Alternative donor substrates for inverting and retaining glycosyltransferases†

Luke L. Lairson,[a] Warren W. Wakarchuk[b] and Stephen G. Withers*[ac]

Received (in Cambridge, MA, USA) 9th October 2006, Accepted 1st November 2006
First published as an Advance Article on the web 7th November 2006
DOI: 10.1039/b614636h

By the use of glycosyl donors containing aromatic leaving groups linked with opposite anomeric configurations compared to those of the natural donor substrates, an inverting (Cst II) and a retaining (LgtC) glycosyltransferase were found to catalyse glycosylation reactions of natural acceptor substrates in the presence of the corresponding nucleotide.

A detailed understanding of how oligosaccharide structures present in normal and pathological biological settings dictate functions ranging from cell–cell communication to the metastasis of cancerous cells remains one of the least explored avenues of chemical biology.[1,2] This lack of understanding limits the potential of carbohydrate-based therapeutic agents and/or therapeutic strategies, although some success has recently been reported in this area.[3,4] This deficiency in understanding, compared to that for protein and nucleic acids research, is not the result of a lack of importance, but rather a lack of physical access to the incredible range of possible oligosaccharide structures. The wide range of regio- and stereo-chemical linkages by which individual carbohydrate units can be connected results in a level of diversity in possible glycosylation patterns that far exceeds the structural diversity found with DNA and even proteins.[1] For this very reason, the application of traditional synthetic chemistry to the synthesis of desired glycan structures requires multiple and selective protection and deprotection schemes, thereby limiting feasibility and economic viability. As such, enzymatic synthesis presents itself as a highly attractive alternative. Traditional use of glycosidases in transglycosylation reactions typically suffers the disadvantage of low yields, although rationally engineered glycosidases (glycosynthases) have provided renewed potential utility to this class of enzyme.[5,6]

Another logical approach is the use of the anabolic enzymes responsible for glycan synthesis in nature – glycosyltransferases. However, the application of this class of enzyme has suffered from both a lack of access to the enzymes themselves, as many exist as membrane-associated species,[7] and also from the cost of glycosyl donor substrates (typically nucleotide sugars). With bacterial enzymes, the former problem is currently being overcome by successful cloning strategies involving truncations to yield recombinant soluble catalytic domains.[8,9] Attempts to overcome the latter problem involve elaborate coupled enzyme recycling schemes involving *in situ* generation of the donor substrate.[10,11] A recently described alternative strategy takes advantage of the reversibility of natural product glycosyltransferases for *in situ* generation of desired nucleotide sugars.[12] The increasing accessibility of the pyrophosphorylases and nucleotidyltransferases responsible for the formation of nucleotide sugars should further facilitate these approaches, however, the technical complexity and problems of product inhibition provide an impetus for exploring potential alternative substrates for glycosyltransferases.[13,14]

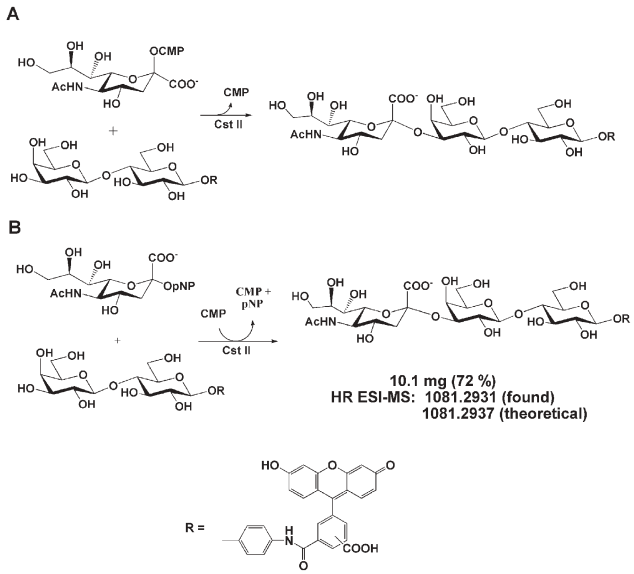
As with the glycosidases, glycosyltransferases are classified as either retaining or inverting depending on the stereochemical outcome at the anomeric centre relative to that of the donor sugar. They are are also classified into sequence similarity-based families which are kept updated on the Carbohydrate Active enZYmes database (CAZY).[15,16] Glycosyltransferases are thought to use mechanistic strategies that directly parallel those used by the well characterized glycosidases and transglycosidases.[17] However, while the mechanism of inverting glycosyltransferases is generally accepted, the mechanism of retaining enzymes of this class remains a topic of considerable debate within the field.[18–20]

The following report describes investigations into the donor substrate specificity of representatives of both retaining and inverting classes of glycosyltransferases. The results provide more fodder for the mechanistic debate surrounding this class of enzyme and present a potential starting point for engineering efforts to provide economical alternatives for their application in enzymatic glycan synthesis.

Cst II is an inverting bifunctional α-2,3/2,8 sialyltransferase from *Campylobacter jejuni*.[21] It uses CMP β-D-sialic acid (CMP = cytidine 5′-monophosphate) as a donor substrate and transfers the sialic acid moiety with net inversion of anomeric configuration to the 3′ hydroxyl of terminal lactose-containing acceptors (Scheme 1A). Subsequently, Cst II will use this 3′-sialyl lactose product as an acceptor and transfer sialic acid to the 8″ hydroxyl. The three-dimensional structure of this enzyme has recently been solved and supports a catalytic mechanism that appears to involve a straightforward direct displacement reaction whereby active site residues facilitate departure of the CMP leaving group and activate the incoming nucleophile of the acceptor.[21]

To determine whether this enzyme could use an alternative source of activated sialic acid as a donor substrate, we chose to investigate the α-linked *para*-nitrophenyl (pNP) derivative (Scheme 1B). Providing potential precedence for their use as alternative substrates for glycosyltransferases, nitrophenyl derivatives of ribose have been successfully employed as alternative

[a]*Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1. E-mail: withers@chem.ubc.ca; Fax: (604) 822-8869; Tel: (604) 822-3402*
[b]*Institute of Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6. E-mail: warren.wakarchuk@nrc.ca*
[c]*Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3*
† Electronic supplementary information (ESI) available: General methods and Figs. S1–S3. See DOI: 10.1039/b614636h

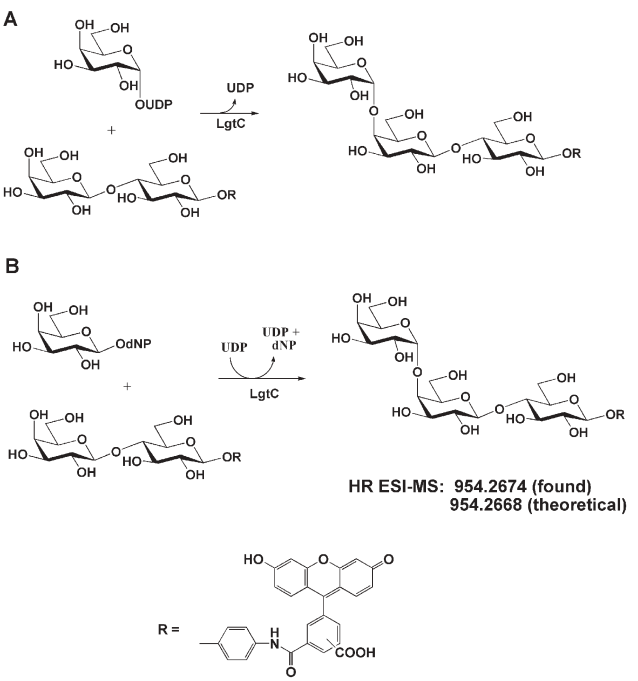


Scheme 1 Reaction catalysed by Cst II using (A) the natural donor substrate CMP β-D-sialic acid or (B) the alternative donor substrate *p*NP α-D-sialic acid.

substrates and used as mechanistic probes for both purine nucleoside phosphorylase and *N*-ribohydrolase.[22] Because Cst II is an inverting enzyme, it was presumed that the aromatic ring of the nitrophenyl substituent could be accommodated within the active site, taking the place of the galactose ring of the acceptor which must be located on the α face of the sialic acid donor. With pNP α-D-sialoside bound in such a manner, CMP could also be accommodated within the active site on the β face of sialic acid, allowing for the direct displacement of *p*-nitrophenolate by CMP and the *in situ* formation of β-linked CMP sialic acid, which could be used as the donor substrate in a subsequent transfer reaction following release of pNP. Catalyzed sialic acid transfer was monitored by TLC using a fluorescein–lactose conjugate acceptor and, as is shown in Fig. S1, Cst II was indeed found to catalyze transfer to the fluorescent acceptor but only in the presence of CMP. Subsequent preparative scale reactions were performed to determine the efficiency of the reaction (Scheme 1B). Treatment of the resulting product with a specific α-2,3-neuraminidase was used to confirm the regio- and stereo-chemistry as being that of the anticipated linkage (Fig. S2).

To further explore the generality of this substrate promiscuity amongst glycosyltransferases, a similar strategy was applied to the retaining α-1,4 galactosyl transferase LgtC from *Neisseria meningitidis*.[18] LgtC catalyzes the transfer of galactose to the 4′ hydroxyl of terminal lactose-containing acceptor substrates using UDP α-D-galactose (UDP = uridine 5′-diphosphate) as the donor with overall net retention of anomeric configuration (Scheme 2A). It has previously been observed that LgtC will use α-galactosyl fluoride as a donor substrate analogue in the presence of UDP, resulting in the observed formation of α-linked UDP Gal.[23] This finding is consistent with either of the proposed mechanisms of retaining glycosyltransferases.[17] The observed activity could occur *via* a double displacement mechanism involving initial displacement of fluoride ion by a catalytic nucleophile, resulting in the transient formation of a covalent glycosyl-enzyme intermediate, followed by subsequent displacement by an incoming nucleophile (either the acceptor or UDP). Alternatively, an oxocarbenium ion is formed, with a lifetime that allows for a degree of departure by the fluoride ion leaving group sufficient for attack by the incoming nucleophile on the same face. To further test the limits of donor substrate promiscuity, LgtC was tested for its ability to use pNP α-D-galactoside or dNP α-D-galactoside (dNP = 2,4-dinitrophenyl) as the surrogate donor substrate. Again, glycosyl transfer was monitored by TLC using a fluorescein–lactose conjugate acceptor. Under these conditions, no trisaccharide product was observed (data not shown). This is not surprising as the dNP leaving group would have to occupy the UDP binding site of the enzyme and the binding of UDP to LgtC is known to induce a conformational change that results in the closure of a loop that makes up a significant portion of the active site.[18] Therefore, an α-linked alternative leaving group must be of sufficiently small size to allow for simultaneous binding of UDP, as is the case with α-galactosyl fluoride.

A strategy analogous to that described above for Cst II was then applied to LgtC in which an alternative donor with an activated leaving group of anomeric configuration *opposite* to that of the natural donor was tested. In this case, because LgtC is a retaining enzyme, the aromatic leaving group would not occupy the acceptor site as both the leaving group and the incoming nucleophile are present on the same face of the donor sugar in the active site of this class of enzyme. As such, the success of this approach would rely upon fortuitous accommodation of an aromatic substituent on the β face of the donor within the LgtC active site. Initially, pNP β-D-galactoside was tested for its ability to act as a surrogate donor in the presence or absence of UDP, but no trisaccharide product was observed in either case, even at high enzyme concentrations (0.5 mg mL^{-1}) and after extended



Scheme 2 Reaction catalysed by LgtC using (A) the natural donor substrate UDP α-D-galactose or (B) the alternative donor substrate 2,4-dinitrophenyl β-D-galactoside.

incubation times (5 days) (data not shown). Dependence on leaving group ability was then probed by repeating the above experiments using the more activated derivative 2,4-dinitrophenyl β-D-galactoside (Scheme 2B). Using this more activated analogue a trisaccharide was observed and again glycosyl transfer only occurred in the presence of the natural nucleotide (UDP) (Fig. S3). Analogously to the inverting enzyme, LgtC presumably catalyzes the direct displacement of the activated β-linked donor analogue by UDP, resulting in the formation of the natural α-linked UDP galactose donor substrate which is then used in a subsequent glycosylation reaction involving the acceptor substrate to yield a trisaccharide product.

From the present understanding of structure and mechanism of the inverting Cst II, it is reasonably clear how the enzyme active site accommodates the observed alternate activity. The α-linked pNP substituent could bind within the acceptor substrate binding site ready for displacement by CMP bound to its natural binding site on the β face of sialic acid. It is expected that this approach should be generally applicable amongst other inverting glycosyltransferases. However, it is more difficult to rationalize the observed activity with the retaining enzyme LgtC. Based on the available ternary complex crystal structure it is not clear how the β-linked dNP substituent can be accommodated within the active site as both the acceptor and leaving group binding sites lie on the α face of the galactose donor. A catalytically relevant covalent glycosyl-enzyme intermediate has recently been observed for an active site mutant of LgtC, however the labelled aspartate residue is separated from the anomeric reaction centre by a distance of 9 Å in the available crystal structure.[20] Even if this residue does not play the role of the catalytic nucleophile in a double displacement mechanism, in order for this covalent species to be formed and turned over a significant conformational change from that observed in the crystal structure must occur during catalysis, thereby indicating a significant degree of structural plasticity. This structural plasticity is further supported by the observed accommodation of β-linked dNP galactose in a catalytically competent conformation. Although neither of these results provides definitive support for a given catalytic mechanism used by retaining glycosyltransferases, they do indicate that crystal structures for glycosyltransferase complexes should be interpreted cautiously.

The results of this work illustrate how the donor substrate promiscuity of both inverting and retaining glycosyltransferases can be harnessed, providing the starting point of an alternative strategy for their application as synthetic tools. Both classes of enzymes appear to be able to use glycosyl donors with alternative activated leaving groups of *opposite* anomeric configuration, compared to the natural donor substrate, in the presence of catalytic amounts of the natural nucleotide. The enzyme functions by catalyzing the *in situ* formation of the natural nucleotide sugar donor, which is then used in a subsequent catalyzed glycosylation reaction involving an acceptor substrate.

The authors would like to thank the Michael Smith Foundation for Health Research for a Senior Graduate Studentship Award (L.L.L.), the Natural Sciences and Engineering Research Council of Canada for a Doctoral Post Graduate Scholarship (L.L.L.) and the Canadian Institutes for Health Research for funding (S.G.W. and W.W.W.).

Notes and references

1 C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357–2364.
2 P. H. Seeberger, *Nature*, 2005, **437**, 1239.
3 J. Alper, *Science*, 2001, **291**, 2338–2343.
4 D. B. Werz and P. H. Seeberger, *Chemistry*, 2005, **11**, 3194–3206.
5 L. F. Mackenzie, Q. P. Wang, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 1998, **120**, 5583–5584.
6 G. Perugino, A. Trincone, M. Rossi and M. Moracci, *Trends Biotechnol.*, 2004, **22**, 31–37.
7 J. C. Paulson and K. J. Colley, *J. Biol. Chem.*, 1989, **264**, 17615–17618.
8 T. R. Henion, B. A. Macher, F. Anaraki and U. Galili, *Glycobiology*, 1994, **4**, 193–202.
9 W. W. Wakarchuk, A. Cunningham, D. C. Watson and N. M. Young, *Protein Eng.*, 1998, **11**, 295–302.
10 I. Yoshitaka, G. J. Shen and C. H. Wong, *J. Am. Chem. Soc.*, 1991, **113**, 4698–4700.
11 M. Gilbert, R. Bayer, A. M. Cunningham, S. DeFrees, Y. Gao, D. C. Watson, N. M. Young and W. W. Wakarchuk, *Nat. Biotechnol.*, 1998, **16**, 769–772.
12 C. Zhang, B. R. Griffith, Q. Fu, C. Alberman, X. Fu, I. K. Lee, L. Li and J. S. Thorson, *Science*, 2006, **313**, 1291–1294.
13 P. A. Ropp and P. W. Cheng, *Anal. Biochem.*, 1990, **187**, 104–108.
14 X. Ma and J. Stockigt, *Carbohydr. Res.*, 2001, **333**, 159–163.
15 Information on Carbohydrate Active enZYmes maintained on the CAZY server at http://afmb.cnrs-mrs.fr/CAZY/.
16 P. M. Coutinho, E. Deleury, G. J. Davies and B. Henrissat, *J. Mol. Biol.*, 2003, **328**, 307–317.
17 L. L. Lairson and S. G. Withers, *Chem. Commun.*, 2004, 2243–2248.
18 K. Persson, H. D. Ly, M. Dieckelmann, W. W. Wakarchuk, S. G. Withers and N. C. J. Strynadka, *Nat. Struct. Biol.*, 2001, **8**, 166–175.
19 L. N. Gastinel, C. Bignon, A. K. Misra, O. Hindsgaul, J. H. Shaper and D. H. Joziasse, *EMBO J.*, 2001, **20**, 638–649.
20 L. L. Lairson, C. P. Chiu, H. D. Ly, S. He, W. W. Wakarchuk, N. C. Strynadka and S. G. Withers, *J. Biol. Chem.*, 2004, **279**, 28339–28344.
21 C. P. Chiu, A. G. Watts, L. L. Lairson, M. Gilbert, D. Lim, W. W. Wakarchuk, S. G. Withers and N. C. Strynadka, *Nat. Struct. Mol. Biol.*, 2004, **11**, 163–170.
22 L. J. Mazzella, D. W. Parkin, P. C. Tyler, R. H. Furneaux and V. L. Schramm, *J. Am. Chem. Soc.*, 1996, **118**, 2111–2112.
23 B. Lougheed, H. D. Ly, W. W. Wakarchuk and S. G. Withers, *J. Biol. Chem.*, 1999, **274**, 37717–37722.