

Expedient Synthesis of O-Glycosylated amino acids

Felicity J. Frank,^a Rebecca A. Lawson^a and Tom E. McAllister^{a*}

Protein glycosylation is the most abundant and complex post-translational modification, necessitating many different approaches to fully understand the biological effects. Investigation using synthetic glycopeptides is limited by the high cost of building blocks; typically >100x more than other modified amino acids e.g. phosphorylation. We report a simple, low cost route to O-glycosylated amino acids suitable for Fmoc-SPPS in 2 steps starting from peracetylated sugars. One set of reagents can furnish either the α - or β -anomer through adjusting the concentration, equivalents and reaction time. Depending on the derivative, the cost of our route is 25-60x less than commercial alternatives and offers scope for producing modified analogues. Overall, this is a convenient and user friendly approach to access O-glycosylated amino acids, urgently required for continued investigation of the manifold roles of glycosylation in biology.

Protein glycosylation is an abundant post-translational modification,¹ with glycans linked to specific amino acid side chains and categorised by the linking atom with C-, O-, N- and S-glycans all observed.² The O-glycans are the most diverse with the covalent attachment occurring through serine (Ser) and threonine (Thr) (as well as reports of tyrosine, Tyr) sidechains. They are subdivided by the identity of initial monosaccharide, which can be glucose (Glc), galactose (Gal), mannose (Man), fucose (Fuc), xylose (Xyl), N-acetylglucosamine (GlcNAc) or N-acetylgalactosamine (GalNAc). Canonical α -O-GalNAc glycosylation is the most abundant and complex form, occurring on secreted and membrane proteins in dense clusters in mucin domains forming the principal component of the mucosal layer in the gut.³ Simultaneously, α -O-GalNAc glycans occur at single sites with specific functions including cellular communication, regulation of protein half-life and host pathogen interactions.⁴ This apparent dichotomy between both generic and specific functions is one of the pressing areas of research for O-GalNAc glycans; there are likely further specific examples yet to be identified.⁵ Dysregulation of the GalNAc transferases and production of truncated glycans is associated with various disease phenotypes including cancer.⁶⁻⁸

A common approach to studying these modifications is through production of synthetic glycopeptides e.g. for *in vitro* enzyme reactions^{9,10} or characterising binding.¹¹ While some O-GalNAc glycosylated amino acid building blocks for Fmoc-SPPS are commercially available, they are typically very expensive. From a survey of UK online prices, they are >7000x more expensive on a molar basis than their non-glycosylated counterparts and >100x more than derivatives of other common modifications

e.g. phosphorylation (details in ESI). Approaches have thus been developed to minimise the quantities needed but this is not feasible for all applications.¹² The high costs are a barrier to progress and particularly likely to deter non-specialist researchers from venturing into the area. Chemical synthesis requires formation of the glycosidic bond and many methods exist though the low reactivity of oxygen nucleophiles makes O-glycosylation more challenging. More generally, a major challenge in carbohydrate chemistry is achieving selective formation of 1,2-*cis*-glycosides (α -stereochemistry in gluco- and galacto-configured pyranoses), as neighbouring-group participation (NGP) from C(2) typically directs glycosylations to the β - (1,2-*trans*) configuration.¹³ This is not always the case and there appear to be more subtleties and nuances to this process than previously realised.¹⁴ There are many existing syntheses (>30) to Fmoc-Thr(GalNAc(Ac)₃- α -D]-OH **1**, recently reviewed by Liu *et al.*¹⁵, but on average these require at least seven steps from commercial reagents and have varying degrees of control over the anomer formed.

As part of our ongoing work to investigate protein O-GalNAc glycosylation, we set out to synthesise the building blocks Fmoc-Thr[GalNAc(Ac)₃- α -D]-OH **1** and Fmoc-Ser[GalNAc(Ac)₃- α -D]-OH **2** (Figure 1).

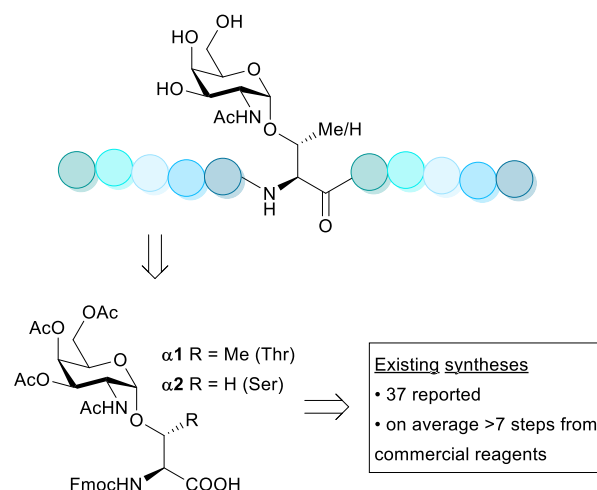


Figure 1: Glycopeptides are useful reagents for understanding protein glycosylation but require compatible building blocks for Fmoc-SPPS. Reported syntheses of **1** were recently reviewed Liu *et al.*¹⁵

We initially deployed the Ni-catalysed synthesis reported by Yu *et al.*¹⁶ using a C(2) imine to overcome NGP but found preparation of the appropriate donor lengthy and the glycosylation inconsistent in our hands. Eager to pursue a shorter synthesis, we investigated the ferric chloride-catalysed reaction reported by Wei *et al.*¹⁷ using commercially available N-acetylgalactosamine tetraacetate β -GalNAc(Ac)₄ **3** and Fmoc-Ser-OME **4** in refluxing 1,2-dichloroethane (DCE). Using

^a School of Natural and Environmental Science, Newcastle University, Newcastle upon Tyne, NE1 7RU.

* Corresponding author: tom.mcallister@newcastle.ac.uk

Supplementary Information available: See DOI: 10.1039/x0xx00000x

this method, we could only produce small quantities of the β -anomer as product. A similar procedure was more recently described by Sommer *et al.*¹⁸ using copper(II) triflate ($\text{Cu}(\text{OTf})_2$) to catalyse glycosylation of simple alcohols with *N*-acetylglucosamine tetraacetate (β -GlcNAc(Ac_4)) **5** in refluxing DCE. This reaction was shown to be stereodivergent as either the α - or β -anomers could be produced using the same reagents but under different reaction conditions; shorter reaction times giving predominantly β -product with more α -product from prolonged reaction times.

In this paper we detail our work to explore the scope of this reaction for producing either anomer for both GalNAc and GlcNAc glycosylated threonine and serine amino acids, to generate building blocks suitable for Fmoc-SPPS. While primarily targeting GalNAc-modified amino acids, the capacity to also use GlcNAc donors would make the reaction more versatile. β -O-GlcNAc addition to serine and threonine is an abundant dynamic modification of intracellular proteins¹⁹ and while β -O-GalNAc and α -O-GlcNAc are not known protein modifications, the potential for UPD-GlcNAc to be used by human GalNAc transferases (generating Ser/Thr- α -O-GlcNAc) has been shown,²⁰ and in other species such as trypanosomes, GlcNAc is used exclusively in place of GalNAc in mucin-type glycans.²¹ Thus, facile access to all these derivatives would be a timely advance to facilitate further study of protein O-glycosylation.

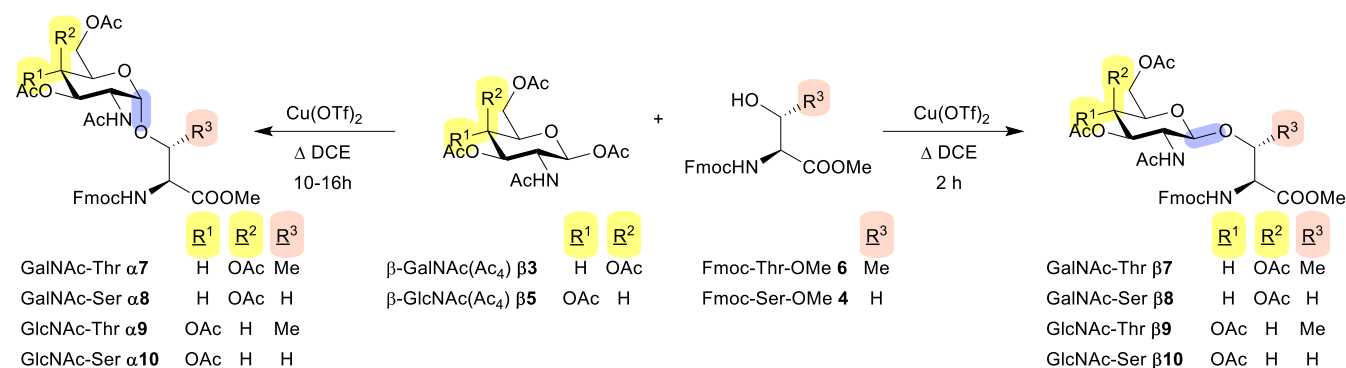
We initially used commercially available Fmoc-Thr-OH (without sidechain protection) as good results had been shown previously with Fmoc-Ser-OH in an indium(III) bromide-catalysed glycosylation yielding exclusively β -anomers.²² while we did observe formation of new products, separation from unreacted acceptor was laborious and anomers proved impossible to resolve (data not shown). Hence, we elected to use esterified amino acids; Fmoc-Ser-OMe **4** is commercially available while Fmoc-Thr-OMe **6** was synthesised from the corresponding methyl ester hydrochloride salt quantitatively (details in ESI).

Initial results were very encouraging; glycosylation of Fmoc-Thr-OMe **6** (5 equiv.) with β -GalNAc(Ac_4) **3** in the presence of $\text{Cu}(\text{OTf})_2$ (1 equiv.) for 1.6 h yielded the β -product Fmoc-Thr(GalNAc(Ac_3)- β -D)-OMe **7** in 82% isolated yield (Table 1,

entry 1). Likewise, the corresponding reaction with serine acceptor **4** gave 66% isolated yield of Fmoc-Ser(GalNAc(Ac_3)- β -D)-OMe **8** (Table 1, entry 2). The corresponding reaction with β -GlcNAc(Ac_4) **5** gave similar yields for Fmoc-Thr(GlcNAc(Ac_3)- β -D)-OMe **9** and Fmoc-Ser(GlcNAc(Ac_3)- β -D)-OMe **10** (Table 1, entries 3-4).

Following the successful isolation of **7-10**, our investigation turned to the synthesis of the equivalent α isomers. Initially, **3** was reacted with 5 equiv. **4** for 16 hours, showing the formation of one major product by TLC. However, after isolation of this compound, it was determined that both products Fmoc-Ser[GalNAc(Ac_3)- α -D]-OMe **8** and **9** have identical R_f values, resulting in an inseparable 1:1 ratio of the two products (Table 1, entry 5). Therefore, a series of screening reactions were carried out to selectively produce **8**. First, we reduced the equivalents of acceptor **4** to 3.5 and 2 equiv. relative to **3**, while maintaining the 1 equiv. of $\text{Cu}(\text{OTf})_2$. After 16 hours reflux in DCE, both reactions showed only **8** as determined by LCMS. However, the yield was so low that **8** could not be isolated. Interestingly, the LCMS also showed a major peak consistent (based on observed mass) with Fmoc-Ser(Ac)-OMe **11** (Figure S1, ESI). Acetate is liberated from the donor and is presumably able to compete for reaction with the acceptor, which is in excess. Close inspection of previous LCMS reaction monitoring data also showed varying (minor) amounts of acetylated amino acid by product in previous reactions with Fmoc-Thr-OMe as well (Figure S2, ESI). We postulated that a shorter reaction time may limit this side reaction occurring, therefore **3** was reacted with 2 equiv. of **4** for 10 hours. Isolation of **8** and **9** in a 3:1 ratio was observed, suggesting that the shorter reaction time did indeed limit the competing acetylation (Table 1, entry 6). Further reduction in acceptor **4** to 1 equiv. enhanced α selectivity, resulting in 12% isolated yield of **8** (Table 1, entry 7) without any co-production of **9**. Although this was an excellent result for selectivity, a reduction in yield was observed, therefore a final screening reaction was carried out using 5 equiv. **3** and 1 equiv. **4** to increase yields of **3** further (Table 1, entry 8). Unfortunately, although LCMS indicated the formation **8**, only trace amounts were formed that could not be isolated with a major product of Fmoc-Ser(Ac)-OMe **11**.

Table 1: Exploration of glycosylation reaction.



Entry ^[a]	Donor	Acceptor	Equiv. acceptor	Time/h	Product (yield/%)	α:β ratio
1	β3	6	5	1.6	β7 (82)	β selective ^[f]
2	β3	4	5	1.6	β8 (66)	β selective ^[f]
3	β5	6	5	1.6	β9 (57)	β selective ^[f]
4	β5	4	5	1.6	β10 (68)	β selective ^[f]
5	β3	4	5	16	β8 + α8 (68) ^b	1:1
6	β3	4	2	10	β8 + α8 (34) ^b	3:1
7	β3	4	1	10	α8 (12)	α selective ^[f]
8	β3	4 ^[c]	0.2	10	β8 + α8 ^[e]	2:1
9	β3	4	0.2	10	α8 ^[e]	α selective ^[f]
10	β5	4	1	10	β10 (19)	β selective ^[f]
11	β5	4	1	24	β10 + α10 (9) ^[b]	2:1
12	β3	6	5	16	α7 (39), β7 (20)	1.9:1
13	β5	6	5	16	α9 (10), β9 (29)	1:2.7
14	β5	6	1	16	α9(6), β9(7)	1:1.2
15 ^[d]	β3	6	5	16	α7 (32)	N.D.
16 ^[d]	β3	6	5	1.6	β7 (75)	N.D.
17 ^[d]	β3	4	1	14	α8 (3)	N.D.
18 ^[d]	β3	4	5	1.6	β8 (95)	N.D.

[a] Unless otherwise stated all reactions were performed with 258 μmol of donor, 1 equiv. of promoter Cu(OTf)₂ and 5 equiv. of acceptor (equiv. related to donor) at 51.6 mM [donor] in refluxing DCE; [b] α and β products were inseparable via silica gel column chromatography, ratio determined by ¹H NMR; [c] 0.2 equiv. of Cu(OTf)₂ was used in this reaction; [d] Reactions were performed on 5.15 mmol scale at 103 mM [donor]; [e] yield too low to recover, product identified by LCMS; [f] none of the other anomer was observed; N.D. – not determined.

After the successful synthesis of α8, these same optimised reaction conditions were utilised, using β5 as an alternative donor (table 1, entry 10). After 10 h, only β10 was isolated, indicating the formation of α10 is significantly slower than α8. The lower reactivity of glucopyranosyl donors relative to galactopyranosyl donors has been reported previously²³ though the apparent complete lack of α10 was unexpected. Extension of the reaction time to 24 hours resulted in a 9% combined yield of α10 and β10 in a 2:1 ratio, calculated by NMR (Table 1, entry 11). Despite further attempts we could not synthesise α10 without co-production of β10 and we were unable to separate them by standard phase chromatography; we conclude that α10 cannot be produced directly by this method in our hands. Next, we investigated the corresponding threonine products. Donor β3 was refluxed with 5 equiv. of 6 for 16 hours, resulting in the formation of a mixture with 39% yield of Fmoc-Thr(GalNAc(Ac)₃-α-D)-OMe α9 and 20% yield β9 isolated; the anomers proving easily separable by silica gel column chromatography (Table 1, entry 12). Further, β5 was reacted with 5 equiv. of 6 for 16 h resulting in 10% and 29% isolated

yields of Fmoc-Thr(GlcNAc(Ac)₃-α-D)-OMe α9 and β9 respectively (Table 1, entry 13). This again indicates that GlcNAc donor β5 reacts slower than its GalNAc counterpart β3. Finally, β5 was reacted with 1 equiv. 6, reasoning this may increase the ratio of α9 (Table 1, entry 14) as was seen previously for α10. Although this resulted in less β9, the yield of the desired α-anomer α9 had also decreased, therefore it was determined that 5 equiv. of acceptor 6 was more suitable in this case as the α- and β- products are separable on silica. To explore the scalability of the reaction, we next performed the glycosylation on 20× larger scale (5 mmol of donor β3) with the previously optimised conditions to yield α7, β7, α10 and β10 in comparable yields to previously (entries 15-18).

While demonstrating that both α- and β- anomers can be produced using this method, yields for the β-anomer were significantly higher (up to 95%; Table 1, entry 18). Sommer and coworkers reported that β-GlcNAc glycosides could be anomerised to the corresponding α-glycosides by refluxing in DCE with 0.05 equiv. of Cu(OTf)₂ and 1 equiv. HOAc.¹⁸ We

attempted anomerization of **β7** using identical conditions, however conversion to **α7** was not observed, showing only degradation to unidentifiable compounds. We subsequently investigated further using a sample containing both **α7** and **β7** (1:7), with 1 equiv. of Cu(OTf)₂ and monitored via LCMS with UV detection at 280 nm. Over the first 5 hours we observed disappearance of **β7** with concomitant generation of amino acid **6**; apparently driving the reaction backwards to the starting materials.* In contrast, the amount of **α7** was unchanged throughout this period, but upon prolonged reaction times (23 h) the proportion of **α7** did increase, suggesting this anomerisation is possible but we also observed formation additional products not due to **6**, **α7** or **β7**. Our data suggest that for GalNAc-derivatives anomerisation proceeds *via* cleavage of the exocyclic bond in the β-glycoside, as has been previously proposed by Ikemoto *et al.*²⁴ Further evidence came from immediate quenching of a sample from the reaction mixture with excess methanol, which showed formation of the methyl-galactoside (from LCMS data; Figure S3, ESI), whose formation only possible if the exocyclic bond is broken.

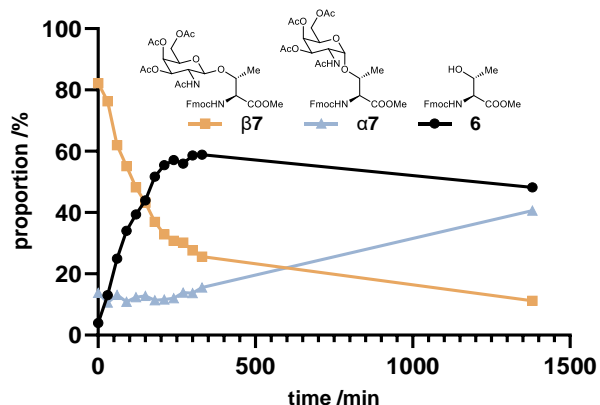


Figure 2: Product distribution over time following treatment of a 7:1 **β7**/**α7** mixture to 1 equiv. Cu(OTf)₂ in refluxing DCE as determined by LCMS (UV absorbance at 280 nm).

Finally, to furnish building blocks suitable for Fmoc-SPPS the methyl esters were removed using the Lil conditions reported by Mayato and coworkers.²⁵ Aqueous workup yielded the corresponding amino acids without the need for additional purification, showing selective demethylation in up to 85% yield (Table 2).

Table 2: Selective methyl ester removal.

Entry	R ¹	R ²	R ³	α/β	Starting Material	Product	yield/%
1	H	OAc	Me	α	α7	α1	44
2	H	OAc	Me	β	β7	β1	84
3	H	OAc	H	α	α8	α2	61
4	H	OAc	H	β	β8	β2	85
5	OAc	H	Me	β	β9	β12	58
6	OAc	H	H	β	β10	β13	66

In conclusion, we present a simple two- or three-step synthesis (for Ser and Thr derivatives respectively) to synthesise Fmoc-SPPS compatible building blocks in up to 68% yield overall. Yields are variable between different products, but this route is significantly more economical than commercially purchased products, using cheap commercially available materials in only two or three steps with only one chromatographic purification required.

For comparison, we calculated the cost of synthesising 100 mg of both Fmoc-Thr(GalNAc(Ac)3-α-D]-OH **α1** and Fmoc-Ser(GalNAc(Ac)3-α-D]-OH **α2** using methods outlined above as being £6.11 and £7.64 respectively, costing 1.5% and 3.5% as much as the cheapest commercial options we could find (based on list prices – see ESI for details).[†] Furthermore, the β-anomers of both GalNAc and GlcNAc on serine and threonine can also be produced cheaply (£1.83 - £3.93, per 100 mg), demonstrating access to a range of building blocks. Thus, we are hopeful that this new route will make glycopeptide synthesis more accessible/affordable and enable further advancements in our understanding of the myriad roles of protein glycosylation.

Author Contributions

CRediT: **FJF** conceptualization, methodology, investigation, formal analysis, data curation, visualisation, writing-original draft, writing-review & editing

RAL methodology, investigation, formal analysis, data curation

TEM conceptualization, methodology, writing-original draft, writing-review & editing, visualisation, supervision, project administration, funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the Supplementary Information.

Notes and references

*a new peak for the GalNAc portion of the molecule was also observed by MS but gives no UV signal so was could not be quantified; data not shown

‡ Our price includes all reagents and reaction solvents but not labour, solvent used for workup and purification or potential reductions in costs through recovery of unreacted acceptor.

- 1 H. J. An, J. W. Froehlich and C. B. Lebrilla, *Curr Opin Chem Biol*, 2009, **13**, 421–426.
- 2 R. G. Spiro, *Glycobiology*, 2002, **12**, 43R–56R.
- 3 T. Lang, G. C. Hansson and T. Samuelsson, *Proc Natl Acad Sci U S A*, 2007, **104**, 16209–16214.
- 4 H. H. Wandall, M. A. I. Nielsen, S. King-Smith, N. de Haan and I. Bagdonaite, *FEBS J*, 2021, **288**, 7183–7212.
- 5 M. de las Rivas, E. Lira-Navarrete, T. A. Gerken and R. Hurtado-Guerrero, *Curr Opin Struct Biol*, 2019, **56**, 87–96.
- 6 E. Scott, K. Hodgson, B. Calle, H. Turner, K. Cheung, A. Bermudez, F. J. G. Marques, H. Pye, E. C. Yo, K. Islam, H. Z. Oo, U. L. McClurg, L. Wilson, H. Thomas, F. M. Frame, M. Orozco-Moreno, K. Bastian, H. M. Arredondo, C. Roustan, M. A. Gray, L. Kelly, A. Tolson, E. Mellor, G. Hysenaj, E. A. Goode, R. Garnham, A. Duxfield, S. Heavey, U. Stopka-Farooqui, A. Haider, A. Freeman, S. Singh, E. W. Johnston, S. Punwani, B. Knight, P. McCullagh, J. McGrath, M. Crundwell, L. Harries, D. Bogdan, D. Westaby, G. Fowler, P. Flohr, W. Yuan, A. Sharp, J. de Bono, N. J. Maitland, S. Wisnovsky, C. R. Bertozzi, R. Heer, R. H. Guerrero, M. Daugaard, J. Leivo, H. Whitaker, S. Pitteri, N. Wang, D. J. Elliott, B. Schumann and J. Munkley, *Oncogene* 2023 **42**:12, 2023, **42**, 926–937.
- 7 S. Julien, P. A. Videira and P. Delannoy, *MDPI AG*, 2012, preprint, DOI: 10.3390/biom2040435.
- 8 T. Ju, V. I. Otto and R. D. Cummings, *Angewandte Chemie International Edition*, 2011, **50**, 1770–1791.
- 9 A. M. Collette, S. A. Hassan, S. I. Schmidt, A. J. Lara, W. Yang and N. L. Samara, *Sci Adv*, DOI:10.1126/sciadv.adj8829.
- 10 M. R. Pratt, H. C. Hang, K. G. Ten Hagen, J. Rarick, T. A. Gerken, L. A. Tabak and C. R. Bertozzi, *Chem Biol*, 2004, **11**, 1009–1016.
- 11 M. De Las Rivas, E. J. Paul Daniel, H. Coelho, E. Lira-Navarrete, L. Raich, I. Compañón, A. Diniz, L. Lagartera, J. Jiménez-Barbero, H. Clausen, C. Rovira, F. Marcelo, F. Corzana, T. A. Gerken and R. Hurtado-Guerrero, *ACS Cent Sci*, 2018, **4**, 1274–1290.
- 12 A. Y. Mehta, R. K. H. Veeraiah, S. Dutta, C. K. Goth, M. S. Hanes, C. Gao, K. Stavenhagen, R. Kardish, Y. Matsumoto, J. Heimbürg-Molinario, M. Boyce, N. L. B. Pohl and R. D. Cummings, *Cell Chem Biol*, 2020, **27**, 1207–1219.e9.
- 13 H. L. Frush and H. S. Isbell, *J Res Natl Bur Stand (1934)*, 1941, **27**, 413.
- 14 P. Basu and D. Crich, *J Am Chem Soc*, 2025, **147**, 5808–5818.
- 15 W. Liu, P. He, S. Shang and Z. Tan, *J Carbohydr Chem*, 2023, **42**, 223–251.
- 16 F. Yu, M. S. McConnell and H. M. Nguyen, *Org Lett*, 2015, **17**, 2018–2021.
- 17 G. Wei, X. Lv and Y. Du, *Carbohydr Res*, 2008, **343**, 3096–3099.
- 18 R. Sommer, D. Hauck and A. Titz, *ChemistrySelect*, 2017, **2**, 4187–4192.
- 19 L. Wells, K. Vosseller and G. W. Hart, *Science (1979)*, 2001, **291**, 2376–2378.
- 20 P. Both, A. P. Green, C. J. Gray, R. Šardžik, J. Voglmeir, C. Fontana, M. Austeri, M. Rejzek, D. Richardson, R. A. Field, G. Widmalm, S. L. Flitsch and C. E. Eyers, *Nat Chem*, 2014, **6**, 65–74.
- 21 L. Mendonça-Previato, L. Penha, T. C. Garcez, C. Jones and J. O. Previato, *Glycoconj J*, 2013, **30**, 659–666.
- 22 C. A. De Leon, G. Lang, M. I. Saavedra and M. R. Pratt, *Org Lett*, 2018, **20**, 5032–5035.
- 23 Z. Zhang, I. R. Ollmann, X. S. Ye, R. Wischnat, T. Baasov and C. H. Wong, *J Am Chem Soc*, 1999, **121**, 734–753.
- 24 N. Ikemoto, O. K. Kim, L. C. Lo, V. Satyanarayana, M. Chang and K. Nakanishi, *Tetrahedron Lett*, 1992, **33**, 4295–4298.
- 25 C. Mayato, R. L. Dorta and J. T. Vázquez, *Tetrahedron Lett*, 2008, **49**, 1396–1398.