



**HAL**  
open science

**A single point mutation alters the  
transglycosylation/hydrolysis partition, significantly  
enhancing the synthetic capability of an  
endo-glycoceramidase**

Julien Durand, Xevi Biarnés, Laurie Watterlot, Cyrielle Bonzom, Vinciane Borsenberger, Antoni Planas, Sophie Bozonnet, Michael O'Donohue, Régis Fauré

► **To cite this version:**

Julien Durand, Xevi Biarnés, Laurie Watterlot, Cyrielle Bonzom, Vinciane Borsenberger, et al.. A single point mutation alters the transglycosylation/hydrolysis partition, significantly enhancing the synthetic capability of an endo-glycoceramidase. *ACS Catalysis*, 2016, 6 (12), pp.8264-8275. 10.1021/acscatal.6b02159 . hal-01606355

**HAL Id: hal-01606355**

**<https://hal.science/hal-01606355v1>**

Submitted on 4 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **A Single Point Mutation Alters the Transglycosylation/Hydrolysis Partition, Significantly Enhancing the Synthetic Capability of an *Endo*-Glycoceramidase**

Julien Durand,<sup>†</sup> Xevi Biarnés,<sup>‡</sup> Laurie Watterlot,<sup>†</sup> Cyrielle Bonzom,<sup>†</sup> Vinciane Borsenberger,<sup>†</sup> Antoni Planas,<sup>‡</sup> Sophie Bozonnet,<sup>†</sup> Michael J. O'Donohue,<sup>†</sup> Régis Faure<sup>\*†</sup>

<sup>†</sup>*LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France*

<sup>‡</sup>*Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Barcelona, Spain*

\* Corresponding Author

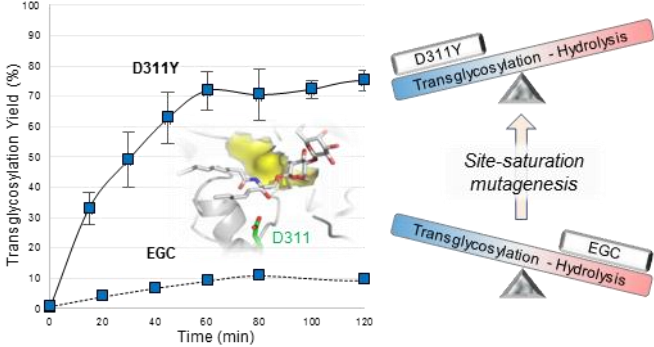
[regis.faure@insa-toulouse.fr](mailto:regis.faure@insa-toulouse.fr)

### **ABSTRACT:**

The mutation of D311 to tyrosine in *endo*-glycoceramidase II from *Rhodococcus* sp. and the use of a poorly recognized substrate, 2-chloro-4-nitrophenyl  $\beta$ -cellobioside, have provided appropriate conditions for the efficient synthesis of alkyl  $\beta$ -cellobioside derivatives. The mutant D311Y was characterized by a lowered  $K_M$  value for the hydrolysis of 2-chloro-4-nitrophenyl  $\beta$ -cellobioside and increased transglycosylation when using aliphatic 1,3-diols or alcohols bearing a  $\delta$ -hydroxy ketone function as acceptors. Closer analysis revealed that the transglycosylation/hydrolysis ratio in reactions catalyzed by the mutant was completely inverted and weak secondary hydrolysis was postponed, thus providing the basis for high transglycosylation yields (between 68 and 93%). Overall, results confirm that the enhancement of transglycosylation in glycoside hydrolases can be achieved by a combination of destabilized transition states and increased recognition for acceptor molecules.

**KEYWORDS:** glycoside hydrolase, mutagenesis, transglycosylation, alkyl cellobiosides, molecular interactions, chemo-enzymatic glycosynthesis

# Table of Contents artwork

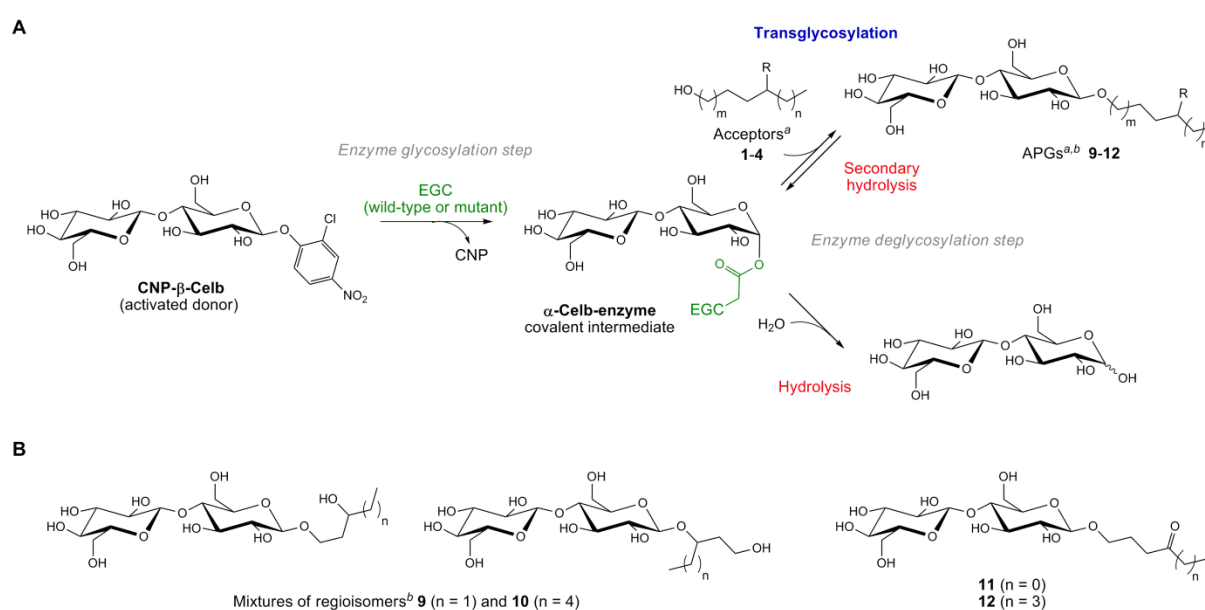


## INTRODUCTION

Alkyl polyglycosides (APGs) are of commercial interest, since they display non-ionic surfactant properties that are exploited for the preparation of a wide variety of products, including detergents and cleaners, personal care products, or food ingredients.<sup>1-5</sup> Remarkably, they also have been shown to reduce skin irritation compared to conventional detergents.<sup>6,7</sup> The market for this family of molecules is expected to grow, because of the drive to reduce the use of petrochemical-based surfactants, favoring the manufacture of bio-based products using eco-friendly technologies. However, most APGs are in reality alkyl glycosides displaying an average degree of polymerization (DP) of the saccharidic moiety close to 1. This is because APGs are currently synthesized through acid-catalyzed Fischer or Koenigs-Knorr glycosidation, reductive amination or using Lewis acid catalysts.<sup>8</sup> However, the products of these reactions are mainly monosaccharide derivatives, because the anomeric centers of oligosaccharide moieties display lower reactivity and their glycosidic bonds are acid-sensitive. An alternative route to obtain a new generation of APGs displaying saccharidic moieties with higher DP is enzymatic synthesis. In this case, it is possible to benefit from the high selectivity that is a hallmark of enzymes. Several examples of enzymatic APG synthesis using glycoside hydrolases (GHs) illustrate the feasibility and efficiency of this route and demonstrate that it is possible to obtain products in good yields.<sup>8,9</sup> Nevertheless, naturally-occurring GHs are mostly enzymes that hydrolyze glycosidic bonds, with bond synthesis occurring either in specific conditions (high substrate concentration and low water activity) that favor so-called reverse hydrolysis, or through a mechanistically-related transglycosylation reaction that is under kinetic control.<sup>10,11</sup> Only a relatively small number of GHs, known as non-Leloir transglycosylases (TGs), naturally display strong innate ability to catalyze transglycosylation reactions. Interestingly, the key molecular determinants that differentiate TGs from hydrolytic GHs are still unclearly defined, despite considerable efforts to discover these and relevant studies on reversed transglycosylation/hydrolysis (T/H) partition.<sup>12-17</sup>

The class of GHs (retaining GHs, hereafter referred to as rGHs) mostly holding the potential to perform transglycosylation reactions are those that catalyze the hydrolysis of glycosidic bonds through a double displacement mechanism that results in the retention (in the product) of the configuration of the anomeric carbon originally present in the donor substrate. The reaction trajectory of rGHs is characterized by the formation of a more or less transient glycosylated enzyme reaction intermediate that can be deglycosylated either by a water molecule (hydrolysis), or by another suitable acceptor moiety (transglycosylation), including

sugars and alcohols (Figure 1).<sup>18,19</sup> This implies that the outcome of rGH-catalyzed reactions is influenced by competition at the deglycosylation step between water molecules and any alternative acceptors that might be present. In most cases, because water is omnipresent at a concentration of 55 M in aqueous reaction media, hydrolysis is the principal outcome of a rGH-catalyzed reaction, although transglycosylation often occurs to some extent, with glycosylated products being obtained at low yields, except in the special case of TGs.<sup>15</sup> Moreover, in hydrolytic rGHs, yields of transglycosylation products are further diminished by secondary hydrolysis, a reaction in which the synthetic product becomes a substrate for the enzyme.



**Figure 1.** (A) Reactions catalyzed by the retaining GH5 *endo*-glycoceramidase II (EGC) from *Rhodococcus* sp. M-777 strain and its evolved mutant. (B) Structures of the transglycosylation products **9-12**. <sup>a</sup> 1,3-Diol **1** and **2** (R = OH, m = 0, n = 1 and 4 respectively) and  $\delta$ -hydroxy ketones **3** and **4** (R = O, m = 1, n = 0 and 3 respectively). <sup>b</sup> The secondary hydroxyl group of 1,3-diol **1** and **2** was also glycosylated leading to an almost stoichiometric mixture of regioisomers of APGs **9** and **10**.

Numerous reports recount the enzymatic synthesis of APGs using rGHs displaying different sugar specificities and various alcohols.<sup>19,20</sup> Overall, previous studies have revealed that rGH-catalyzed alkylation is mainly restricted to short-chain primary alcohols, probably because the longer chains are not well accommodated in the active site of GHs and are increasingly

insoluble in aqueous reaction media. To counter the latter problem and increase the yield of longer aliphatic chain APGs, some authors have deployed specific strategies, such as the use of co-solvents and vigorous shaking, or the use of immobilized enzyme systems.<sup>21</sup> However, an alternative strategy is to perform protein engineering, altering the enzyme's aglycone binding subsites in order to increase affinity for longer chain alcohols. In the current study, this option has been applied to type II *endo*-glycoceramidase (EGC) from *Rhodococcus* sp. M-777 strain, an enzyme that hydrolyzes the  $\beta$ -glycosidic bonds of gangliosides, releasing an oligosaccharide moiety and a ceramide.<sup>22-24</sup> According to the CAZy classification,<sup>25</sup> this rGH belongs to family 5 (subfamily 28)<sup>26</sup> that is in turn assigned to clan A, the members of which bear a catalytic domain that displays  $(\beta/\alpha)_8$  structural architecture. Regarding the active site topology of EGC, previous studies have shown that this enzyme presents a large polar cavity that accommodates the polyhydroxylated oligosaccharide moiety of the substrate and a narrow hydrophobic tunnel that binds the lipidic chains of the ceramide.<sup>27</sup>

Previous works on *endo*-glycoceramidases have revealed their natural ability to catalyze transglycosylation reactions,<sup>28-33</sup> although the reported reaction yields are not high, no doubt because the products of transglycosylation are invariably subjected to so-called secondary hydrolysis by EGC. To attenuate hydrolysis, a catalytically-impotent version of EGC has been created, using the principles of the glycosynthase methodology,<sup>34-36</sup> which involve substitution of the catalytic nucleophile residue (E351 in EGC) by one that does not display requisite nucleophile character (e.g., Ala, Cys, Gly, or Ser). This work resulted in the creation of a hydrolytically-inactive enzyme (containing the mutation E351S) that can only catalyze transglycosylation reactions in the presence of fluorinated donor substrates displaying  $\alpha$ -configured anomeric centers.<sup>37</sup> To improve the overall performance of this EGC-based glycosynthase, it was then submitted to random mutagenesis and subsequent screening using an ELISA-based assay. This approach ultimately procured an evolved mutant E351S-D314Y, which displayed the ability to synthesize complex glycosphingolipids in 45-86% yields starting from the acceptor D-erythro-sphingosine and tri- or penta-saccharyl donors.<sup>38-40</sup>

In the current study, to develop new enzyme tools for APG synthesis and possibly to prepare glycosphingolipids,<sup>41,42</sup> we have engineered EGC, targeting its ability to perform transglycosylation using a cellobiosyl moiety as the donor and simple aliphatic alcohols or derivatives (aliphatic 1,3-diols and  $\delta$ -hydroxy ketones) as acceptors (Figure 1). Unlike the aforementioned study that used glycosynthase methodology, we employed a site-directed saturation mutagenesis approach, targeting amino acids that form part of the long narrow hydrophobic tunnel that constitutes the aglycone binding subsites of EGC. This strategy

procured a mutant (D311Y) that displays significantly enhanced ability to catalyze transglycosylation and the means to prepare a series of alkyl cellobioside derivatives.

## EXPERIMENTAL SECTION

**Substrates, Materials and Chemicals.** 2-Chloro-4-nitrophenyl  $\beta$ -cellobioside (CNP- $\beta$ -Celb) was purchased from Megazyme. Pentane-1,3-diol (**1**), octane-1,3-diol (**2**), 5-hydroxypentan-2-one (**3**) and 1-hydroxyoctan-4-one (**4**) were purchased from FCH Group. Molecular biology products were purchased from New England Biolabs. The synthetic gene encoding EGC (GenScript), optimized for expression in *E. coli* (Figure S1),<sup>43</sup> was cloned into the vector pET28 (Novagen) between the *NdeI* and *XhoI* sites. The resulting plasmid, pET28-EGC, was used to transform *E. coli*, selecting for its presence in cells by the inclusion of kanamycin (50  $\mu\text{g.mL}^{-1}$ ) in the growth medium.

**Creation of Site-Saturation Mutant Libraries and Site-Directed Mutagenesis.** Saturation mutagenesis was carried out using the Quick-Change mutagenesis kit (Agilent). The DNA template for mutagenesis was the plasmid pET28-EGC and mutations were introduced using a series of oligonucleotide primers (Eurogentec), designed using NNK degeneracy at positions E179, N279, Q280, and D311 (Table 1).

**Table 1.** List of primers used for the construction of the mutated sequences

Primer name	Sequence (5' to 3')
E179	GAACCCCAACCCCGCTGG <u>NNK</u> CTGTATTACATCCAGCCAG
N279	CCGCAGGCCATTGGCGTT <u>NNK</u> CAAGGTTTACCATCGGGC
Q280	CCGCAGGCCATTGGCGTTAAT <u>NNK</u> GGTTTACCATCGGGCT
D311	GTACCCGCTGCCATTG <u>NNK</u> ATCGGCGACGGCCACG
D314Y	GCCATTGGACATCGGCTATGGCCACGAAGGACTTG

Mismatched bases are underlined; N is A, G, C, or T; K is G or T; M is A or C; and MNN the reverse complement of NNK.

To perform mutagenesis an appropriate reaction mixture containing all of the necessary reagents (template DNA, a single primer pair, buffer, deoxyribonucleotides and thermostable DNA polymerase) was placed in a thermocycler and submitted to 16 repetitions of the following protocol: 95 °C for 30 s, 55 °C for 1 min and 68 °C for 13 min for the four NNK libraries and 7 min in the case of D314Y. After, any remaining parental DNA template was

eliminated by the addition of *Dpn* I to the reaction mixture (1 h at 37 °C with 10 units of *Dpn*I) before using 2 µL to transform chemically competent *E. coli* TOP10 or *E. coli* XL10-Gold cells. After growth, transformants (always > 200 cfus) were pooled and total plasmid DNA was extracted and stored at -20 °C. In this way, four mutant libraries were prepared. The successful introduction of mutations and their diversity as well as conservation of sequence integrity were confirmed by full-length gene sequencing of five to ten clones per library (GATC Biotech, Mulhouse, France).

**Mutant Library Screening.** For library screening, an aliquot of the pooled plasmid DNA was used to transform *E. coli* BL21 DE3 cells. After growth, 96 cfus were selected and dispersed in 200 µL LB medium in microplates and grown at 37 °C with shaking (700 rpm) for 16 h, thus generating pre-cultures that were used to inoculate (30 µL) 1500 µL of fresh TB medium in deep wells of 96-microplates. Following growth for 2.5 h at 37 °C with shaking (700 rpm) protein production was induced by the addition of IPTG (to a final concentration of 0.6 mM) and incubation resumed for 16 h at 20 °C. At the end of this period, cells were recovered using centrifugation (2,250g, 30 min) and submitted to the combined lytic effects of lysozyme (0.5 g/L) and freeze/thaw (-80 °C and 37 °C respectively) treatments. After centrifugation, the supernatants were submitted to a dual enzyme assay, measuring activity on the substrate CNP-β-Celb (5 mM) in the absence and presence of acceptors **1** and **3** (at 25 mM). The ratios between the two activities were calculated to search potential activated clones in presence of acceptor.

**Expression and Purification of EGC and Mutants.** *E. coli* BL21 DE3 cells containing the plasmid pET28-EGC were grown to saturation in selective LB medium (containing kanamycin, 50 µg.mL<sup>-1</sup>) for 16 h at 37 °C with shaking (140 rpm), then 0.5 mL was used to inoculate a fresh aliquot (50 mL) of culture medium and growth was pursued. Induction of protein expression was achieved by addition of IPTG (1 mM final) during the exponential growth phase (OD<sub>600</sub> nm of 0.5-0.6), and was allowed to continue for 18 h at 20 °C and 140 rpm. Afterwards, the culture was subjected to centrifugation (5,000g, 15 min, 10 °C) to recover the cell pellet, which was suspended in 1 mL of TALON buffer (20 mM Tris-HCl, pH 8.0, with 300 mM NaCl). Sonication (Vibracell Bioblock Scientific 72434) of the cell suspension was performed on ice during 8 min (0.5 s 'on', 1.5 s 'off', with the probe operating at 30% of maximal power), allowing the recovery of a cell free extract after centrifugation (11,000g, 30 min, 4 °C) and filtration (0.22 µm), which was then loaded onto a Cobalt resin (Talon® Metal Affinity Resin, Clontech Laboratories, Inc.) contained in a disposable column (Bio-Spin columns, Bio-Rad). Protein elution was achieved under gravity



using TALON buffer containing increasing imidazole concentrations (one volume each of 5 and 60 mM imidazole, and then two volumes of 150 mM imidazole). Fractions were analyzed using SDS-PAGE (Bio-Rad) and those containing purified enzyme were pooled and dialyzed against sodium acetate buffer (25 mM, pH 5.0) overnight at 4 °C, before being filtered (0.22 µm) and stored at 4 °C. Average enzymes yields per liter of culture were 80 (EGC), 60 (D311Y), and 70 mg (D341Y).

**Evaluating Enzymatic Activities by Thin Layer Chromatography.** Reactions (20 µL) were performed in 50 mM sodium acetate buffer (pH 4.0), containing 1 mg.mL<sup>-1</sup> BSA, Triton X100 (0.4%, v/v), 2 mM CNP-β-Celb, an acceptor (**1-4** at concentration of 2.5 to 25 mM), and enzyme (wild-type, D311Y or D314Y mutants at 150 µg/mL), for 1 h at 37°C in PCR microplates sealed with an aluminum membrane. Heat inactivation of the enzyme (5 min at 95 °C) was used to stop reactions and then water was removed by evaporation. The solid residue was dissolved in methanol (10 µL) and the solution was spotted onto thin layer chromatography plates (TLC - silica gel 60 F254 precoated plates, Merck). Migration was achieved using an acetonitrile/water (8:2, v/v) mobile phase and molecular species were revealed by soaking in an orcinol solution (1 g in 1 L of sulphuric acid/ethanol/water solution, 3:72.5:22.5, v/v/v), under UV illumination (254 nm) and finally by charring.

**Enzyme Kinetics.** Kinetic analyses were prepared in microplates (96-well format, 250 µL-well polystyrene plates) using 200 µL of 50 mM sodium acetate buffer (pH 4.0) containing Triton X100 (0.4%, v/v), CNP-β-Celb (5 mM), BSA (1 mg.mL<sup>-1</sup>) and enzyme (25 µg.mL<sup>-1</sup>). For the assays, microplates were incubated in an Eon spectrophotometer (BioTek) for 5 min at 37 °C prior to enzyme addition, then reactions were conducted for 15 min, monitoring at 1-min intervals the release of 2-chloro-4-nitrophenol (CNP) at 405 nm. Enzyme activity was determined by comparing data to a standard curve prepared using CNP (in the range (0 to 1 mM). Negative controls containing all of the reactants except the enzyme were always included in order to correct for spontaneous hydrolysis of the substrate and all reactions were performed at least in triplicate. The kinetic parameters  $K_M$ ,  $k_{cat}$  and the efficiency coefficient  $k_{cat}/K_M$  were determined by measuring enzyme activity at various concentrations of CNP-β-Celb (0.5 to 18 mM) and then by modeling data using the Michaelis-Menten equation and the global curve fitting module in SigmaPlot 11.0 software (Systat software Inc., Ritme, Paris, France). One unit of enzyme was designated the amount of enzyme required to release 1 µmol of CNP from CNP-β-Celb per minute.

To determine the optimum pH for activity the standard assay with 2 mM CNP-β-Celb was performed in buffers (50 mM sodium acetate, pH 3.5 to 6.0, and 50 mM sodium phosphate,

pH 5.0 to 7.0) covering the pH range 3.5 to 7.0, ensuring that the volume of enzyme solution did not exceed 30  $\mu\text{L}$  (final volume of 200  $\mu\text{L}$ ) to prevent a pH change. For each pH value, a standard CNP curve was prepared. Similarly, at the optimal pH for enzyme activity, reactions were performed at eight temperatures in the range 30 to 65  $^{\circ}\text{C}$  to determine optimal temperature. Finally, the thermostability of enzymes was evaluated by incubating enzymes at 37 or 50  $^{\circ}\text{C}$ , removing samples over a 7 h-period. Afterwards, residual activity was measured at 37  $^{\circ}\text{C}$  using the assay described above containing 2 mM CNP- $\beta$ -Celb.

**Transglycosylation Reactions.** 20 mL of 50 mM sodium acetate buffer (pH 4.0), containing Triton X100 (0.4%, v/v), BSA (1  $\text{mg}\cdot\text{mL}^{-1}$ ), CNP- $\beta$ -Celb (8 mM), an acceptor molecule (100 mM of **1-4**) was equilibrated at 37  $^{\circ}\text{C}$  for 10 min before adding D311Y (160  $\mu\text{g}\cdot\text{mL}^{-1}$ ). Afterwards, the reaction was incubated for 45 min at 37  $^{\circ}\text{C}$ , under stirring and then stopped by heat inactivation (10 min at 95  $^{\circ}\text{C}$ ), freezing in liquid nitrogen and lyophilisation. The APGs **5-8** were purified after acetylation (in a 1:1 v/v mixture of acetic anhydride/pyridine and in the presence of a catalytic amount of 4-dimethylaminopyridine) using automated Reveleris® flash column chromatography and a 20 to 100% (v/v) gradient of ethyl acetate in petroleum ether. Deacetylation, performed at 0  $^{\circ}\text{C}$  in a mixture of anhydrous methanol and dichloromethane (1:1, v/v) with sodium methoxide (1 M in methanol, 0.5 eq) and acidic workup with Amberlite IR-120 ( $\text{H}^{+}$ ) resin afforded the expected APGs **9-12**. The compounds were purified by automated flash column chromatography using a 0 to 60% gradient of methanol in dichloromethane. NMR spectra were recorded at 298 K on a Bruker Avance II 500 spectrometer. Coupling constants ( $J$ ) are reported in Hz, chemical shifts ( $\delta$ ) are given in ppm. Multiplicities are reported as follows: s = singlet, d= doublet, t= triplet, m= multiplet, br = broad, dd= doublet of doublets. Roman numerals in ascending order are given to the D-Glcp residues from the reducing end. Analysis and assignments were made using 1D ( $^1\text{H}$ ,  $^{13}\text{C}$  and Jmod) and 2D (COSY, HSQC and HMBC) spectra. High-resolution mass spectrometry experiment was performed at the CRMPO (Centre Régional de Mesures Physiques de l'Ouest, University of Rennes, France) on a Thermo Fisher Q-Exactive or a Waters Q-ToF 2 spectrophotometers, equipped with an electrospray ionization source operating in positive ion detection mode (ESI $^{+}$ ).

*3-Acetate-pent-1-yl and 1-ethyl-(3-acetate)-prop-1-yl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 4)-3,4,6-tri-O-acetyl- $\beta$ -D-glucopyranoside (5).* Mixture of regioisomers.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz, 298K):  $\delta$  5.19-5.12 (m, 2H, H-3 of Glc $^{\text{I}}$  and Glc $^{\text{II}}$ ), 5.06 (t,  $J_{3,4} = J_{4,5} = 9.7$  Hz, 1H, H-4 of Glc $^{\text{II}}$ ), 4.94-4.84 (m, 3H, H-2 of Glc $^{\text{II}}$  and I and CHO), 4.52-4.50 (m, 2H, H-6a of Glc $^{\text{I}}$  and H-1 of Glc $^{\text{II}}$ ), 4.44 (d,  $J_{1,2} = 7.9$  Hz, 0.4H, H-1 of Glc $^{\text{I}}$ ), 4.42 (d,  $J_{1,2} = 8.0$  Hz,

0.6H, H-1 of Glc<sup>I</sup>), 4.36 (br dd,  $J_{5,6a} = 4.4$  and  $J_{6a,6b} = 12.5$  Hz, 1H, H-6a of Glc<sup>II</sup>), 4.10-4.03 (m, 2H, H-6b of Glc<sup>I and II</sup>), 3.89-3.79 (m, 1H, OCH<sub>2</sub>), 3.76 (br t,  $J_{3,4} = J_{4,5} = 9.4$  Hz, 1H, H-4 of Glc<sup>I</sup>), 3.67-3.63 (m, 1H, H-5 of Glc<sup>II</sup>), 3.59-3.56 (m, 1H, H-5 of Glc<sup>I</sup>), 3.54-3.49 (m, 0.4H, OCH<sub>2</sub>), 3.45-3.40 (m, 0.6H, OCH<sub>2</sub>), 2.13-1.98 (8s, 24H, CH<sub>3</sub> of acetyl), 1.86-1.72 (m, 2H, OCH<sub>2</sub>CH<sub>2</sub>CH), 1.60-1.53 (m, 2H, CHCH<sub>2</sub>CH<sub>3</sub>), 0.88-0.85 (m, 3H, CH<sub>3</sub> of aglycone); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 125 MHz, 298K):  $\delta$  170.6-169.0 (C(=O)CH<sub>3</sub>), 100.8, 100.8, 100.7 and 100.3 (C-1 of Glc<sup>I and II</sup>), 76.5 and 76.4 (C-4 of Glc<sup>I</sup>), 72.9 (C-3 of Glc<sup>II</sup>), 72.6, 72.6, 72.6, 72.5, 72.4 and 72.3 (C-3 and C-5 of Glc<sup>I</sup> and CHO), 72.0 (C-5 of Glc<sup>II</sup>), 71.6, 71.6, 71.5 and 71.4 (C-2 of Glc<sup>I and II</sup>), 67.8 (C-4 of Glc<sup>II</sup>), 66.4 and 66.2 (OCH<sub>2</sub>), 61.8 (C-6 of Glc<sup>I</sup>), 61.5 (C-6 of Glc<sup>II</sup>), 33.6 and 33.3 (OCH<sub>2</sub>CH<sub>2</sub>CH), 27.2 and 27.0 (CHCH<sub>2</sub>CH<sub>3</sub>), 21.1-20.5 (CH<sub>3</sub> of acetyl), 9.4 and 9.4 (CH<sub>3</sub> of aglycone). HR-ESI-MS  $m/z$  787.2631 (0 ppm, calcd for C<sub>33</sub>H<sub>48</sub>O<sub>20</sub>Na [M+Na]<sup>+</sup> 787.26311).

*3-Hydroxy-pent-1-yl and 1-ethyl-(3-hydroxy)-prop-1-yl  $\beta$ -D-glucopyranosyl-(1  $\rightarrow$  4)-  $\beta$ -D-glucopyranoside (9).* Mixture of regioisomers. <sup>1</sup>H NMR (CD<sub>3</sub>OD, 500 MHz, 298K):  $\delta$  4.41 and 4.40 (2d,  $J_{1,2} = 7.9$  Hz, 1H, H-1 of Glc<sup>II</sup>), 4.30 (2d,  $J_{1,2} = 7.9$  Hz, 1H, H-1 of Glc<sup>I</sup>), 4.05-3.98 (m, 1H, OCH<sub>2</sub>), 3.92-3.83 (m, 3H, H-6a of Glc<sup>I and II</sup> and H-6b of Glc<sup>I or II</sup>), 3.73-3.62 (m, 4H including 1H of OCH<sub>2</sub>, H-6b of Glc<sup>I or II</sup> and CHO), 3.58-3.49 (m, 2H including H-3 of Glc<sup>I or II</sup>), 3.42-3.30 (m, 3H including H-3 of Glc<sup>I or II</sup>), 3.28-3.20 (m, 2H, H-2 of Glc<sup>I and II</sup>), 1.84-1.77 and 1.67-1.58 (2m, 1H each, OCH<sub>2</sub>CH<sub>2</sub>CH), 1.54-1.40 (m, 2H, CH<sub>2</sub>CH<sub>3</sub>), 0.95 (t,  $J = 7.4$  Hz, 3H, CH<sub>3</sub>); <sup>13</sup>C NMR (CD<sub>3</sub>OD, 125 MHz, 298K):  $\delta$  104.6 (C-1 of Glc<sup>II</sup>), 104.3 and 104.1 (C-1 of Glc<sup>I</sup>), 80.8, 80.7 and 78.1 (C<sub>sugar</sub>), 77.9 (C-3 of Glc<sup>I or II</sup>), 76.5 and 76.5 (C<sub>sugar</sub> including C-3 of Glc<sup>I or II</sup>), 74.9, 74.9 and 74.8 (C-2 of Glc<sup>I and II</sup>), 71.4, 71.1 and 70.9 (C<sub>sugar</sub> and CHO), 68.0 (OCH<sub>2</sub>), 62.4 and 61.9 (C-6 of Glc<sup>I and II</sup>), 37.8 and 37.7 (OCH<sub>2</sub>CH<sub>2</sub>CH), 33.3 and 31.3 (CH<sub>2</sub>CH<sub>3</sub>), 10.3 and 10.3 (CH<sub>3</sub>). HR-ESI-MS  $m/z$  451.1790 (0 ppm, calcd for C<sub>17</sub>H<sub>32</sub>O<sub>12</sub>Na [M+Na]<sup>+</sup> 451.17915).

*3-Acetate-oct-1-yl and 1-(2-acetate)-ethyl-hex-1-yl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-glucopyranosyl-(1  $\rightarrow$  4)-3,4,6-tri-O-acetyl- $\beta$ -D-glucopyranoside (6).* Mixture of regioisomers. <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz, 298K):  $\delta$  5.18-5.11 (m, 2H, H-3 of Glc<sup>I and II</sup>), 5.05 (t,  $J_{3,4} = J_{4,5} = 9.7$  Hz, 1H, H-4 of Glc<sup>II</sup>), 4.96-4.85 (m, 3H, H-2 of Glc<sup>II and I</sup> and CHO), 4.51-4.48 (m, 2H, H-1 of Glc<sup>II</sup> and H-6a of Glc<sup>I</sup>), 4.43 (d,  $J_{1,2} = 7.9$  Hz, 0.6H, H-1 of Glc<sup>I</sup>), 4.41 (d,  $J_{1,2} = 8.0$  Hz, 0.4H, H-1 of Glc<sup>I</sup>), 4.36 and 4.35 (2dd,  $J_{5,6a} = 4.5$  and  $4.4$  and  $J_{6a,6b} = 12.5$  and  $12.4$  Hz respectively, 1H, H-6a of Glc<sup>II</sup>), 4.10-4.02 (m, 2H, H-6b of Glc<sup>I and II</sup>), 3.87-3.73 (m, 2H, 1H of OCH<sub>2</sub> and H-4 of Glc<sup>I</sup>), 3.66-3.63 (m, 1H, H-5 of Glc<sup>II</sup>), 3.58-3.55 (m, 1H, H-5 of Glc<sup>I</sup>), 3.52-

3.47 (m, 0.6H of OCH<sub>2</sub>), 3.43-3.38 (m, 0.4H of OCH<sub>2</sub>), 2.12-1.97 (8s, 24H, CH<sub>3</sub> of acetyl), 1.84-1.70 (m, 2H, OCH<sub>2</sub>CH<sub>2</sub>CH), 1.56-1.44 (m, 2H, CHCH<sub>2</sub>(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>), 1.32-1.19 (m, 6H, CHCH<sub>2</sub>(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>), 0.88-0.85 (m, 3H, CH<sub>3</sub> of aglycone); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 125 MHz, 298K): δ 170.6-169.0 (C(=O)CH<sub>3</sub>), 100.8, 100.7, 100.7 and 100.3 (C-1 of Glc<sup>I and II</sup>), 76.5 and 76.4 (C-4 of Glc<sup>I</sup>), 72.9, 72.6, 72.6, 72.6 and 72.4 (C-3 of Glc<sup>I and II</sup> and C-5 of Glc<sup>I</sup>), 71.9 (C-5 of Glc<sup>II</sup>), 71.6, 71.6, 71.5, 71.4, 71.4 and 71.1 (C-2 of Glc<sup>I and II</sup> and CHO), 67.8 (C-4 of Glc<sup>II</sup>), 66.3 and 66.2 (OCH<sub>2</sub>), 61.8 (C-6 of Glc<sup>I</sup>), 61.5 (C-6 of Glc<sup>II</sup>), 34.4, 34.2, 34.1 and 33.8 (OCH<sub>2</sub>CH<sub>2</sub>CH and CHCH<sub>2</sub>(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>), 31.6, 31.5, 24.8 and 22.5 (CHCH<sub>2</sub>(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>), 21.1-20.5 (CH<sub>3</sub> of acetyl), 13.9 (CH<sub>3</sub> of aglycone). HR-ESI-MS *m/z* 829.3098 (0 ppm, calcd for C<sub>36</sub>H<sub>54</sub>O<sub>20</sub>Na [M+Na]<sup>+</sup> 829.31006).

*3-Hydroxy-oct-1-yl and 1-(2-hydroxy)-ethyl-hex-1-yl β-D-glucopyranosyl-(1 → 4)- β -D-glucopyranoside (10)*. Mixture of regiosomers. <sup>1</sup>H NMR (CD<sub>3</sub>OD, 500 MHz, 298K): δ 4.41 and 4.40 (2d, *J*<sub>1,2</sub> = 7.9 Hz, 1H, H-1 of Glc<sup>II</sup>), 4.30 (2d, *J*<sub>1,2</sub> = 7.9 Hz, 2H, H-1 of Glc<sup>I</sup>), 4.05-3.97 (m, 1H of OCH<sub>2</sub>), 3.92-3.83 (m, 3H, H-6a of Glc<sup>I and II</sup> and H-6b of Glc<sup>I or II</sup>), 3.77-3.63 (m, 4H including 1H of OCH<sub>2</sub>, CHO and H-6b of Glc<sup>I or II</sup>), 3.58-3.49 (m, 2H including H-3 of Glc<sup>I or II</sup>), 3.42-3.33 (m, 3H including H-3 of Glc<sup>I or II</sup>), 3.28-3.21 (m, 2H, H-2 of Glc<sup>I and II</sup>), 1.83-1.77 and 1.66-1.59 (2m, 1H each, OCH<sub>2</sub>CH<sub>2</sub>CH), 1.49-1.42 (m, 3H, CHCH<sub>2</sub>(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub> and 1H of CHCH<sub>2</sub>(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>), 1.39-1.25 (m, 5H of CHCH<sub>2</sub>(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>), 0.91 (t, *J* = 7.0 Hz, 3H, CH<sub>3</sub> of aglycone); <sup>13</sup>C NMR (CD<sub>3</sub>OD, 125 MHz, 298K): δ 104.6 (C-1 of Glc<sup>II</sup>), 104.3 and 104.1 (C-1 of Glc<sup>I</sup>), 80.8, 80.7, 78.1, 77.9 and 76.5 (C<sub>sugar</sub>), 74.9, 74.9 and 74.8 (C-2 of Glc<sup>I and II</sup>), 71.4 (C<sub>sugar</sub>), 69.7 and 69.5 (CHO), 68.0 (OCH<sub>2</sub>), 62.4 and 61.9 (C-6 of Glc<sup>I and II</sup>), 38.7, 38.6, 38.3 and 38.2 (OCH<sub>2</sub>CH<sub>2</sub>CH and CHCH<sub>2</sub>CH<sub>3</sub>), 33.1, 33.1, 26.5 and 23.7 (CHCH<sub>2</sub>(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>), 14.4 (CH<sub>3</sub> of aglycone). HR-ESI-MS *m/z* 493.2263 (0 ppm, calcd for C<sub>20</sub>H<sub>38</sub>O<sub>12</sub>Na [M+Na]<sup>+</sup> 493.2261).

*2-Oxo-pent-5-yl 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl-(1 → 4)-3,4,6-tri-O-acetyl-β-D-glucopyranoside (7)*. <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz, 298K): δ 5.16 (dd, *J*<sub>3,4</sub> = 9.3 Hz, 1H, H-3 of Glc<sup>I</sup>), 5.14 (t, *J*<sub>3,4</sub> = 9.4 Hz, 1H, H-3 of Glc<sup>II</sup>), 5.06 (t, *J*<sub>4,5</sub> = 9.8 Hz, 1H, H-4 of Glc<sup>II</sup>), 4.92 (dd, *J*<sub>2,3</sub> = 9.4 Hz, 1H, H-2 of Glc<sup>II</sup>), 4.88 (dd, *J*<sub>2,3</sub> = 9.7 Hz, 1H, H-2 of Glc<sup>I</sup>), 4.50 (d, *J*<sub>1,2</sub> = 7.9 Hz, 1H, H-1 of Glc<sup>II</sup>), 4.50 (dd, *J*<sub>5,6a</sub> = 2.1 Hz, 1H, H-6a of Glc<sup>I</sup>), 4.42 (d, *J*<sub>1,2</sub> = 8.0 Hz, 1H, H-1 of Glc<sup>I</sup>), 4.36 (dd, *J*<sub>5,6a</sub> = 4.4 et *J*<sub>6a,6b</sub> = 12.5 Hz, 1H, H-6a of Glc<sup>II</sup>), 4.09 (dd, *J*<sub>5,6b</sub> = 5.1 et *J*<sub>6a,6b</sub> = 12.0 Hz, 1H, H-6b of Glc<sup>I</sup>), 4.04 (dd, *J*<sub>5,6b</sub> = 2.3 Hz, 1H, H-6b of Glc<sup>II</sup>), 3.82-3.78 (m, 1H of OCH<sub>2</sub>), 3.75 (br t, *J*<sub>4,5</sub> = 9.8 Hz, 1H, H-4 of Glc<sup>I</sup>), 3.67-3.63 (m, 1H, H-5 of Glc<sup>II</sup>), 3.59-3.55 (m, 1H, H-5 of Glc<sup>I</sup>), 3.52-3.48 (m, 1H of OCH<sub>2</sub>), 2.48 (t, *J*<sub>3,4'</sub> = 7.1 Hz, 2H, CH<sub>2</sub>C(O)),

2.12 (s, 3H, CH<sub>3</sub> of aglycone), 2.12-1.97 (7s, 21H, CH<sub>3</sub> of acetyl), 1.87-1.78 (m, 2H, CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 125 MHz, 298K): δ 208.1 (C(=O)CH<sub>3</sub> of aglycone), 170.5-169.0 (C(=O)CH<sub>3</sub> of acetyl), 100.8 (C-1 of Glc<sup>II</sup>), 100.6 (C-1 of Glc<sup>I</sup>), 76.5 (C-4 of Glc<sup>I</sup>), 72.9 (C-3 of Glc<sup>II</sup>), 72.7 (C-5 of Glc<sup>I</sup>), 72.5 (C-3 of Glc<sup>I</sup>), 72.0 (C-5 of Glc<sup>II</sup>), 71.6 (C-2 of Glc<sup>II</sup>), 71.6 (C-2 of Glc<sup>I</sup>), 69.0 (OCH<sub>2</sub>), 67.8 (C-4 of Glc<sup>II</sup>), 61.8 (C-6 of Glc<sup>I</sup>), 61.6 (C-6 of Glc<sup>II</sup>), 39.7 (CH<sub>2</sub>C(=O)), 29.9 (CH<sub>3</sub> of aglycone), 23.6 (CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>), 20.8-20.5 (CH<sub>3</sub> of acetyl). HR-ESI-MS *m/z* 743.2368 (0 ppm, calcd for C<sub>31</sub>H<sub>44</sub>O<sub>19</sub>Na [M+Na]<sup>+</sup> 743.2369).

*2-Oxo-pent-5-yl β-D-glucopyranosyl-(1 → 4)- β-D-glucopyranoside (11)*. <sup>1</sup>H NMR (CD<sub>3</sub>OD, 500 MHz, 298K): δ 4.41 (d, *J*<sub>1,2</sub> = 7.9 Hz, 1H, H-1 of Glc<sup>II</sup>), 4.26 (d, *J*<sub>1,2</sub> = 7.9 Hz, 1H, H-1 of Glc<sup>I</sup>), 3.92-3.83 (m, 4H, 1H of OCH<sub>2</sub>, H-6a of Glc<sup>I and II</sup> and H-6b of Glc<sup>I or II</sup>), 3.66 (dd, *J*<sub>5,6a</sub> = 5.6 and *J*<sub>6a,6b</sub> = 11.9 Hz, 1H, H-6b of Glc<sup>I or II</sup>), 3.58-3.54 (m, 2H including 1H of OCH<sub>2</sub>), 3.50 (t, *J*<sub>2,3</sub> = *J*<sub>3,4</sub> = 8.9 Hz, 1H, H-3 of Glc<sup>I or II</sup>), 3.40-3.28 (m, 4H including H-3 of Glc<sup>I or II</sup>), 3.24-3.20 (m, 2H, H-2 of Glc<sup>I and II</sup>), 2.63 (br t, *J*<sub>3',4'</sub> = 7.2 Hz, 2H, CH<sub>2</sub>C(=O)), 2.15 (s, 3H, CH<sub>3</sub> of aglycone), 1.88-1.82 (m, 2H, CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>); <sup>13</sup>C NMR (CD<sub>3</sub>OD, 125 MHz, 298K): δ 211.8 (C(=O)CH<sub>3</sub> of aglycone), 104.6 (C-1 of Glc<sup>II</sup>), 104.2 (C-1 of Glc<sup>I</sup>), 80.7, 78.1 and 77.9 (C<sub>sugar</sub>), 76.4 and 76.4 (C-3 of Glc<sup>I and II</sup>), 74.9 and 74.8 (C-2 of Glc<sup>I and II</sup>), 71.4 (C<sub>sugar</sub>), 69.9 (OCH<sub>2</sub>), 62.4 and 61.9 (C-6 of Glc<sup>I and II</sup>), 40.7 (CH<sub>2</sub>C(=O)), 29.9 (CH<sub>3</sub> of aglycone), 24.9 (CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>). HR-ESI-MS *m/z* 449.1635 (0 ppm, calcd for C<sub>17</sub>H<sub>30</sub>O<sub>12</sub>Na [M+Na]<sup>+</sup> 449.1635).

*4-Oxo-oct-1-yl 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl-(1 → 4)-3,4,6-tri-O-acetyl-β-D-glucopyranoside (8)*. <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz, 298K): δ 5.17-5.11 (m, 2H, H-3 of Glc<sup>I and II</sup>), 5.05 (br t, *J*<sub>3,4</sub> = *J*<sub>4,5</sub> = 9.7 Hz, 1H, H-4 of Glc<sup>II</sup>), 4.91 (dd, *J*<sub>2,3</sub> = 9.3 Hz, 1H, H-2 of Glc<sup>II</sup>), 4.87 (dd, *J*<sub>2,3</sub> = 9.7 Hz, 1H, H-2 of Glc<sup>I</sup>), 4.50 (d, *J*<sub>1,2</sub> = 8.0 Hz, 1H, H-1 of Glc<sup>II</sup>), 4.49 (dd, *J*<sub>5,6a</sub> = 2.1 and *J*<sub>6a,6b</sub> = 12.0 Hz, 1H, H-6a of Glc<sup>I</sup>), 4.42 (d, *J*<sub>1,2</sub> = 8.0 Hz, 1H, H-1 of Glc<sup>I</sup>), 4.36 (dd, *J*<sub>5,6a</sub> = 4.5 et *J*<sub>6a,6b</sub> = 12.5 Hz, 1H, H-6a of Glc<sup>II</sup>), 4.08 (dd, *J*<sub>5,6b</sub> = 5.2 et *J*<sub>6a,6b</sub> = 12.0 Hz, 1H, H-6b of Glc<sup>I</sup>), 4.04 (dd, *J*<sub>5,6b</sub> = 2.3 and *J*<sub>6a,6b</sub> = 12.4 Hz, 1H, H-6b of Glc<sup>II</sup>), 3.82-3.77 (m, 1H of OCH<sub>2</sub>), 3.75 (t, *J*<sub>3,4</sub> = *J*<sub>4,5</sub> = 9.4 Hz, 1H, H-4 of Glc<sup>I</sup>), 3.66-3.63 (m, 1H, H-5 of Glc<sup>II</sup>), 3.58-3.55 (m, 1H, H-5 of Glc<sup>I</sup>), 3.51-3.47 (m, 1H of OCH<sub>2</sub>), 2.44 (t, *J*<sub>2',3'</sub> = 7.2 Hz, 2H, OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>C(O)), 2.37 (t, *J*<sub>5',6'</sub> = 7.5 Hz, 2H, C(O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 2.12-1.97 (7s, 21H, CH<sub>3</sub> of acetyl), 1.85-1.78 (m, 2H, OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>C(O)), 1.56-1.50 (m, 2H, C(O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 1.32-1.27 (m, 2H, C(O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 0.89 (t, *J*<sub>7',8'</sub> = 7.4 Hz, 3H, CH<sub>3</sub> of aglycone); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 125 MHz, 298K): δ 210.6 (C(=O) of aglycone), 170.5-169.0 (C(=O)CH<sub>3</sub> of acetyl), 100.8 (C-1 of Glc<sup>II</sup>), 100.6 (C-1 of Glc<sup>I</sup>), 76.5 (C-4 of Glc<sup>I</sup>), 72.9

(C-3 of Glc<sup>II</sup>), 72.7 (C-5 of Glc<sup>I</sup>), 72.5 (C-3 of Glc<sup>I</sup>), 72.0 (C-5 of Glc<sup>II</sup>), 71.6 and 71.6 (C-2 of Glc<sup>I and II</sup>), 69.1 (OCH<sub>2</sub>), 67.8 (C-4 of Glc<sup>II</sup>), 61.8 (C-6 of Glc<sup>I</sup>), 61.6 (C-6 of Glc<sup>II</sup>), 42.6 (C(O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 38.6 (OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>C(O)), 25.9 (C(O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 23.6 (OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>C(O)), 22.3 (C(O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 20.8-20.5 (CH<sub>3</sub> of acetyl), 13.8 (CH<sub>3</sub> of aglycone). HR-ESI-MS *m/z* 785.2836 (0 ppm, calcd for C<sub>34</sub>H<sub>50</sub>O<sub>19</sub>Na [M+Na]<sup>+</sup> 785.28385).

*4-Oxo-oct-1-yl β-D-glucopyranosyl-(1→4)-β-D-glucopyranoside (12)*. <sup>1</sup>H NMR (CD<sub>3</sub>OD, 500 MHz, 298K): δ 4.41 (d, *J*<sub>1,2</sub> = 7.9 Hz, 1H, H-1 of Glc<sup>II</sup>), 4.29 (d, *J*<sub>1,2</sub> = 7.9 Hz, 1H, H-1 of Glc<sup>I</sup>), 3.90-3.83 (m, 4H, 1H of OCH<sub>2</sub>, H-6a of Glc<sup>I and II</sup> and H-6b of Glc<sup>I or II</sup>), 3.68-3.62 (m, 2H including H-6b of Glc<sup>I or II</sup>), 3.58-3.53 (m, 2H including 1H of OCH<sub>2</sub>), 3.50 (t, *J*<sub>2,3</sub> = *J*<sub>3,4</sub> = 8.9 Hz, 1H, H-3 of Glc<sup>I or II</sup>), 3.40-3.33 (m, 3H including H-3 of Glc<sup>I or II</sup>), 3.24-3.20 (m, 2H, H-2 of Glc<sup>I and II</sup>), 2.60 (t, *J*<sub>2',3'</sub> = 7.2 Hz, 2H, OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>C(=O)), 2.47 (t, *J*<sub>5',6'</sub> = 7.4 Hz, 2H, C(=O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 1.88-1.82 (m, 2H, OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>C(=O)), 1.57-1.51 (m, 2H, C(=O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 1.36-1.28 (m, 2H, C(=O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 0.92 (s, 3H, CH<sub>3</sub>); <sup>13</sup>C NMR (CD<sub>3</sub>OD, 125 MHz, 298K): δ 214.0 (C(O)CH<sub>3</sub> of aglycone), 104.6 (C-1 of Glc<sup>II</sup>), 104.2 (C-1 of Glc<sup>I</sup>), 80.7, 78.1 and 77.9 (C<sub>sugar</sub>), 76.4 and 76.4 (C-3 of Glc<sup>I and II</sup>), 74.9 and 74.8 (C-2 of Glc<sup>I and II</sup>), 71.4 (C<sub>sugar</sub>), 69.9 (OCH<sub>2</sub>), 62.4 and 61.9 (C-6 of Glc<sup>I and II</sup>), 43.3 (C(=O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 39.9 (OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>C(=O)), 27.0 (C(=O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 25.0 (OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>C(=O)), 23.4 (C(=O)CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 14.2 (CH<sub>3</sub>). HR-ESI-MS *m/z* 491.2100 (1 ppm, calcd for C<sub>20</sub>H<sub>38</sub>O<sub>12</sub>Na [M+Na]<sup>+</sup> 491.21045).

**<sup>1</sup>H NMR Monitoring of Enzyme Reactions and Selection of Significant Protons.** Reactions (5 mL) in 20 mM sodium acetate buffer (pH 4.0), containing Triton X-100 (0.2%, v/v) BSA (0.5 mg.mL<sup>-1</sup>), CNP-β-Celb or F-β-Celb as donor (5 mM) and an acceptor (62.5 mM of **1**, **2**, **3** or **4**), were incubated at 37 °C for 5 min then 200 μL aliquot was removed and designated t = 0. Afterwards, enzyme was added to start the reaction (wild-type enzyme at a final concentration of 150 μg.mL<sup>-1</sup> when using **1**, **2**, and **4**, and 120 μg.mL<sup>-1</sup> for the reaction containing **3**, or D311Y at a final concentration of 100 μg.mL<sup>-1</sup> irrespective of the acceptor) and 200 μL samples were removed at intervals over 180 min, inactivating the enzyme using a heat (95 °C for 5 min), liquid nitrogen freezing and lyophilization routine. At 120 min, more enzyme was added reaching 2x the original concentration. In the case of D311Y, all reactions were performed in triplicate, but those involving EGC were performed once only. The lyophilized samples were dissolved in 250 μL of CD<sub>3</sub>OD and <sup>1</sup>H NMR spectra were recorded at 298 K on a Bruker Ascend Avance III 800 spectrometer equipped with a samplejet autosampler. Donor conversion, hydrolysis and transglycosylation activities were quantified by integrating the relevant anomeric and aromatic proton signals that are reported in Table 2.

**Table 2.** List of proton signals used for the tracking of the different species involved during enzymatic reactions catalyzed by EGC and mutant thereof

Compounds	Integrated proton	Signal multiplicity, $\delta$ (ppm); $J$ (Hz)	Enzymatic reaction monitoring
CNP- $\beta$ -Celb	H-6' of CNP	d, 7.43 (9.2)	Donor consumption
	H-5' of CNP	dd, 8.19 (2.8 and 9.2)	
	H-3' of CNP	d, 8.32 (2.8)	
CNP	H-6' of CNP	d, 6.68 (9.2)	CNP release (donor consumption)
	H-5' of CNP	dd, 7.94 (2.9 and 9.2)	
	H-3' of CNP	d, 8.15 (2.9)	
CNP- $\beta$ -Celb	H-1 of Glc <sup>I</sup>	d, 5.23 (7.4)	Donor consumption
	H-1 of Glc <sup>II</sup>	d, 4.46 (7.9)	
Cellobiose	H-1 of Glc <sup>Ia</sup>	d, 5.10 (3.8)	Hydrolysis
	H-1 of Glc <sup>Ib</sup>	d, 4.50 (7.8)	
APG <b>9</b>	H-1 of Glc <sup>I</sup>	d, 4.30 (7.9)	Transglycosylation
APG <b>10</b>	H-1 of Glc <sup>I</sup>	d, 4.30 (7.9)	Transglycosylation
APG <b>11</b>	H-1 of Glc <sup>I</sup>	d, 4.26 (7.9)	Transglycosylation
APG <b>12</b>	H-1 of Glc <sup>I</sup>	d, 4.29 (7.9)	Transglycosylation

**Evaluation of the Modulation from Hydrolase to Transglycosidase.** The transfer rate  $R_T$  (%) of the donor substrate to the acceptor, expressed in terms of production ( $\mu\text{mol}$  of product/ $100 \mu\text{mol}$  of consumed substrate) was derived from the plot of product yield as a function of donor substrate conversion. As a rate,  $R_T$  is independent of the duration of the reaction and thus provides a convenient way to compare different enzymes.<sup>14</sup> Initial rates were derived from the linear regions corresponding to maximum donor consumption, which was between 16 and nearly 100% depending on the reaction, with root-mean-square deviations above 0.98.

**Modeling D311Y complexed with substrates.** The 3D structure of EGC (GenBank accession code: AAB67050.1, aminoacids 43 to 490) bearing the mutation D311Y was modelled using MODELLER 9v8.<sup>44</sup> Available EGC structures in the Protein Data Bank were selected for threading with multiple templates: 2OSX (EGC II E351S from *Rhodococcus* sp. in complex with ganglioside GM3), 2OYM (EGC II from *Rhodococcus* sp. in complex with

iminocyclitol) and 2OSY (EGC II E233A from *Rhodococcus* sp. lactosyl-enzyme covalent intermediate). Both the *apo*-enzyme and lactosyl-enzyme were modelled. A slow refinement protocol based on simulated annealing molecular dynamics was applied to generate an ensemble of 10 different structures for each state. CNP- $\beta$ -Celb and acceptors **2** and **4** were docked separately to each modelled D311Y structure in the ensemble with AUTODOCK VINA.<sup>45</sup> The docking search space was confined in a 27.8 x 31.5 x 31.5 Å<sup>3</sup> box centered in the active site. The exhaustiveness level was set to 24 and for each docking calculation, 20 binding modes of the ligands were generated. 200 structures of D311Y in complex with each ligand were thus obtained. Only low energy binding poses were considered for analysis. Distances and surfaces were measured, and pictures of the complexes generated with VMD.<sup>46</sup>

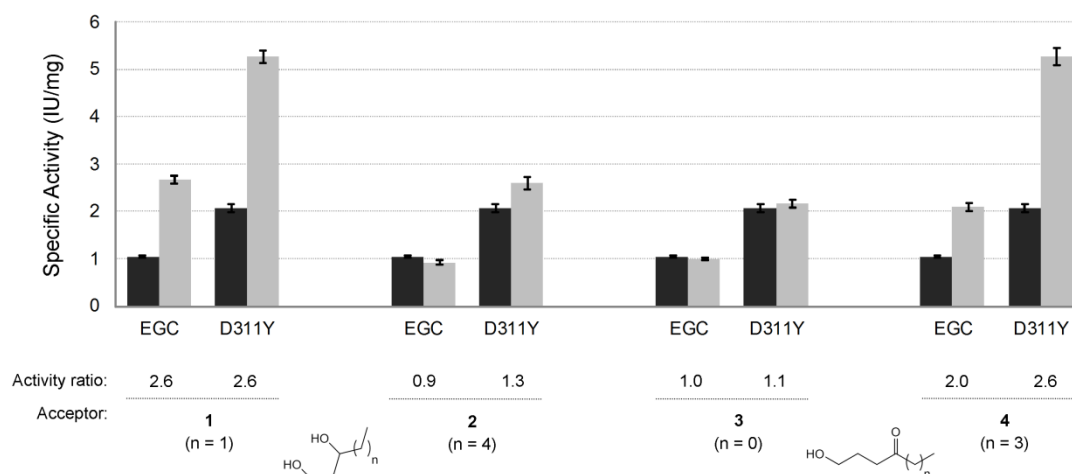
## RESULTS

**Acceptor Preference for Transglycosylation with the Wild-Type Enzyme.** The inherent ability of EGC to transfer cellobiosyl units onto target acceptors was evaluated. Transglycosylation activity was first tested with pentan-1-ol and octan-1-ol, in the presence of 2.5% (v/v) acetonitrile, but both tests were unsuccessful. Similar tests were performed with pentan-1,3-diol (**1**) and 5-hydroxypentan-2-one (**3**), two hydroxylated molecules displaying either hydroxyl or ketone groups that can be considered to some extent mimic parts of the natural ceramide. These tests revealed that although the main reaction was hydrolysis, small amounts of transglycosylation products were obtained (Figure S2). Based on this observation, an enzyme engineering strategy was then deployed with the aim to increase the transglycosylation/hydrolysis (T/H) ratio.

**Site-Saturation Mutagenesis and Detection of Transglycosylating Mutant.** Resolution of the 3D structure of the EGC complexed with ganglioside revealed four amino acids that are located in the aglycone subsite of EGC, where they are thought to interact either directly with the carbonyl group of the ceramide part of glycosphingolipids (N279 and Q280), or through a water molecule (E179 and D311) with the secondary hydroxyl function of the diol from the sphingosine moiety (Figure S3).<sup>27</sup> Therefore these four residues were targeted for site-saturation mutagenesis, creating four libraries that were screened using four different potential acceptor substrates **1-4** (Figure 2) that display different carbon chain lengths and 1,3-diol and  $\delta$ -hydroxyketone functions. Primary screening in 96-well microtiter plates using CNP- $\beta$ -Celb as the donor substrate revealed one isolate (D311Y) that showed 2-fold higher CNP-releasing activity (i.e., a hydrolytic specific activity of  $2.07 \pm 0.08$  IU/mg) when compared to EGC, and increased CNP-releasing activity in the presence of acceptors **1-4** (i.e., conditions favoring



transglycosylation) relative to the activity in the absence of acceptors. This result is similar to that obtained with the wild-type enzyme in the presence of 5-carbon chain acceptors **1** and **3**, but slightly more than EGC in the presence of 8-carbon acceptors **2** and **4** (Figure 2).



**Figure 2.** Measurement of the specific activities (i.e., CNP release) of EGC and D311Y on CNP- $\beta$ -Celb (5 mM) alone (dark bars), or in the presence of 25 mM acceptor **1-4** (grey bars). The activity ratio is the comparison of enzyme activity in reactions performed in the presence or absence of a given acceptor.

**Kinetic Analysis of the Hydrolytic Activities of EGC, D311Y and D314Y.** Determination of the pH and temperature optima and thermal stability of EGC, D311Y and D314Y confirmed that compared to EGC these parameters were mostly unaltered (pH<sub>opt</sub> of 4, T<sub>opt</sub> of approximately 60 °C) in the mutant enzymes (Figure S4). Nevertheless, the half-life of D311Y was more than 2 h at 37 °C, compared to 45 min for EGC and 3 h for D314Y, but at 50 °C it only retained approximately 5% of its activity after 30 min (Figure S4C and Table S1), observations that are consistent with structural data suggesting that mutations at these flexible positions might reduce the B-factor.<sup>47</sup>

The substitution of D314 by tyrosine (i.e., D314Y) has been shown to increase the catalytic activity of the EGC glycosynthase, E351S, when this crippled mutant is used with G<sub>M1</sub>-oligosaccharyl  $\alpha$ -fluoride donor and sphingosine acceptor.<sup>38</sup> Therefore, D314Y was included in our study to compare with EGC and D311Y (Table 3). Apart from a slight increase of  $k_{cat}$  (1.5-fold), the catalytic parameters describing D314Y-mediated hydrolysis of CNP- $\beta$ -Celb

were similar to that of EGC. However, in comparison the  $K_M$  value measured when using D311Y (7.0 mM) was 3.2-fold lower than that observed for EGC (22.4 mM), and the performance constant  $k_{cat}/K_M$  value was higher, even if this increase was attenuated by the slight decrease (0.8-fold) of  $k_{cat}$ .

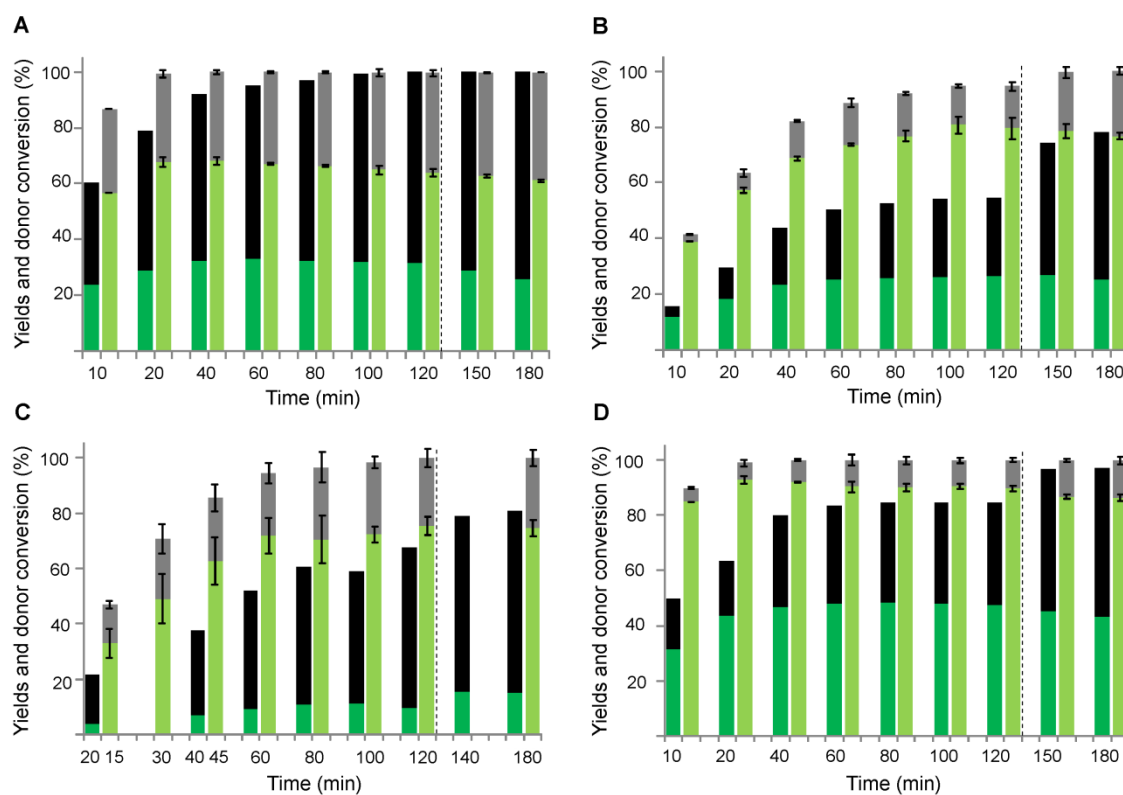
**Table 3.** Steady-state kinetic parameters for the hydrolysis of CNP- $\beta$ -Celb by EGC and mutants D311Y and D314Y<sup>a</sup>

Enzyme	$K_M$ (mM)	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_M$ (s <sup>-1</sup> .mM <sup>-1</sup> )
EGC	22.4 ± 3.8	3.7 ± 0.6	0.17 ± 0.05
D314Y	22.7 ± 3.1	5.4 ± 0.7	0.24 ± 0.06
D311Y	7.0 ± 0.8	2.9 ± 0.3	0.42 ± 0.08

<sup>a</sup> All assays were carried out in triplicate at 37 °C in 50 mM sodium acetate buffer (pH 4).

**Transglycosylation Activity of the Mutant D311Y.** The aptitude of EGC, D311Y and D314Y to perform transglycosylation was first qualitatively assessed using various acceptors and TLC analysis (Figure S2). Although in the case of alkan-1-ol acceptors it was not possible to detect synthesis products when using either D311Y or EGC, in the case of acceptors **1-3** the transglycosylation ability of D311Y was improved compared to that of EGC. It is noteworthy that mutation D314Y, which is thought to modify the position of residue D311,<sup>38</sup> did not reveal any improvement of the transglycosylation ability. Having ascertained this, the donor/acceptor ratio was optimized in favor of transglycosylation (2:25, mM/mM) (Figure S5) and large-scale transglycosylation reactions were performed, procuring APGs **9-12** at yields of 50-81%. Analysis of products **9-12** by NMR spectroscopy and mass spectrometry confirmed that these were all cellobiosides (correlation from HMBC spectra between H-1 of Glc<sup>I</sup> and OCH<sub>n</sub> of the aglycone and/or C-1 of Glc<sup>I</sup> and OCH<sub>n</sub>, with n = 1 or 2) and revealed that APGs **9** and **10** were in each case an equimolar mixture of two regioisomers, meaning that in the presence of diol acceptors D311Y was able to transfer the cellobiosyl moiety onto both the primary and secondary hydroxyl groups. Logically, in the case of APGs **11** and **12** only one product was present. Moreover, the  $\beta$ -stereochemistry of the newly formed *O*-glycosidic linkages in APGs **9-12** was confirmed through the determination of the

experimental coupling constants ( $J_{1,2} = 7.9$  Hz) and the observation of chemical shifts of H-1 of Glc<sup>I</sup> ( $\delta = 4.26$ -4.30 ppm).



**Figure 3.** <sup>1</sup>H NMR monitoring of time-dependent CNP- $\beta$ -Celb donor consumption and T/H modulation in reactions catalyzed by EGC and D311Y with acceptor (A) **1**, (B) **2**, (C) **3** and (D) **4**. Blacks and grey bars represent hydrolysis yields (% cellobiose produced) and dark and light green bars represent transglycosylation yields (% synthesis products **9-12**) for EGC or D311Y respectively. The sum (black + dark green grey + light green), corresponds to the overall progress of the reaction (i.e., donor conversion in %). Enzyme concentration is doubled after sampling at 120 min (black dotted lines), enabling to appreciate secondary hydrolysis. Reactions with D311Y and EGC were performed in triplicate and monoplicate, respectively.

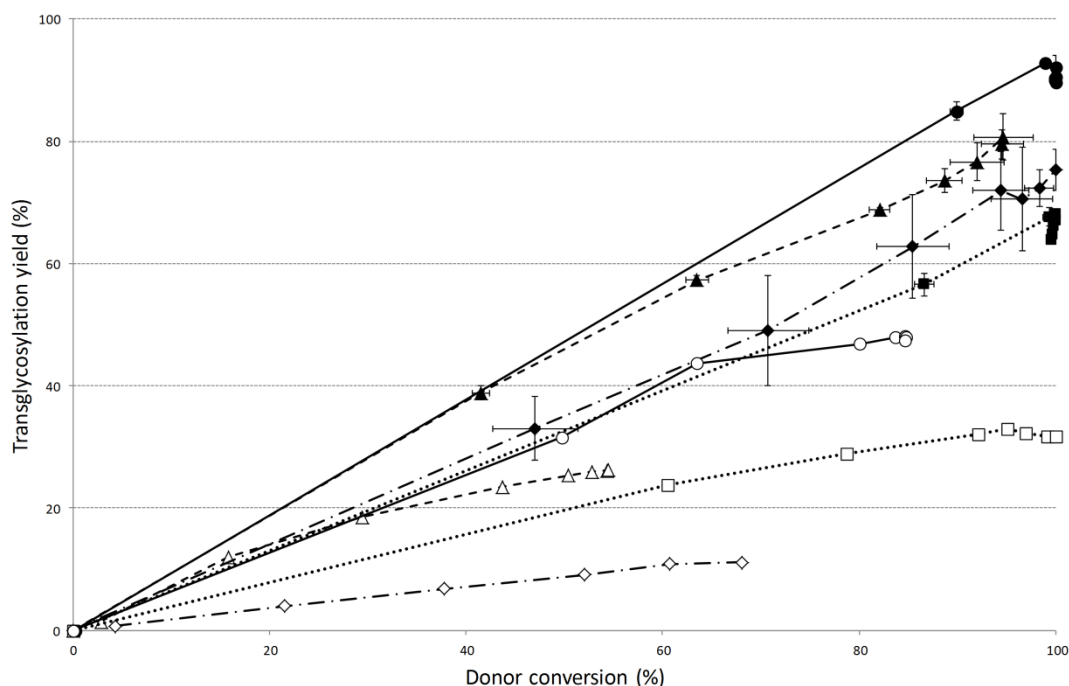
<sup>1</sup>H NMR monitoring of the time-dependent progression of both D311Y-catalyzed transglycosylation and hydrolysis revealed a marked increase of the T/H ratio when compared to EGC (Figures 3 and S6). It is noteworthy that whereas EGC fails to fully consume the donor (even when a fresh enzyme aliquot is added) D311Y quantitatively consumed all the donor in the presence of almost all the acceptors. Moreover, although D311Y-catalyzed

transglycosylation was generally more efficient (i.e., faster product accumulation and higher yields) than that of EGC, irrespective of the acceptor, this was particularly evident when the shorter  $\delta$ -hydroxy ketone **3** was used as the acceptor, since this reaction led to 6.8-fold increase in the transglycosylation yield (75%) compared to the equivalent EGC-catalyzed reaction (Table 4). It is noteworthy that high final yields of APGs **9-12** in reactions catalyzed by D311Y were also the result of the enzyme's inherent low secondary hydrolysis ability (< 6% of their maximal yield), this being underlined by the fact that the addition of fresh enzyme at the end of the reaction did not affect final yields. Likewise, D311Y catalysis was characterized by increased transfer rates ( $R_T$ ), in the range 1.1 to 1.8-fold, for acceptors **1**, **2** and **4**, and up to 4.2-fold for acceptor **3** (Table 4). Concerning the linearity of transglycosylation, in EGC-catalyzed reactions this was conserved up to 80% donor consumption, whereas in D311Y-catalyzed reactions this linear phase was maintained up to 95% donor consumption (Figure 4 and Table S2), underlining the fact that the replacement of D311 by a tyrosine residue confers strong transglycosylase character to EGC.

**Table 4.** Comparison of maximum transglycosylation yields and cellobiosyl transfer rates in EGC and D311Y catalyzed reactions containing acceptors **1-4**

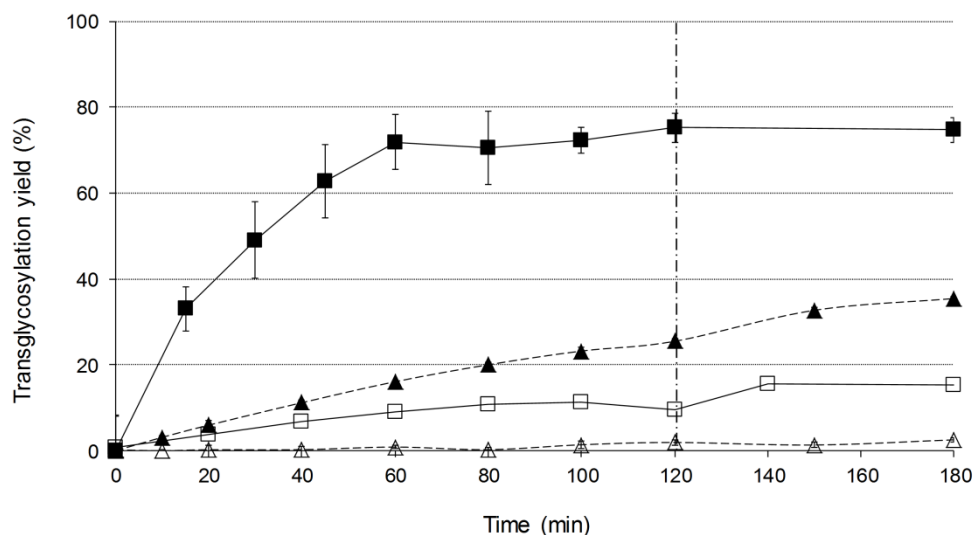
Acceptor	Enzyme	Maximal yield (%) <sup>a</sup>	$R_T$ <sup>b</sup>
Pentan-1,3-diol ( <b>1</b> )	EGC	33	38
	D311Y	68	67
Octan-1,3-diol ( <b>2</b> )	EGC	26	75
	D311Y	81	85
5-Hydroxypentan-2-one ( <b>3</b> )	EGC	11	18
	D311Y	75	74
1-Hydroxyoctan-4-one ( <b>4</b> )	EGC	48	63
	D311Y	93	93

<sup>a</sup> The maximum transglycosylation yield was not reached at the same time for all products. <sup>b</sup> The transfer rate ( $R_T$ , in %) represents the number of  $\mu$ moles of cellobiosyl transferred onto an acceptor for 100  $\mu$ moles of *p*NP- $\beta$ -Celb consumed.



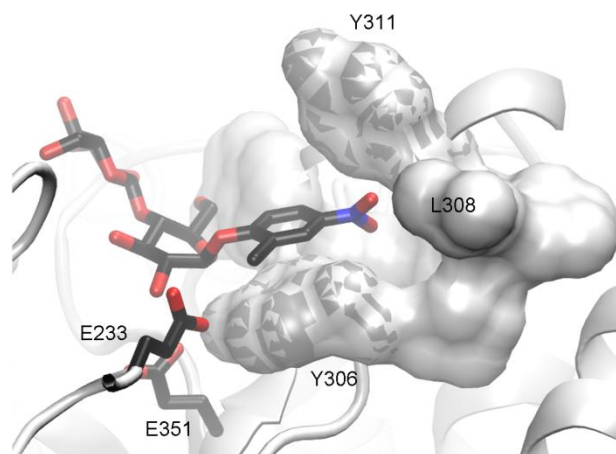
**Figure 4.** Transglycosylation yield evolution as a function of donor conversion with CNP- $\beta$ -Celb (5 mM) as donor substrate and acceptor (62.5 mM): **1** (square), **2** (triangle), **3** (diamond) or **4** (circle), for EGC (open symbols) and D311Y (closed symbols). Transfer rates were calculated up to 120 min of reaction. Reactions with D311Y and EGC were performed in triplicate and monoplicate, respectively.

**Transglycosylation using  $\beta$ -Cellobiosyl Fluoride as Activated Donor.** To compare the impact of the leaving group on transglycosylation,  $\beta$ -cellobiosyl fluoride (F- $\beta$ -Celb)<sup>48</sup> was used as the donor in reactions catalyzed by EGC and D311Y respectively and containing acceptor **3**. The transglycosylation product **11** resulting from these reactions was barely detectable (less than 2% yield) in the EGC-catalyzed reaction, whereas its yield reached 25% when using D311Y (Figure 5). Moreover, this yield could be further improved after 120 min simply by adding a fresh aliquot of D311Y. After 180 min, TLC analysis confirmed that the D311Y-catalyzed reaction was complete, whereas at the same time point in the reaction containing EGC almost all the F- $\beta$ -Celb donor was either unconsumed or hydrolyzed (data not shown).



**Figure 5.** Transglycosylation yield evolution as a function of time with 5-hydroxypentan-2-one (**3**, at 62.5 mM) as acceptor and F-β-Celb (triangle) or CNP-β-Celb (square) as donor substrate (5 mM) for EGC (open symbols) and D311Y (closed form). Enzyme concentration is doubled after sampling at 120 min (dash-dotted lines).

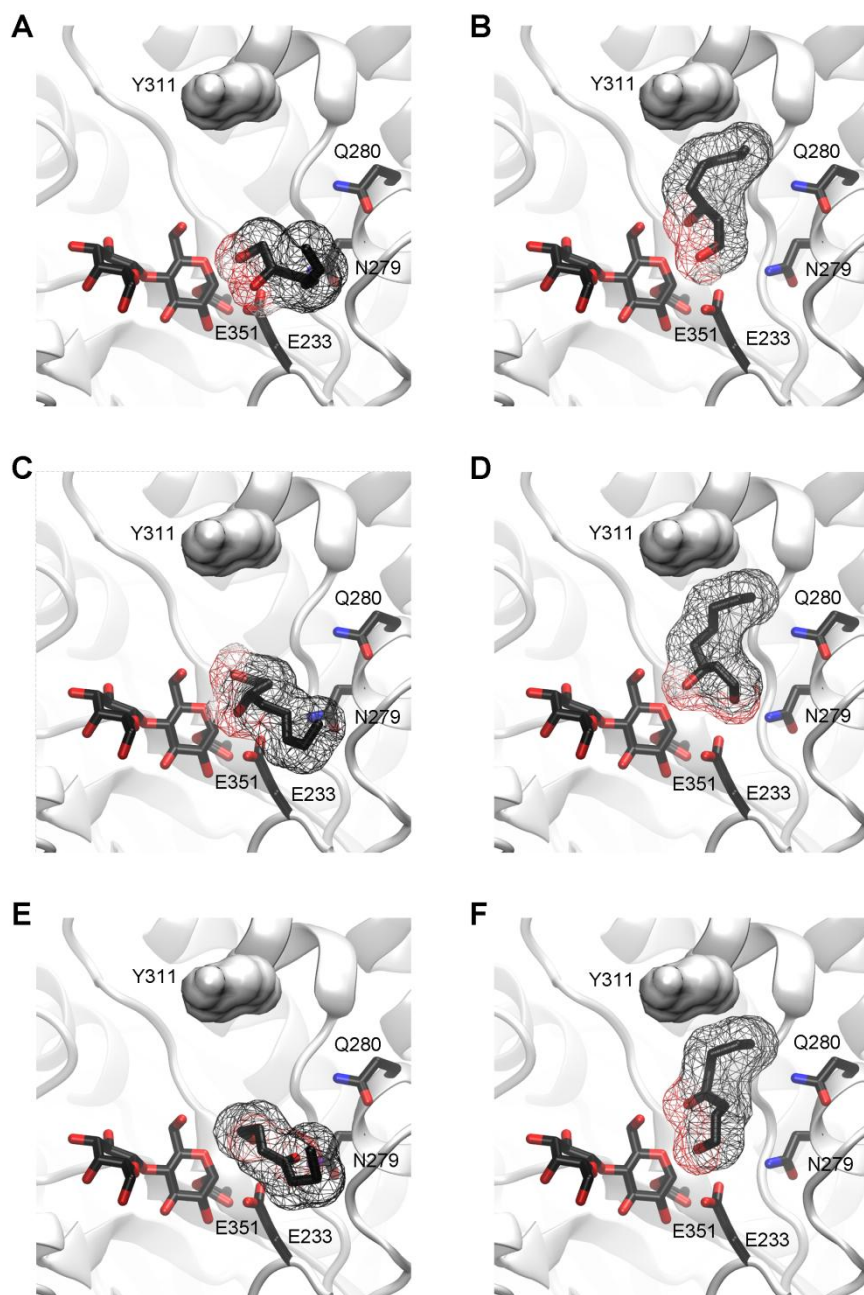
**Modeling D311Y in complex with donor substrate.** For computational docking, the 3D structure of the D311Y:CNP-β-Celb complex was modeled using available crystallographic information (PDB ID: 2OSX), procuring a “Michaelis” complex geometry competent for catalysis (Figure 6). Analysis of this complex revealed that conformational distortion ( ${}^1S_3$ ) of the D-glucopyranosyl unit occupying subsite -1 places the anomeric carbon of the scissile bond at a distance of 3.9 Å from the nucleophile (E351) and the glycosidic oxygen at 4.4 Å from the general acid/base (E233). The leaving group (CNP) is thus located in axial orientation at subsite +1 establishing stacking interactions with the enzyme (Y306 and the introduced Y311).



**Figure 6.** Model of the Michaelis complex D311Y:CNP-β-Celb. Substrate and catalytic residues are drawn using thick lines. The hydrophobic pincer formed by residues Y306, L308 and Y311 is shown in grey surface mode.

**Modeling the covalent intermediate lactosyl-D311Y bound to acceptor molecules.** To rationalize the outcomes of the transglycosylation reactions, the binding modes of acceptors **2** and **4** to the covalent lactosyl-D311Y intermediate were investigated. This enzyme state was chosen because it immediately precedes the transglycosylation reaction. Six different catalytically-competent geometries were obtained (Figure 7), all of which displayed correct orientations for plausible glycosyl transfer (i.e., short distances to the anomeric carbon and general acid/base) onto the alcohol moieties of the two acceptors. Regarding both acceptor **2** and **4**, in all scenarios the single aliphatic chains can adopt two different binding configurations that correspond to the two binding channels that are revealed in the crystallographic structure of EGC bound to the GM3 ganglioside (Figure S3A). Channel I is defined by hydrophobic sidechains of Y182, F235 and I276, whereas channel II is defined by the more hydrophilic sidechains R177, E179, L308 and D311. Logically, in the mutant D311Y the hydrophobicity of channel II is increased compared to that of the wild-type enzyme. This leads to the formation of a clear hydrophobic interaction between the tyrosine ring and the aliphatic chain of both acceptors **2** and **4** when these occupy channel II (right-hand side of Figure 7). When the aliphatic chain occupies channel I (left-hand side of Figure 7), hydrophobic interactions are established with Y182, F235 and I276, but not with Y311. In all scenarios, binding of the acceptor molecule is accompanied by H-bonding to either N279

or the *HO-6* of the D-glucopyranosyl unit at subsite -1, mimicking the interactions already described between EGC and the ceramide part of GM3 ganglioside.



**Figure 7.** Model of the covalent glycosyl-D311Y intermediate:acceptors 2 and 4 complexes. Substrates and catalytic residues are drawn using thick lines. The introduced D311Y mutation is shown in a grey surface mode. Binding modes of acceptor 2 for glycosyl transfer at primary alcohol in channel I (A) and II (B). Binding modes of acceptor 2 for glycosyl transfer at



secondary alcohol in channel I (C) and II (D). Binding modes of acceptor **4** in channel I (E) and II (F).

## DISCUSSION

*Endo*-glycoceramidase II from *Rhodococcus* sp. strain M-777 is a retaining GH that hydrolyzes complex gangliosides, which are composed of ceramide chains linked to oligosaccharides. Accordingly, this rGH has already been exploited for chemistry purposes, converting it into a glycosynthase for the synthesis of glycosphingolipids.<sup>37–40,49</sup> In the present study, the inherent properties of the active site of EGC that allow it to accommodate both sphingolipids and sugar moieties led us to consider EGC as a promising prototype for the design of enzymes for APG synthesis.

Based on previous studies aimed at improving transglycosylation in rGHs,<sup>14,16,17,50–54</sup> it is well-established that mutations that lower the hydrolytic component of the reaction and others that increase acceptor binding can lead to overall improvements in transglycosylation. Therefore, having ascertained that EGC only poorly transfers  $\beta$ -cellobiosyl donor onto alcohol acceptors, it was decided to target the aglycone subsites of this enzyme using a site-saturation mutagenesis strategy to mutate four different amino acids that are thought to interact with the aglycone moiety (E179, N279, Q280 and D311).<sup>27</sup> Like in previous studies, we deployed a two-stage screening strategy that was intended to isolate mutants displaying lowered ability to catalyze the first step of the reaction (i.e., the release of CNP from the donor CNP- $\beta$ -Celb in the presence of this one only), but increased ability to achieve this in the presence of additional target acceptors. The final outcome of our strategy was the isolation of D311Y. Though this clone displays greater ability to release CNP from the donor CNP- $\beta$ -Celb than EGC, in contradiction to what we look for, it was the only one that exhibited an even higher activity (i.e., leaving group release) in the presence of two out of the four acceptors tested.

Previous work has revealed that rGHs that display high or even exclusive transglycosylation activity are mostly characterized by low or moderate  $k_{cat}/K_M$  values. This finding has led us to propose that so-called non-Leloir transglycosylases, which are a specific type of rGHs, are intrinsically poor enzymes that display impaired ability to generate efficient transition states.<sup>14–17</sup> In the case of D311Y, the ability of the mutant to perform the first step of the double displacement mechanism is altered, but is apparently slightly improved when compared to EGC. However, although this appears to contradict expectations, we propose that EGC is an intrinsically very poor catalyst, because its performance constant  $k_{cat}/K_M$  is very low. Consistently, although D311Y displays enhanced  $k_{cat}/K_M$  value, the actual improvement

is modest and the resulting enzyme remains an intrinsically poor catalyst for the hydrolysis of CNP- $\beta$ -Celb. In fact, the modification of the T/H ratio provoked by the mutation of D311 to tyrosine is mainly due to a 3.2-fold decrease in the  $K_M$  value, which is indicative of improved donor substrate binding for transglycosylation.

Recent studies suggested that the most highly conserved residues within or around the subsite -1 are the major determinants of donor substrate binding and thus those that have most influence on the T/H ratio.<sup>14,16,17</sup> In contrast, D311 is not a determinant of subsite -1, although the glutamic acid sidechain is conserved at this position in ten out of the twelve enzymes that currently form subfamily GH5\_28 (Figure S7). Regarding the nature of the mutant selected in this study, it is well-established that aromatic residues are often key molecular determinants of subsites in GHs, since aromatic sidechains can stack with sugar rings. For instance, the substitution of a serine by a tryptophan in the +1 subsite of a GH5 *endo*- $\beta$ -1,4-mannanase conveniently illustrates how such a mutagenesis strategy can procure higher transglycosylation capability in GHs.<sup>55</sup> Additionally, many studies have described positive binding sites as determinant for the T/H partition.<sup>13,14,56-59</sup> Therefore, the fact that the substitution of D311 by a tyrosine procured higher transglycosylation capability in EGC is consistent with previous findings and undoubtedly results from the increased hydrophobicity and bulkiness in the aglycone subsite, notably +1.<sup>60-63</sup> These specific properties of tyrosine could alter the enzyme's behavior in several ways. Firstly, the Y311 and Y306 might stack sandwich-wise with the CNP moiety of the donor (Figure 6), an interaction that would explain the lowered  $K_M$  value of D311Y-catalyzed hydrolysis of CNP- $\beta$ -Celb and, inversely, is inferred by the lowered transglycosylation level of D311Y when using F- $\beta$ -Celb, a donor that lacks an aromatic aglycon moiety. Secondly, increased hydrophobicity in subsite +1 probably favors binding of longer acceptor aliphatic chains. This is clearly implied from the models of D311Y with acceptors **2** or **4** bound in channel II (right-hand side of Figure 7). Thirdly, the bulkiness of the tyrosinoyl sidechain might induce a donor/acceptor substrate interaction between the *HO-6*<sup>1</sup> of cellobiosyl moiety and the extra alcohol or ketone function of the acceptor. Moreover, D311 forms a salt bridge with R177. Loss of this feature in Y311 probably leads to a relative greater conformational flexibility within the active site; an important characteristic of glycosyltransferases and GHs working in transglycosylation-mode.<sup>15,58,64</sup> Finally, it is noteworthy that crystallographic data reveals that D311 in EGC interacts with an ordered water molecule (c.f. structure of EGC:ganglioside complex),<sup>27</sup> thus facilitating the positioning of a nucleophilic water molecule on the  $\beta$ -face of the donor. Presumably, since this interaction is lost in D311Y, water is probably less well-positioned to

perform hydrolysis.<sup>65</sup>

It is noteworthy that in this work we used a catalytically-inefficient enzyme-substrate pair that is characterized by high transition state energy barriers (i.e., low  $k_{\text{cat}}/K_{\text{M}}$  value for the hydrolysis of CNP- $\beta$ -Celb by EGC). In this respect, the dual effect of the mutation D311Y led to an overall improvement of catalysis that specifically favored transglycosylation. First, it procured a slightly higher - yet modest -  $k_{\text{cat}}/K_{\text{M}}$  value, indicative of an improved enzyme-substrate interaction and thus a lowered energy barrier associated with the glycosylation step. Second, it improved acceptor-mediated deglycosylation of the glycosyl-enzyme intermediate, thus probably lowering the energy barrier of this step.

## CONCLUSION

EGC has been successfully engineered, to create a quite efficient transglycosylase that is able to synthesize APG derivatives. Remarkably this result was achieved through a single point mutation that mostly influences interactions in the aglycone subsite, favoring first donor substrate and then acceptor substrate binding. Beyond the specific nature of our achievement, this study further underlines the fact that the switch from hydrolytic to transglycosylation activity in retaining glycoside hydrolases is brought about by fulfilling two key conditions: (i) impair the ability of the enzyme to generate efficient transition states, thus lowering its overall catalytic potency (i.e., low  $k_{\text{cat}}/K_{\text{M}}$ ), and (ii) improve the binding of target acceptor molecules in subsite +1, thus favoring acceptor-mediated deglycosylation of the glycosyl-enzyme intermediate.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the [ACS Publications website](#) at DOI: .

Sequence information and alignment, representation of active site, activity profiles (TLC and NMR), physico-chemical, kinetic and stability characteristics as well as transglycosylation performances.

## ACKNOWLEDGMENTS

This work was performed in the framework of ERA-NETBIOSURF (grant no. 0315928A, ERA-IB10.039). NMR analyses were performed using facilities at MetaToul (Metabolomics & Fluxomics Facilities, Toulouse, France, [www.metatoul.fr](http://www.metatoul.fr)), which is part of the national

infrastructure MetaboHUB (The French National infrastructure for metabolomics and fluxomics, [www.metabohub.fr](http://www.metabohub.fr)), and is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, SICOVAL, IBSa-France, CNRS and INRA. Work requiring automated liquid handling was performed using the ICEO facility that is part of the PICT infrastructure and is financially supported by IBSa-France, CNRS and INRA. Contribution from R.F. was partially supported by the Région Midi-Pyrénées grant DESR-Recherche 14052246 (CTP-B) and the research mobility grants from INSA Toulouse (2015). A.P acknowledges grant BIO2013-49022-C2-1-R from MINECO, Spain. The authors thank Prof. S. Cottaz and Dr S. Fort (CERMAV-CNRS) for the gift of  $\beta$ -cellobiosyl fluoride heptaacetate, and Prof. S. Withers for insightful discussions.

## ABBREVIATIONS

APG, alkylpolyglycoside; CNP, 2-chloro-4-nitrophenol; CNP- $\beta$ -Celb, 2-chloro-4-nitrophenyl  $\beta$ -D-Glcp-(1 $\rightarrow$ 4)-D-Glcp (i.e. CNP- $\beta$ -cellobioside); F- $\beta$ -Celb,  $\beta$ -cellobiosyl fluoride; EGC, *endo*-glycoceramidase II from *Rhodococcus* sp. M-777 strain; GH, glycoside hydrolase; T/H, transglycosylation/hydrolysis partition; LG, leaving group; Ph, phenol; rGH, retaining GH, TG, non-Leloir transglycosylase.

## REFERENCES

- (1) von Rybinski, W.; Hill, K. *Angew. Chem. Int. Ed.* **1998**, 37, 1328–1345.
- (2) Greffe, L.; Bessueille, L.; Bulone, V.; Brumer, H. *Glycobiology* **2004**, 15, 437–445.
- (3) Heyd, M.; Kohnert, A.; Tan, T.-H.; Nusser, M.; Kirschhöfer, F.; Brenner-Weiss, G.; Franzreb, M.; Berensmeier, S. *Anal. Bioanal. Chem.* **2008**, 391, 1579–1590.
- (4) Reis, R. S.; Pacheco, G. J.; Pereira, A. G.; Freire, D. M. G. Biosurfactants: production and applications. In *Biodegradation - Life of Science*; Chamy, R.; Rosenkranz, F. Eds.; InTech, Croatia, 2013.
- (5) Gudiña, E. J.; Rangarajan, V.; Sen, R.; Rodrigues, L. R. *Trends Pharmacol. Sci.* **2013**, 34, 667–675.
- (6) Löffler, H.; Happle, R. *Contact Dermatitis* **2003**, 48, 26–32.
- (7) Schwarz, J. C.; Klang, V.; Hoppel, M.; Mahrhauser, D.; Valenta, C. *Eur. J. Pharm. Biopharm.* **2012**, 81, 557–562.
- (8) Rather, M.; Mishra, S. *Sustain. Chem. Process.* **2013**, 1, 7.
- (9) Wang, L.-X.; Huang, W. *Curr. Opin. Chem. Biol.* **2009**, 13, 592–600.
- (10) Crout, D.; Vic, G. *Curr. Opin. Chem. Biol.* **1998**, 2, 98–111.
- (11) van Rantwijk, F.; Woudenberg-van Oosterom, M.; Sheldon, R. A. *J. Mol. Catal. B Enzym.* **1999**, 6, 511–532.
- (12) Lundemo, P.; Karlsson, E. N.; Adlercreutz, P. *Appl. Microbiol. Biotechnol.* **2016**, doi:10.1007/s00253-016-7833-9 (accessed Sep 27, 2016)
- (13) Talens-Perales, D.; Polaina, J.; Marín-Navarro, J. *J. Agric. Food Chem.* **2016**, 64, 2917–2924.

- (14) Bissaro, B.; Durand, J.; Biarnés, X.; Planas, A.; Monsan, P.; O'Donohue, M. J.; Fauré, R. *ACS Catal.* **2015**, 5, 4598–4611.
- (15) Bissaro, B.; Monsan, P.; Fauré, R.; O'Donohue, M. J. *Biochem. J.* **2015**, 467, 17–35.
- (16) Teze, D.; Hendrickx, J.; Czjzek, M.; Ropartz, D.; Sanejouand, Y.-H.; Tran, V.; Tellier, C.; Dion, M. *Protein Eng. Des. Sel.* **2014**, 27, 13–19.
- (17) Teze, D.; Daligault, F.; Ferrieres, V.; Sanejouand, Y.-H.; Tellier, C. *Glycobiology* **2015**, 25, 420–427.
- (18) Danby, P. M.; Withers, S. G. *ACS Chem. Biol.* **2016**, 11, 1784–1794.
- (19) Lundemo, P.; Adlercreutz, P.; Karlsson, E. N. *Appl. Environ. Microbiol.* **2013**, 79, 3400–3405.
- (20) Rémond, C.; Ferchichi, M.; Aubry, N.; Plantier-Royon, R.; Portella, C.; O'Donohue, M. J. *Tetrahedron Lett.* **2002**, 43, 9653–9655.
- (21) Ochs, M.; Muzard, M.; Plantier-Royon, R.; Estrine, B.; Rémond, C. *Green Chem.* **2011**, 13, 2380–2388.
- (22) Ito, M.; Yamagata, T. *J. Biol. Chem.* **1986**, 261, 14278–14282.
- (23) Ito, M.; Tatsuya, Y. *J. Biol. Chem.* **1989**, 264, 9510–9519.
- (24) Izu, H.; Izumi, Y.; Kurome, Y.; Sano, M.; Kondo, A.; Kato, I.; Ito, M. *J. Biol. Chem.* **1997**, 272, 19846–19850.
- (25) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Res.* **2014**, 42, D490–D495. Carbohydrate Active Enzymes database ([www.cazy.org](http://www.cazy.org))
- (26) Aspeborg, H.; Coutinho, P. M.; Wang, Y.; Brumer, H.; Henrissat, B. *BMC Evol. Biol.* **2012**, 12, 186.
- (27) Caines, M. E. C.; Vaughan, M. D.; Tarling, C. A.; Hancock, S. M.; Warren, R. A. J.; Withers, S. G.; Strynadka, N. C. J. *J. Biol. Chem.* **2007**, 282, 14300–14308.
- (28) Li, Y. T.; Carter, B. Z.; Rao, B. N.; Schweingruber, H.; Li, S.-C. *J. Biol. Chem.* **1991**, 266, 10723–10726.
- (29) Ashida, H.; Tsuji, Y.; Yamamoto, K.; Kumagai, H.; Tochikura, T. *Arch. Biochem. Biophys.* **1993**, 305, 559–562.
- (30) Horibata, Y.; Higashi, H.; Ito, M. *J. Biochem.* **2001**, 130, 263–268.
- (31) Ishibashi, Y.; Kiyohara, M.; Okino, N.; Ito, M. *J. Biochem.* **2007**, 142, 239–246.
- (32) Ishibashi, Y.; Nakasone, T.; Kiyohara, M.; Horibata, Y.; Sakaguchi, K.; Hijikata, A.; Ichinose, S.; Omori, A.; Yasui, Y.; Imamura, A.; Ishida, H.; Kiso, M.; Okino, N.; Ito, M. *J. Biol. Chem.* **2007**, 282, 11386–11396.
- (33) Ishibashi, Y.; Nagamatsu, Y.; Meyer, S.; Imamura, A.; Ishida, H.; Kiso, M.; Okino, N.; Geyer, R.; Ito, M. *Glycobiology* **2009**, 19, 797–807.
- (34) Moracci, M.; Trincone, A.; Perugino, G.; Ciaramella, M.; Rossi, M. *Biochemistry* **1998**, 37, 17262–17270.
- (35) Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, 120, 5583–5584.
- (36) Malet, C.; Planas, A. *FEBS Lett.* **1998**, 440, 208–212.
- (37) Vaughan, M. D.; Johnson, K.; DeFrees, S.; Tang, X.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **2006**, 128, 6300–6301.
- (38) Hancock, S. M.; Rich, J. R.; Caines, M. E. C.; Strynadka, N. C. J.; Withers, S. G. *Nat. Chem. Biol.* **2009**, 5, 508–514.
- (39) Rich, J. R.; Cunningham, A.-M.; Gilbert, M.; Withers, S. G. *Chem. Commun.* **2011**, 47, 10806–10808.
- (40) Rich, J. R.; Withers, S. G. *Angew. Chem. Int. Ed.* **2012**, 51, 8640–8643.
- (41) Stocker, B. L.; Timmer, M. S. M. *ChemBioChem* **2013**, 14, 1164–1184.

- (42) Wennekes, T.; van den Berg, R. J. B. H. N.; Boot, R. G.; van der Marel, G. A.; Overkleeft, H. S.; Aerts, J. M. F. G. *Angew. Chem. Int. Ed.* **2009**, 48, 8848–8869.
- (43) Johnson, K. F.; Defrees, S.; Withers, S.; Vaughan, M. Mutant endoglycoceramidases with enhanced synthetic activity. Patent US20140147891 A1, May 29, 2014.
- (44) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, 243, 779–815.
- (45) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, 31, 455–461.
- (46) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, 14, 33–38.
- (47) Zhang, X.-F.; Yang, G.-Y.; Zhang, Y.; Xie, Y.; Withers, S. G.; Feng, Y. *Sci. Rep.* **2016**, 6, 33797.
- (48) Konstantinidis, A. K.; Marsden, I.; Sinnott, M. L. *Biochem. J.* **1993**, 291, 883–888.
- (49) Yang, G.-Y.; Li, C.; Fischer, M.; Cairo, C. W.; Feng, Y.; Withers, S. G. *Angew. Chem. Int. Ed.* **2015**, 54, 5389–5393.
- (50) Arab-Jaziri, F.; Bissaro, B.; Dion, M.; Saurel, O.; Harrison, D.; Ferreira, F.; Milon, A.; Tellier, C.; Fauré, R.; O'Donohue, M. J. *New Biotechnol.* **2013**, 30, 536–544.
- (51) Arab-Jaziri, F.; Bissaro, B.; Tellier, C.; Dion, M.; Fauré, R.; O'Donohue, M. J. *Carbohydr. Res.* **2015**, 401, 64–72.
- (52) Feng, H.-Y.; Drone, J.; Hoffmann, L.; Tran, V.; Tellier, C.; Rabiller, C.; Dion, M. *J. Biol. Chem.* **2005**, 280, 37088–37097.
- (53) Kone, F. M. T.; Le Behec, M.; Sine, J.-P.; Dion, M.; Tellier, C. *Protein Eng. Des. Sel.* **2009**, 22, 37–44.
- (54) Osanjo, G.; Dion, M.; Drone, J.; Solleux, C.; Tran, V.; Rabiller, C.; Tellier, C. *Biochemistry* **2007**, 46, 1022–1033.
- (55) Dilokpimol, A.; Nakai, H.; Gotfredsen, C. H.; Baumann, M. J.; Nakai, N.; Abou Hachem, M.; Svensson, B. *BBA-Proteins Proteom.* **2011**, 1814, 1720–1729.
- (56) Armand, S.; Andrews, S. R.; Charnock, S. J.; Gilbert, H. J. *Biochemistry* **2001**, 40, 7404–7409.
- (57) van der Veen, B. A.; Leemhuis, H.; Kralj, K.; Uitdehaag, J. C. M.; Dijkstra, B. W.; Dijkhuizen, L. *J. Biol. Chem.* **2001**, 276, 44557–44562.
- (58) Mark, P.; Baumann, M. J.; Eklöf, J. M.; Gullfot, F.; Michel, G.; Kallas, Å. M.; Teeri, T. T.; Brumer, H.; Czjzek, M. *Proteins Struct. Funct. Bioinf.* **2009**, 75, 820–836.
- (59) Zakariassen, H.; Hansen, M. C.; Jøranli, M.; Eijsink, V. G. H.; Sørli, M. *Biochemistry* **2011**, 50, 5693–5703.
- (60) Rivera, M. H.; López-Munguía, A.; Soberón, X.; Saab-Rincón, G. *Protein Eng. Des. Sel.* **2003**, 16, 505–514.
- (61) Rosengren, A.; Hägglund, P.; Anderson, L.; Pavon-Orozco, P.; Peterson-Wulff, R.; Nerinckx, W.; Ståhlbrand, H. *Biocatal. Biotransfor.* **2012**, 30, 338–352.
- (62) Rosengren, A.; Reddy, S. K.; Sjöberg, J. S.; Aurelius, O.; Logan, D. T.; Kolenová, K.; Ståhlbrand, H. *Appl. Microbiol. Biotechnol.* **2014**, 98, 10091–10104.
- (63) Ochs, M.; Belloy, N.; Dauchez, M.; Muzard, M.; Plantier-Royon, R.; Rémond, C. *J. Mol. Catal. B: Enzym.* **2013**, 96, 21–26.
- (64) Rojas-Cervellera, V.; Ardèvol, A.; Boero, M.; Planas, A.; Rovira, C. *Chem. - Eur. J.* **2013**, 19, 14018–14023.
- (65) Hidaka, M.; Fushinobu, S.; Honda, Y.; Wakagi, T.; Shoun, H.; Kitaoka, M. *J. Biochem.* **2010**, 147, 237–244.