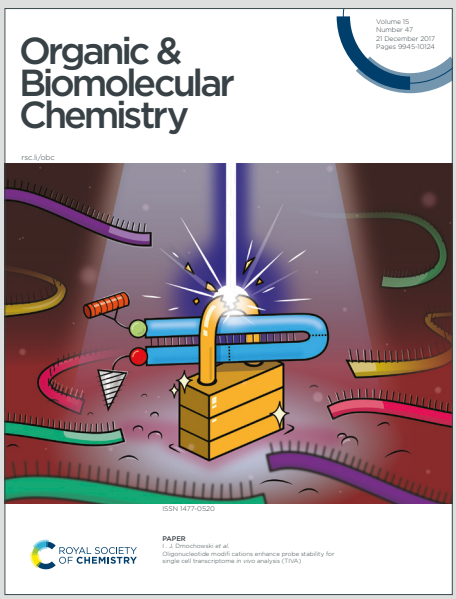
Organic & Biomolecular Chemistry

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: J. Gao, P. Lin, S. T. Nick, K. Liu, K. Yu, E. Hohenester and X. Huang, *Org. Biomol. Chem.*, 2021, DOI: 10.1039/D1OB00317H.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/obc

View Article Online
DOI: 10.1039/D1OB00317H

Exploration of Human Xylosyltransferase for Chemoenzymatic Synthesis of Proteoglycan Linkage Region

Jia Gao,[a] Po-han Lin,[a] Setare Tahmasebi Nick,[a] Kunli Liu,[a] Kefei Yu,[b] Erhard Hohenester,[c] Xuefei Huang[a,d*]

[a]Department of Chemistry and Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, MI 48824 USA

[b]Department of Microbiology & Molecular Genetics, Michigan State University, East Lansing, MI 48824 USA

[c]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

[d]Department of Biomedical Engineering, Michigan State University, East Lansing, Michigan 48824, USA

*Email: huangxu2@msu.edu

Abstract

Proteoglycans (PGs) play important roles in many biological processes including tumor progression, cell adhesion, and regulation of growth factor activities. With glycosaminoglycan chains attached to the core proteins in nature, PGs are highly challenging synthetic targets due to the difficulties in integrating the sulfated glycans with the peptide backbone. To expedite the synthesis, herein, the utility of human xylosyltransferase I (XT-I), the enzyme responsible for initiating PG synthesis, has been explored. XT-I was found to be capable of efficiently installing the xylose unit onto a variety of peptide structures on mg scales. Furthermore, an unnatural sugar, i.e., 6-azidoglucose can be transferred by XT-I introducing a reactive handle onto the glycopeptide for selective functionalization. XT-I can be coupled with β-4-galactosyl transferase-7 for one pot synthesis of glycopeptides bearing galactose-xylose disaccharide, paving the way toward efficient chemoenzymatic synthesis of PG glycopeptides and glycoproteins.

Keywords

Carbohydrates; Enzymes; Glycopeptides; Glycosylation; Synthesis Design

View Article Online
DOI: 10.1039/D1OB00317H

Proteoglycans (PGs) are an essential class of glycoproteins that are ubiquitous in the mammalian systems. They are directly involved in numerous biological processes including tumor progression, cell adhesion, and regulation of growth factor activities.[1-3] Structurally, PGs consist of a core protein and one or more glycosaminoglycan (GAG) chains, which are linked through glucuronic acid (GlcA)-β-1,3-galactose (Gal)-β-1,3-Gal-β-1,4-xylose (Xyl) tetrasaccharide linkage region attached to serine residues of serine-glycine dipeptides.[4] Due to the complexity of post-translational modifications on the GAG chains, PGs from natural sources are highly heterogeneous. To date, structurally defined proteoglycan glycopeptides can only be prepared through chemical synthesis. However, the chemical synthesis process is highly challenging and tedious owing to the presence of many sensitive functional groups, thus requiring meticulous designs of the protective group strategy and the synthetic route.[5-7] To expedite the PG preparations, we have become interested in developing a synthetic strategy deploying the enzymes involved in biosynthetic assembly of the tetrasaccharide linkage. Herein, we report our results on the utility of human xylosyltransferase I (XT-I), the enzyme responsible for initiating PG synthesis in nature.

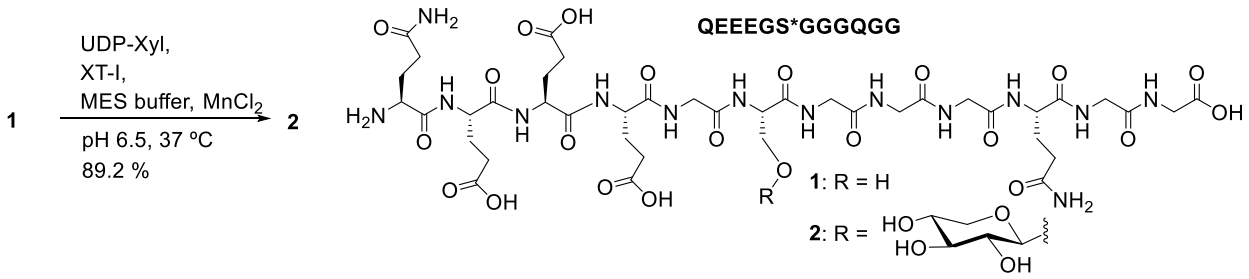
XT-I natively transfers a Xyl unit from a UDP-Xyl donor to the side chain of certain serine residues in the PG core protein.[8-10] A consensus preferred sequence for peptide acceptors has been deduced as Gly-Ser-Gly or Ser-Gly-x-Gly (x being any natural amino acid) with acidic residues commonly present near the GAG attachment site.[8, 11, 12] Till now, XT-I has not been utilized for synthetic studies of the PG. We report for the first time that human XT-I enzyme can be used to efficiently synthesize native xylosylated PG glycopeptides at milligram scales, and the combination of XT-I with β-4-galactosyl transferase 7 (β4GalT7)[13-15] enabled one pot synthesis of glycopeptides bearing Gal-Xyl disaccharides. Moreover, we have investigated XT-I donor promiscuity. Its ability to transfer an unnatural donor such as 6-azidoglucose (6AzGlc) opens the door to introduce a biorthogonal handle to label peptide and protein substrates.

To explore the synthetic potential and capability of XT-I, we selected a bikunin-like peptide sequence QEEEGSGGGQGG as the initial peptide substrate.[16, 17] The preparation of QEEEGSGGGQGG was achieved with Fmoc-based solid-phase peptide synthesis (SPPS) using Cl-MPA ProTide resin under microwave condition. Acidic treatment of the peptide loaded resins cleaved the peptide off the resin, which was followed by Fmoc-removal from the *N*-terminus leading to 43.2% isolated yield of bikunin like peptide **1** (**Scheme S1**).

To express the polyhistidine-tagged human XT-I (EC 2.4.2.26),[12] a plasmid encoding signal peptide-His6-XT-I was constructed and used to transfect Expi293F cells. Secreted XT-I protein was purified using Ni Sepharose affinity column with an expression yield of 5 mg/L. Xylosylation was then initiated by sequentially adding UDP-Xyl (1.2 eq), peptide **1** (1 eq), and XT-I (0.025 mol%) to the 2-(*N*-morpholino)ethanesulfonic acid (MES) reaction buffer. After overnight incubation at 37 ºC, quantitative conversion of **1** to xylosylated glycopeptide **2** was confirmed with high-resolution mass spectrometry (HRMS) and high performance liquid chromatography (HPLC). The desired glycopeptide product **2** was isolated *via* G-10 size exclusion chromatography in 89.2% yield at milligram scales (**Scheme 1**). HRMS and nuclear magnetic resonance (NMR) analyses confirmed the structure of the β-glycosylated product ($^1J_{C1,\,H1}$ =159.5 Hz),[18] which was identical to the chemically synthesized glycopeptide **2**.[19]

View Article Online
DOI: 10.1039/D1OB00317H

Scheme 1. XT-I-catalyzed xylosylation of bikunin peptide **1**.



Investigation furthered with multiple peptide substrates **3-6** (**Figure 1**),[20-23] which contain diverse amino acid residues, including hydrophilic or hydrophobic residues flanking the glycosylation site. In addition, peptides **4** and **5** have two potential sites of glycosylation, while peptide **6** has three sites. Excitingly, XT-I enzyme smoothly converted all the peptide substrates to the glycosylated products with the desired stereoselectivity under the same reaction condition as for the formation of glycopeptide **2** (**Table 1**). All glycopeptide structures were confirmed through HPLC, NMR and MS comparisons with glycopeptides synthesized chemically.[19] In addition, a recombinant polyhistidine-tagged human CD44 hyaluronic acid binding domain protein ($hCD44_{20\text{-}178}$)[24] was successfully xylosylated by XT-I demonstrating that XT-I can utilize a protein as an acceptor as well (**Figures S12** and **S13**).

View Article Online
DOI: 10.1039/D1OB00317H

Figure 1. Structures of peptides **3-6** and glycopeptides **7-10** with the serine xylosylation site highlighted.

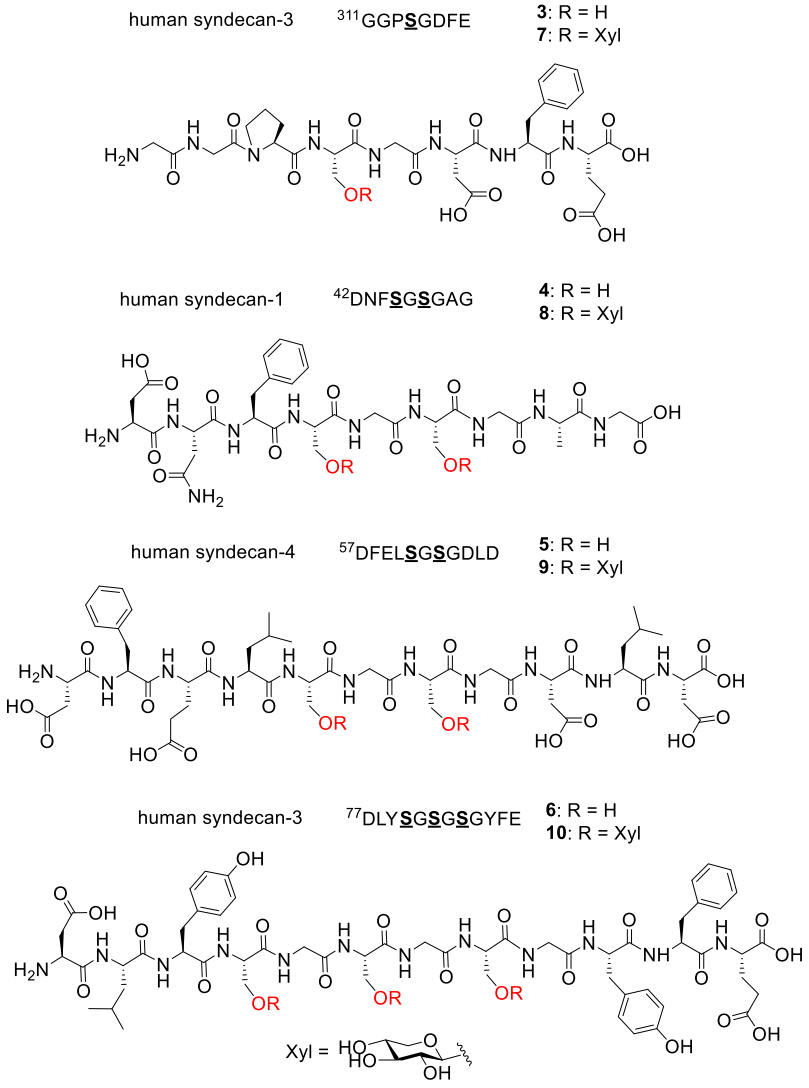


Table 1. Yields of XT-I catalyzed glycosylation of peptides **3-6**.

Acceptor	Product	Reaction Yield (%)
3	**7**	100
4	**8**	68.6
5	**9**	73.8
6	**10**	86.5

View Article Online
DOI: 10.1039/D1OB00317H

To attain more in-depth understandings on XT-I activity and substrate preferences, enzyme kinetics were analyzed with multiple peptide acceptors using a modified phosphatase-coupled glycosyltransferase assay.[25] Among the peptides examined, XT-I demonstrated the highest affinity and catalytic efficiency towards the bikunin peptide **1** (**Table 2**). The differential k_{cat}/K_m values with various peptide sequences suggest that presence of acidic residues *N*-terminal to the xylose attachment site may enhance enzyme activities.

Table 2. Summary of kinetic data from peptide substrates **1**, **3-5**.

Substrate	K_m (μM)	V_{max} (pmol/min/μg)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1} mM^{-1})
1	49.8±4.9	350.9±30.6	28	562
3	308.0±69.4	45.4±4.8	3	10
4	133.8±27.0	196.7±16.3	16	120
5	164.4±23.0	183.8±10.7	15	91

We investigated next the donor selectivity of XT-I. XT-I was believed to be monofunctional to UDP-Xyl.[26] A variety of UDP-sugars was tested as XT-I donors, including UDP-Xyl, UDP-glucose (Glc), UDP-galactose (Gal), UDP-*N*-acetyl glucosamine (GlcNAc) and UDP-6-azidoglucose (6AzGlc) with peptide **1** as the acceptor. UDP-Gal or UDP-GlcNAc was not transferred to any detectable amounts. Examination of the crystal structure of XT-I (PDB code: 6EJ7)[12] shows that axial 4-OH of galactoside would clash with Asp494 and Glu529 (the catalytic base) in the active site of the enzyme (**Figure S14**). For UDP-GlcNAc, the 2-*N*-acetyl group of UDP-GlcNAc could be accommodated, but it could not form the hydrogen bond to Arg598 as present when UDP-Xyl was bound.

Interestingly, besides UDP-Xyl, noticeable enzymatic activities were observed with UDP-Glc and UDP-6AzGlc (**Table 3**). The successful transfer of 6AzGlc to bikunin peptide **11** (**Scheme S2**) by XT-I indicates its potential to be developed as a valuable biolabeling tool. As a proof of concept, azido-tagged glycopeptide **12** and Cy5-alkyne were subject to copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC). The desired Cy5 conjugated glycopeptide **13** was successfully produced (**Scheme S2**).

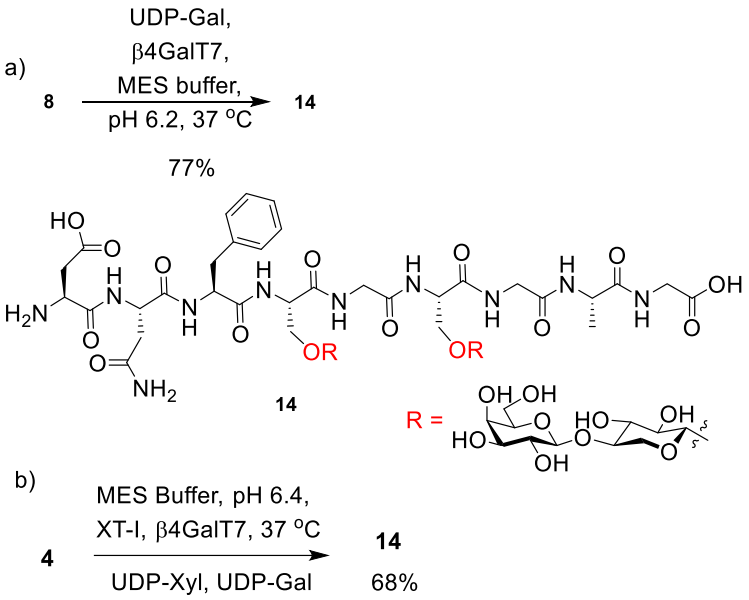
Table 3. Summary of kinetic results from UDP-sugar donors.

Substrate	K_m (μM)	V_{max} (pmol/min/μg)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}mM^{-1}$)
UDP-Xyl	43.4±6.9	165.9±6.4	13	266
UDP-Glc	84.0±26.6	20.4±1.9	2	20
UDP-6AzGlc	23.4±10.5	11.0± 1.0	1	39

View Article Online
DOI: 10.1039/D1OB00317H

As proteoglycans can contain long glycan chains, it is important that the glycan of the synthetic xylosyl peptides can be extended. In nature, a galactosyl transferase such as the β4GalT7 is responsible for adding a galactose unit to the xylose from the UDP galactose (UDP-Gal) donor.[13-15] Recently, β4GalT7 has been shown to be able to galactosylate chemically synthesized xylosylated peptides.[19] To test whether the enzymatically prepared xylosyl peptide is a viable substrate for β4GalT7, xylosylated peptide **8** was treated with β4GalT7 and UDP-Gal (**Scheme 2a**). The glycopeptide **14**[19] with two Gal-Xyl disaccharides was successfully produced in 77% yield. Thus, the overall yield for the stepwise conversion of **4** to **14** with XT-I glycosylation followed by β4GalT7 reaction was 53%. To further improve the synthetic efficiency, one pot synthesis was explored with XT-I and β4GalT7. Peptide **4**, UDP-Xyl, UDP-Gal, XT-I, and β4GalT7 were incubated together in the MES reaction buffer at 37 ºC overnight (**Scheme 2b**). Encouragingly, a full conversion of acceptor peptide **4** was observed with an isolated yield of 68% for glycopeptide **14**. Besides peptide **4**, this one-pot two-enzyme (OP2E) protocol smoothly converted peptides **3, 5**, and **6** to the corresponding glycopeptides **15-17**[19] (**Figure 2**) with higher yields compared to the stepwise synthesis (**Table 4**). The polyhistagged $hCD44_{20-178}$ protein was also glycosylated by the OP2E method to yield the Gal-Xyl modified CD44 (**Figure S15**).

Scheme 2. A) Galactosylation of glycopeptide **8** by β4GalT7 to form glycopeptide **14** bearing galactose-xylose disaccharide; b) OP2E synthesis of **14** from peptide **4** by one pot reaction with XT-I and β4GalT7.



View Article Online
DOI: 10.1039/D1OB00317H

Figure 2. Structures of glycopeptide products **15-17** from OP2E reactions. Glycosylated serine sites are highlighted.

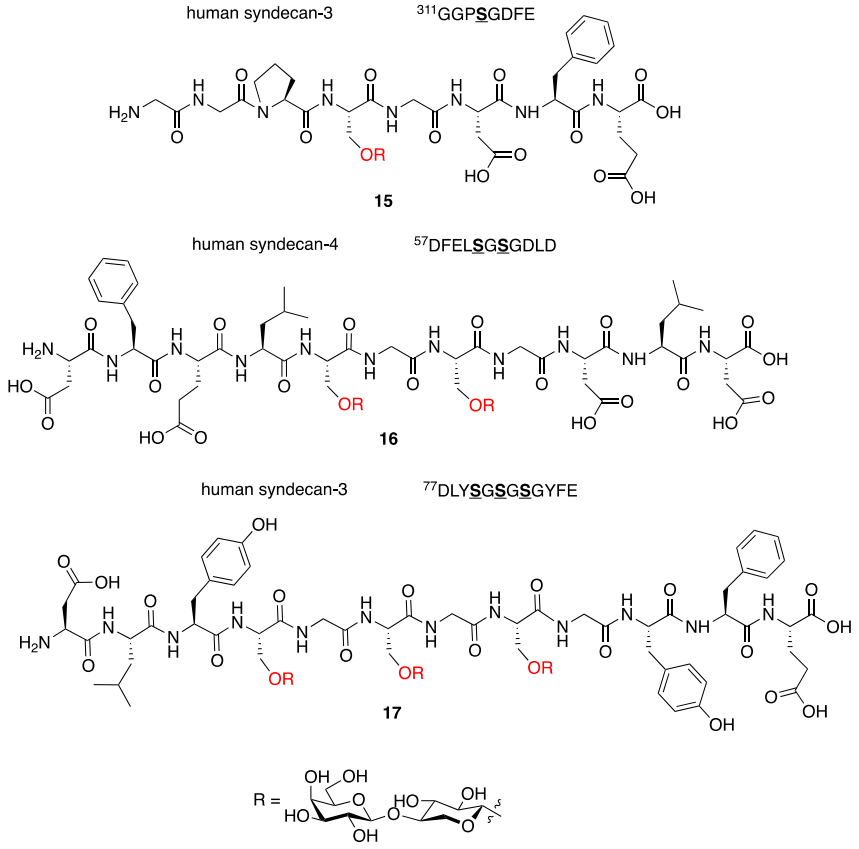


Table 4. Yield summary of OP2E synthesis.

Acceptor	Product	Overall Yield (%) from Stepwise Synthesis	OP2E Reaction Yield (%)
3	**15**	91	94
4	**14**	53	68
5	**16**	60	89
6	**17**	68	91

Enzymatic synthesis of glycopeptide such as **14** is more efficient than the corresponding chemical synthesis. Due to the need for multiple protective group manipulation to prepare the two strategically protected monosaccharide building blocks followed by the technically challenging chemical glycosylations and deprotection reactions, it would have taken over 20 synthetic steps to access a glycopeptide such as **14** *via* chemical glycosylation from commercially available monosaccharides.[5] Thus, the OP2E protocol can significantly improve the overall synthetic efficiencies.

In the OP2E protocol for glycopeptide synthesis, XT-I presumably installed the xylose onto the peptide first, followed by β4GalT7 promoted galactosylation of the xylosylated peptide as in the case for stepwise synthesis. Alternatively, β4GalT7 may galactosylate UDP-Xyl first with subsequent transfer of the UDP disaccharide donor to the peptide acceptor catalyzed by XT-I.

View Article Online
DOI: 10.1039/D1OB00317H

However, the formation of disaccharide donor in OP2E reaction is unlikely to occur at an appreciable rate as β4GalT7 prefers β-xyloside acceptors.[27] The UDP-Xyl has an α-anomeric configuration and the UDP moiety would clash with the active site of β4GalT7 enzyme. Furthermore, the crystal structure of XT-I (PDB code: 6EJ7)[12] shows that the active site of XT-I (**Figure S14**) is not sufficiently large to accommodate a disaccharide donor.

In conclusion, human XT-I (EC 2.4.2.26) enzyme has been expressed and utilized to efficiently synthesize structurally diverse xylosylated glycopeptides at milligram scales with a range of peptide acceptors as well as the His tag bearing $hCD44_{20\text{-}178}$ protein. XT-I was found tolerant toward several non-native UDP-sugar donors, particularly UDP-6AzGlc, rendering it potentially a valuable tool to label biological proteins. The one-pot two-enzyme method developed further enhanced the synthetic efficiency and the overall yield, paving the way toward efficient chemoenzymatic synthesis of PG glycopeptides and glycoproteins.

Author Contributions

XH and JG designed the project. JG, PL, SN expressed the protein, synthesized and characterized the peptides and glycopeptides. KL expressed the CD44 protein. KY and EH guided the plasmid design and enzyme expression. All authors contributed to the preparation of the manuscript.

Acknowledgments

We are grateful to the National Institute of General Medical Sciences, National Institutes of Health (Grant R01GM072667 to XH), Michigan State University, the Wellcome Trust (101748/Z/13/Z to EH) and the UK Biotechnology and Biological Sciences Research Council (BB/T01279X/1 to EH) for financial support of our work.

Supporting Information: The Supporting Information includes detailed experimental procedures, characterization data and spectra, supplementary figures, and supplementary schemes.

References

1. M. Bernfield, M. Götte, P. W. Park, O. Reizes, M. L. Fitzgerald, J. Lincecum and M. Zako, *Annu. Rev. Biochem.*, 1999, **68**, 729-777.
2. J. R. Couchman, *Nat. Rev. Mol. Cell Biol.*, 2003, **4**, 926-937.
3. L. Schaefer and R. M. Schaefer, *Cell Tissue Res.*, 2010, **339**, 237-246.
4. J. D. Esko and L. Zhang, *Curr. Opin. Struct. Biol.*, 1996, **6**, 663-670.
5. W. Yang, K. Yoshida, B. Yang and X. Huang, *Carbohydr. Res.*, 2016, **435**, 180-194 and references cited therein.
6. W. Yang, Y. Eken, J. Zhang, L. Cole, S. Ramadan, Y. Xu, Z. Zhang, J. Liu, A. K. Wilson and X. Huang, *Chem. Sci.*, 2020, **11**, 6393-6404 and references cited therein.
7. T. Li, W. Yang, S. Ramadan and X. Huang, *Org. Lett.*, 2020, **22**, 6429-6433.
8. T. Brinkmann, C. Weilke and K. Kleesiek, *J. Biol. Chem.*, 1997, **272**, 11171-11175.
9. I. B. H. Wilson, *J. Biol. Chem.*, 2002, **277**, 21207-21212.
10. A. E. Kearns, S. C. Campbell, J. Westley and N. B. Schwartz, *Biochemistry*, 1991, **30**, 7477-7483.
11. C. Weilke, T. Brinkmann and K. Kleesiek, *Clin. Chem.*, 1997, **43**, 45-51.
12. D. C. Briggs and E. Hohenester, *Structure*, 2018, **26**, 801-809.

View Article Online
DOI: 10.1039/D1OB00317H

13. R. Almeida, S. B. Levery, U. Mandel, H. Kresse, T. Schwientek, E. P. Bennett and H. Clausen, *J. Biol. Chem.*, 1999, **274**, 26165-26171.
14. F. Daligault, S. Rahuel-Clermont, S. Gulberti, M. T. Cung, G. Branlant, P. Netter, J. Magdalou and V. Lattard, *Biochem. J.*, 2009, **418**, 605-614.
15. I. Talhaoui, C. Bui, R. Oriol, G. Mulliert, S. Gulberti, P. Netter, M. W. Coughtrie, M. Ouzzine and S. Fournel-Gigleux, *J. Biol. Chem.*, 2010, **285**, 37342-37358.
16. U. Pfeil and K. Wenzel, *Glycobiology*, 2000, **10**, 803-807.
17. I. B. H. Wilson, *Cell. Mol. Life Sci.*, 2004, **61**, 794-809.
18. K. Bock and C. Pedersen, *J. Chem. Soc., Perkin Trans. 2*, 1974, 293-297.
19. J. Gao, P.-H. Lin, S. T. Nick, J. Huang, E. Tykesson, U. Ellervik, J. Li and X. Huang, *Org. Lett.*, 2021, **23**, 1738–1741.
20. E. E. Grebner, C. W. Hall and E. F. Neufeld, *Arch. Biochem. Biophys.*, 1966, **116**, 391-398.
21. H. C. Robinson, A. Telser and A. Dorfman, *Proc. Natl. Acad. Sci. U. S. A.*, 1966, **56**, 1859-1866.
22. A. C. Stoolmiller, A. L. Horwitz and A. Dorfman, *J. Biol. Chem.*, 1972, **247**, 3525-3532.
23. P. Campbell, I. Jacobsson, L. Benzing-Purdie, L. Roden and J. H. Fessler, *Anal. Biochem.*, 1984, **137**, 505-516.
24. L.-K. Liu and B. C. Finzel, *J. Med. Chem.*, 2014, **57**, 2714-2725.
25. Z. L. Wu, C. M. Ethen, B. Prather, M. Machacek and W. Jiang, *Glycobiology*, 2011, **21**, 727-733.
26. J. Kuhn, C. Gotting, B. J. Beahm, C. R. Bertozzi, I. Faust, P. Kuzaj, C. Knabbe and D. Hendig, *Biochem. Biophys. Res. Commun.*, 2015, **459**, 469-474.
27. Y. Tsutsui, B. Ramakrishnan and P. K. Qasba, *J. Biol. Chem.*, 2013, **288**, 31963-31970.