of the amino acid, plays no role in the formation of the bridged disaccharide since 3-mercaptopropionic acid reacts in a similar manner to give the 3-mercaptopropionic acid-bridged disaccharide in 52 % yield.**

KEYWORDS

Glycopeptides, glycosylation, bridged-disaccharides.

1. Introduction

Glycopeptides play an important role in biological systems such as cellular differentiation, cell signalling, intracellular transport of enzymes, and adhesion processes.[1] In addition, the presence of carbohydrates in glycopeptides increase the solubility of the parent peptides, protects against enzymatic degradation and can be used for the delivery of biologically active compounds to target sites.[1] The *S*-glycopeptides have particularly attracted much attention due to their chemical and enzymatic stability; preparation of these glycopeptides requires easy access to *S*-glycosylated amino acids.[2] The *S*-glycosylated amino acids have been synthesized from the condensation of 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosylisothiouronium salt and the iodide or tosyl derivatives of L-serine,[3] the desulfurization of disulfide-linked glycosyl cysteine derivatives,[4] Lewis acid-catalyzed glycosylation,[5,6] and solid phase glycosylation.[7]

Glycosylation of amino acids has previously relied on the use of amino acids protected at both the amino and carboxyl groups. This can be cumbersome as it requires several steps of protecting group manipulations which may ultimately lead to low yields. Kihlberg and co-workers have shown that 3-mercaptopropionic acid and *N*-Fmoc-L-cysteine with an unprotected carboxyl and sulfhydryl groups can be used as a glycosyl acceptors with disarmed peracetylated glycosyl donors using $BF_3{\cdot}OEt_2$ or $SnCl_4$ as Lewis acid catalysts.[8] It has been argued that the presence of a Lewis acid catalyst in this reaction promotes the rearrangement of the intermediate acyl glycoside to the target *S*-glycosylated L-cysteine derivative.[5] We have previously reported that glycosylation of a peracetylated glucopyranosyl donor with *N*-acetyl-L-cysteine bearing an unprotected carboxyl group as a glycosyl acceptor afforded *N*-acetyl *S*-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl)-L-cysteine **4** in 64 % yield.[9]

2. Results and Discussion

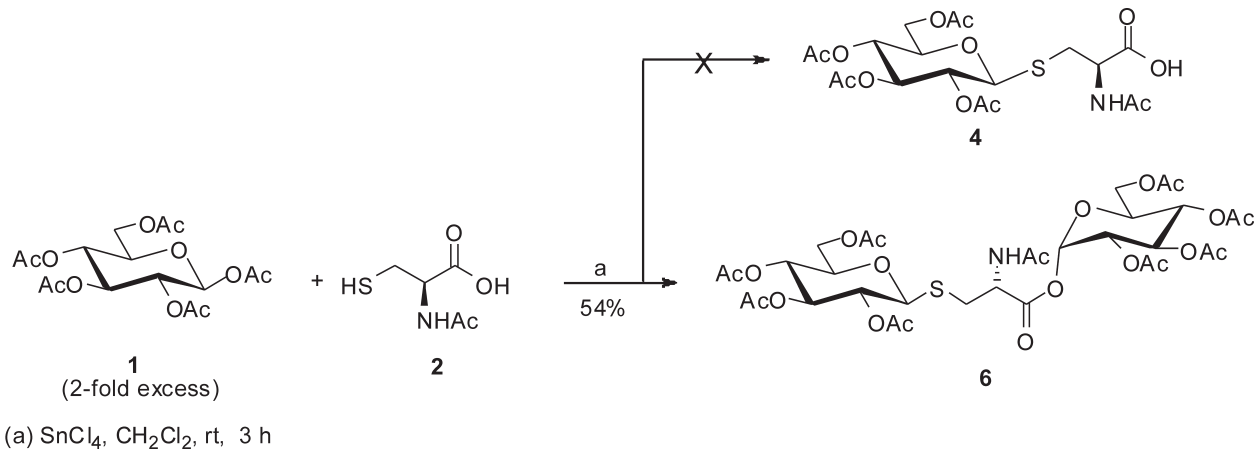
In an attempt to improve the previously reported 64 % yield of the L-cysteine *S*-glycoside **4**, a 2-fold excess of 1,2,3,4,6-Penta--*O*-acetyl-β-D-glucopyranoside **1** was used in the glycosylation of *N*-acetyl-L-cysteine. However, a previously unreported cysteine-bridged dodisaccharide **6** was isolated in 54 % yield instead of the target L-cysteine thioglycoside **4** (Scheme 1).

The presence of the cysteine-bridged disaccharide **6** was indicated by the presence of two anomeric protons at δ 6.27 (d, J = 3.6 Hz) and at δ 4.51 (d, J = 10 Hz) in the ^{1}H NMR spectrum (Fig. 1). The downfield doublet at δ 6.27 is due to the proton on the acyl glycosylated anomeric carbon and the small coupling constant of 3.6 Hz confirms the α-stereochemistry at this end of the molecule. Furthermore, the upfield doublet at δ 4.51 with a large coupling constant of 10 Hz is typical of a thioglycosydic bond with a β-stereochemistry. Other key signals in the ^{1}H NMR spectrum of **6** are: δ 6.50 (d, 1 H, J = 7.6 Hz) for the cysteinyl NH which was assigned on the basis of HMQC data, which shows that this proton does not correlate with any carbon and the COSY (H-H) which shows that this proton is only coupled to the methine on the α-carbon of the amino acid residue at δ 4.81 (m, 1H). The ^{13}C NMR spectrum displayed diagnostic signals at δ 90.3 ppm and δ 83.3. The identity of **6** was also confirmed by HRMS which showed the protonated parent peak at 824.2268 (calculated 824.2283).

It was also important to establish whether the acetamido group on cysteine which affects the pK_a of cysteine (pK_a = 3.24 compared to 4.34 for 3-mercaptopropionic acid) has any role to play in the formation of **6**. Thus, 3-mercaptopropionic acid **3** was glycosylated with **1** in the presence of $SnCl_4$ in dichloromethane at room temperature (Scheme 2). The results of this reaction showed that the product formed depended on whether an excess amount of the donor was used or not, and not on the pK_a of the mercaptopropionic acid used.

When excess donor **1** (2 equivalents) was used, the 3-mercaptopropionic acid bridged disaccharide **7** was obtained as a sole product in 52 % yield. The formation of **7** was supported by the presence of two anomeric protons at δ 5.65 (d, J = 5.6 Hz) and δ 4.50 (d, J = 9.6 Hz) in the ^{1}H NMR spectrum (Fig. 2). The

* To whom correspondence should be addressed. E-mail: nokwequmg@tut.ac.za



Scheme 1
Glucosylation of *N*-acteyl-L-cysteine.

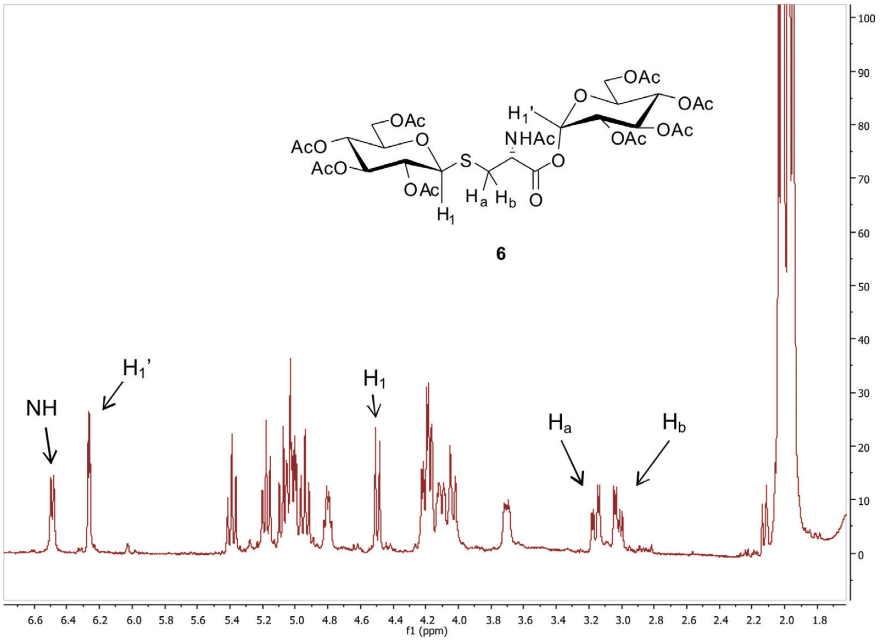
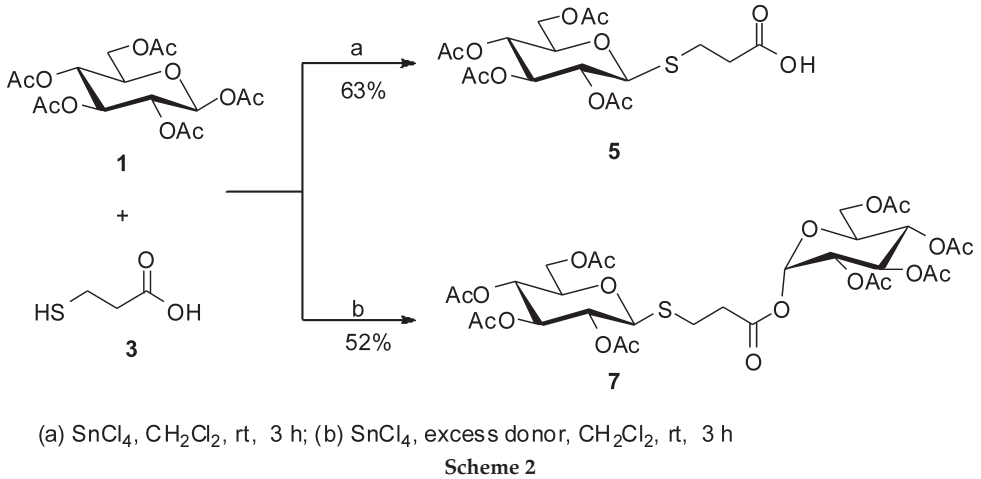


Figure 1 ^{1}H NMR spectrum of bridged disaccharide **6.**

downfield signal at δ 5.65 (d, J = 5.6 Hz) is due to the proton on the acyl glycosylated anomeric carbon and the small coupling constant (J = 5.6 Hz) confirms the α-stereochemistry at this end of the molecule (Fig. 2). The high field doublet at δ 4.69 (J = 9.2 Hz) is due to the thioglycosidic anomeric proton and the large coupling constant confirms the β-stereochemistry at this end of the molecule. Furthermore, the ^{1}H NMR spectrum displayed a multiplet at δ 2.91–2.64 for the four methylene



Scheme 2
Glucosylation of 3-mercaptopropionic acid.

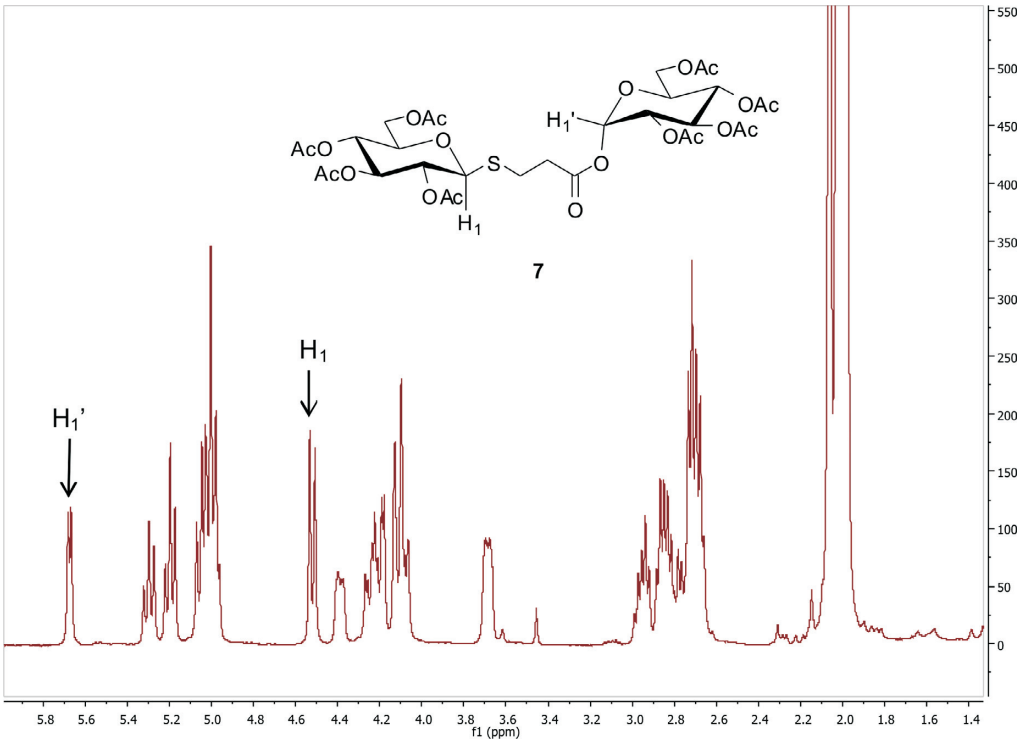


Figure 2 ^{1}H NMR spectrum of bridged disaccharide **7**.

protons on the mercaptopropionic acid link. The ^{13}C NMR spectrum displayed diagnostic signals at δ 83.9 and δ 82.7 for the two anomeric carbons. The identity of **7** was also confirmed by HRMS which showed the protonated parent peak at 767.2037 (calculated 767.2024).

In order to establish the mechanism of formation of these pseudodisaccharides, and specifically the order of glycosylation, model reactions were set-up: when cysteine thioglycoside **4** was glycosylated with **1** in the presence of $SnCl_4$ as a promoter, no glycosylation product was detected. This is in line with earlier observations that cysteine glycosyl esters are difficult to form under mild conditions. Furthermore, the glycosylation of *S*-protected *S*-benzyl-*N*-acetyl-L-cysteine with donor **1** under similar conditions failed to form any glucosyl ester. This result was not unexpected because it has been shown that glycosyl esters are difficult to form when the donor does not have a more labile leaving group at the anomeric centre of the donor.[10,11] These results suggest something unique is happening in the milieu of the reaction and we are currently performing further experiments to establish the exact order of events during the formation of these bridged disaccharides.

3. Conclusions

Overall, the formation of pseudodisaccharides **6** and **7** depend on the reaction conditions: when excess donor is used, both the sulfyhydryl and the carboxylic acid groups react to give a unique mercaptopropionic acid-bridged glycopeptide structure. These bridged disaccharides have two glucose units, one linked in an α-configuration *via* the acyl group and the other linked in a β-configuration *via* the sulfhydryl group. On the other hand, when the mercaptopropionic acids are used in excess, only the thioglucosides **4** and **5** are formed. Furthermore, we conclude that the acetamido group on L-cysteine, which affects the pK_a of cysteine (pK_a = 3.24 compared to 4.34 for 3-mercaptopropionic acid), has no role to play in the formation of the bridged disaccharide since 3-mercaptopropionic acid reacts in a similar manner to *N*-acetyl-L-cysteine. The reaction described herein represents the first reported case of a one-pot, sequential installation of two sugar units in different anomeric configurations, with one a thioglycoside and the other an acyl glycoside. We are currently conducting further experiments to elucidate the mechanism by which these pseudodisaccharides are formed.

4. Experimental

4.1. General Methods

4.1.1. *Preparative*

All reactions were carried out under an inert N_2 atmosphere. Dichloromethane was dried by distilling from P_2O_5 and all commercially available reagents were used without further purification. Reactions were monitored by TLC using Silica gel 60 UV254 (Alugram) pre-coated silica gel plates; detection was by means of a UV lamp and by heating the plate after spraying with a solution of Ceric ammonium sulfate (CAS) [Preparation: 63 g CAS dissolved in 500 mL of 6 % H_2SO_4 and diluted to 1 L mark with distilled H_2O]. Organic layers were dried over anhydrous $MgSO_4$ prior to evaporation on a Buchi rotary evaporator B-490 with a bath temperature of 40 °C. Column chromatography was carried out on Machery Nagel silica gel 60.

4.1.2. *Analytical*

IR spectra were recorded on a Perkin Elmer UATR Spectrum Two spectrophotometer. ^{1}H and ^{13}C NMR spectra were recorded on Varian Gemini 400 at ambient temperature, in $CDCl_3$. The splitting patterns are reported as follows: singlet (s), doublet (d), triplet (t), doublet of doublets (dd), multiplet (m) and broad singlet (br s). Mass spectra were obtained on a Waters Synapt G2 mass spectrometer.

N-acetyl-S-(2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl)-L-cysteine (**4**): 1,2,3,4,6-penta-*O*-acetyl-β-D-glucopyranoside (2.14 g, 5.48 mmol) and *N*-acetyl-L-cysteine (1.68 g, 10.3 mmol) were dissolved in dry CH_2Cl_2 (20 mL) under N_2 flow. This was

followed by dropwise addition of $SnCl_4$ (1.3 mL, 10 mmol). The mixture was stirred at room temperature for 3 h, then diluted with CH_2Cl_2 (20 mL) and washed with HCl solution (2 × 20 mL, 1 M). The organic phase was dried over anhydrous $MgSO_4$, filtered and concentrated. The crude was purified by column chromatography (1:9 MeOH:CH_2Cl_2) to afford a white foam (1.73 g, 64 %). The ^{1}H and ^{13}C NMR spectra data matched that of literature.[3] IR (cm^{-1}): 3289, 2957, 2731, 1745, 1678, 1543, 1457, 1375, 1228, 1043; ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.91 (d, 1H, N-H, J = 7.6 Hz); 5.20 (t, 1H, H-3, J = 9.2Hz); 5.05 (t, 1H, H-4, J = 9.6Hz); 4.94 (dd, 1H, H-2, J = 9.2, 10 Hz); 4.75 (m, 1H, H-2′); 4.56 (d, 1H, H-1, J = 10 Hz); 4.21–4.13 (m, H-6, 2H); 3.73–3.68 (m, 1H, H-5), 3.21 (dd, 1H, H-3′$_a$, J = 4.8 Hz, 14Hz); 3.07 (dd, 1H, H-3′b, J = 6Hz, 14Hz); 2.04 (s, 3H), 2.02 (s, 3H), 1.99 (s, 3H), 1.97 (s, 3H), 1.94 (s, 3H); ^{13}C (100 MHz, $CDCl_3$): δ 170.9, 170.6, 160.0, 169.9, 169.5, 169.4, 83.3, 76.2, 73.5, 69.7, 68.0, 61.8, 31.7, 20.7, 20.5; HR-ESIMS (*m*/*z*) calculated for $C_{19}H_{28}NO_{12}S$ (M+H$^+$): 494.1332; found: 494.1335

S-*(2,3,4,6-tetra*-O-*acetyl-β-D-glucopyranosyl)-mercaptopropionic acid* (**5**) (procedure same as for **4**; 63 % yield): IR (cm^{-1}): 2987, 2731, 1741, 1702, 1532, 1446; ^{1}H NMR (400 MHz, $CDCl_3$): 5.22 (dd, 1H, H-3, J = 9.2, 9.6 Hz); 5.06 (t, 1H, H-2, J = 9.2, 10.4 Hz); 5.02 (dd, 1H, H-4, J = 9.2, 10 Hz); 4.54 (d, 1H, H-1, J = 10.4 Hz), 4.21 (m, 1H), 4.15(m, 1H), 3.72 (m, 1H), 2.99–2.84 (m, 4H); ^{13}C (100 MHz, $CDCl_3$): δ 171.8, 170.8, 169.9, 169.5, 169.4, 82.8 (C-1), 73.5, 72.2, 69.5, 69.2, 68.7, 61.8, 35.4, 25.4, 20.70, 20.69, 20.62, 20.6

N-*acetyl*-S-*(2,3,4,6-tetra*-O-*acetyl-β-D-glucopyranosyl)-O-(2,3,4,6-tetra*-O-*acetyl-α-D-glucopyranosyl)-L-cysteinoate* (**6**): 1,2,3,4,6-penta-*O*-acetyl-*β*-D-glucopyranoside (2.00 g, 5.12 mmol) and *N*-acetyl-L-cysteine (0.42 g, 2.6 mmol) were dissolved in dry CH_2Cl_2 (20 mL) under N_2 flow. This was followed by dropwise addition of $SnCl_4$ (0.7 mL, 6 mmol). The mixture was stirred at room temperature for 3 h, then diluted with CH_2Cl_2 (20 mL) and washed with HCl solution (2 × 20 mL, 1 M). The organic phase was dried over anhydrous $MgSO_4$, filtered and concentrated. The crude was purified by column chromatography (1:9 MeOH: CH_2Cl_2) to afford **6** as a clear oil (1.12 g, 54 %): R_f = 0.22 (1:9 MeOH: CH_2Cl_2); IR (cm^{-1}): 2950, 2731, 1742, 1735, 1675, 1542, 1457, 1375, 1228; ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.49 (d, 1H, N-H, J = 7.6Hz), 6.26 (d, 1H, H-1, J = 3.6 Hz), 5.39 (t, 1H, J = 10 Hz), 5.18 (dd, 1H, J = 9.2 Hz, 9.6 Hz), 5.10–5.00 (m, 3H), 4.96 (t, 1H, J = 9.6 Hz), 4.81 (m, 1H, H-2′), 4.51 (d, 1H, H-1, J = 10 Hz), 4.20–4.02 (m, 5H), 3.70 (m, 1H), 3.16 (dd, 1H, H-3′$_a$, 4.8 Hz, 14.4 Hz), 3.02 (dd, 1H, H-3′$_b$, J = 6.4 Hz, 14.4 Hz), 2.04–1.95 (m, 28H); ^{13}C (100 MHz, $CDCl_3$): δ 170.6, 170.0, 169.9, 169.7, 169.4, 169.3, 168.8, 90.3 (C-1), 83.3 (C-1), 76.2, 73.4, 70.1, 69.6, 68.0, 61.8, 61.3, 52.2, 31.3, 29.7, 22.7, 22.66, 20.7, 20.65, 20.6, 20.55, 20.5; HR-ESIMS (*m*/*z*) calculated for $C_{33}H_{46}NO_{21}S$ (M+H$^+$): 824.2283; found: 824.2268

S-*(2,3,4,6-tetra*-O-*acetyl-β-D-glucopyranosyl)-O-(2,3,4,6-tetra*-O-*acetyl-α-D-glucopyranosyl) mercaptopropionoate* (**7**) (procedure same as for **6**; clear oil, 52 % yield): R_f = 0.15 (1:19 MeOH: CH_2Cl_2); IR (cm^{-1}): 2980, 2731, 1728, 1720, 1542, 1462; ^{1}H NMR (400 MHz, $CDCl_3$): δ 5.65 (d, 1H, H-1, J = 5.6 Hz), 5.26 (dd, 1H, J = 9.6 Hz, 10 Hz), 5.16 (dd, 1H, J = 9.6 Hz, 10 Hz), 5.03–4.92 (m, 4H), 4.50 (d, 1H, H-1, J = 9.6 Hz), 4.36 (m, 1H), 4.19 (m, 2H), 4.07 (dd, 2H, J = 12.4 Hz, 13.2Hz), 3.85 (m, 1H), 2.91 (m, H-2′, 1H), 2.70 (m, H-2′, H-3′, 4H), 2.03–1.94 (m, 25H); ^{13}C (100 MHz, $CDCl_3$): δ 176.9, 176.8, 170.8, 170.7, 170.2, 169.9, 169.6, 169.4, 83.3 (C-1), 82.6 (C-1), 76.7, 75.9, 73.9,73.7, 70.6, 70.3, 69.6, 68.5, 68.3, 67.7, 62.1, 61.9, 35.2, 34.5, 25.3, 25.1, 20.7, 20.65, 20.62, 20.58, 20.56; HR-ESIMS (*m*/*z*) calculated for $C_{31}H_{43}NO_{20}S$ (M+H$^+$): 767.2024, found 767.2037

Acknowledgements

We acknowledge the institutional and financial support of the Tshwane University of Technology and the financial support from the National Research Foundation (NRF-South Africa) for Thuthuka Grant for Researchers in Training (GUN 66173) to MGN and Human and Institutional Capacity Development Programmes grants (IRDP GUN 62464) to CMN.

Supplementary Material

^{1}H, ^{13}C, COSY, HMQC and high resolution mass spectra of bridged disaccharides **6** and **7** are provided as supplementary material

References

1 Varki, A. *Glycobiology* 1993, **3**, 97–130.
2 Z.J. Witczak, R. Chhabra, H. Chen and X.-Q. Xie, *Carbohydr. Res.*, 1997, **301**, 167–175.
3 M.L.P. Monsigny, D. Delay and M.Vaculik, *Carbohydr. Res.*, 1977, **59**, 589–593.
4 G.J.L. Bernardes, E.J. Grayson, S. Thompson, J.M. Chalker, J.C. Errey, F. El Oualid, T.D.W. Claridge and B.G. Davis, *Angew. Chem. Int. Ed.*, 2008, **47**, 2244–2247.
5 L.A. Salvador, M. Elofsson and J. Kihlberg, *Tetrahedron*, 1995, **51**, 5643–5656.
6 L. Käsbeck and H. Kessler, *Liebigs Ann.*, 1997, **1997**, 165–167.
7 L. Jobron and G. Hummel, *Org. Lett.*, 2000, **2**, 2265–2267.
8 M. Elofsson, B. Walse and J. Kihlberg, *Tetrahedron Lett.*, 1991, **32**, 7613–7616.
9 M.G. Nokwequ, C.M. Nkambule and D.W. Gammon, *Carbohydr. Res.*, 2012, **359**, 18–23.
10 E. Valepyn, J. Nys, A. Richel, P. Laurent, N. Berezina, O. Talon and M. Paquot, *Biocatal. Biotransfor.*, 2011, **29**, 25–29.
11 G.H. Veeneman, S.H. van Leeuwen and J.H. van Boom, *Tetrahedron Lett.*, 1990, **31**, 1331–1334.