Rational Enzyme Design Without Structural Knowledge: A Sequence-Based Approach for Efficient Generation of Glycosylation Catalysts

David Teze, Jiao Zhao, Mathias Wiemann, Kazi Zubaida Gulshan Ara, Rossana Lupo, Mette Errebo Rønne, Göran Carlström, Jens Duus, Yves-Henri Sanejouand, Michael O'Donohue, et al.

HAL Id: hal-02906907
https://hal.inrae.fr/hal-02906907v1

Preprint submitted on 26 Jul 2020 (v1), last revised 25 Jun 2021 (v2)

Rational enzyme design without structural knowledge: a sequence-based approach for efficient generation of glycosylation catalysts

David Teze[1,*], Jiao Zhao[2], Mathias Wiemann[3], Kazi Zubaida Gulshan Ara[4], Rossana Lupo[1], Mette Errebo Rønne[1], Göran Carlström[5], Jens Duus[6], Yves-Henri Sanejouand[7], Michael J. O'Donohue[2], Eva Nordberg-Karlsson[4], Régis Fauré[2], Henrik Stålbrand[3] and Birte Svensson[1,*]

Glycobiology is dogged by the relative scarcity of synthetic, defined oligosaccharides. Enzyme-catalysed glycosylation using glycoside hydrolases is feasible, but is hampered by the innate hydrolytic activity of these enzymes. Protein engineering methods are applicable, though usually require prior structural knowledge of the target enzyme and the use of powerful computing methods, and/or relies on extensive screening methodologies. Here we describe a straightforward strategy that involves rapid *in silico* analysis of protein sequences. The method pinpoints a small number (<10) of mutant candidates aimed at diminishing hydrolysis and thus tipping the reaction balance toward transglycosylation. Requiring no other significant prior knowledge of the target enzyme than its sequence, the results reveal that the method is quite generic, allowing the improvement of glycoside hydrolases that act on different α-/β-pyranosides or furanosides. Moreover, the presented data support that mutational hotspots that are validated in one enzyme can be transposed to other related enzymes.

Glycosides are ubiquitous and abundant in Nature, being essential for biological interactions and processes. Nevertheless, progress in glycobiology is hampered by the lack of synthetic carbohydrates, an issue related to their complexity. Carbohydrates are composed of polyhydroxylated units that exist in different forms (e.g. pyranoside or furanoside), interlinked in a variety of regioselectivities, with the anomeric centres displaying either α- or β-anomeric configurations[1]. Faced with this high degree of complexity organic chemistry has developed numerous glycosylation methodologies[2,3]. These involve several synthetic steps, including protection-deprotection cycles, are characterised by relatively poor overall yields, and generate considerable amounts of waste. This is in stark contrast to polynucleotides and polypeptides, both of which are accessible *via* automated chemical synthesis processes and through *in vivo* biological synthesis. Unfortunately, unlike these biopolymers, most carbohydrates cannot be obtained using straightforward, generic technologies[4] amenable to automation.

Enzyme-catalysed glycosylation offers an alternative to chemical methods. The natural choice for this are glycosyltransferases (GTs) that are well-represented in a variety of families in the CAZy database (http://www.cazy.org/)[5], with each family potentially harbouring numerous specificities[6]. Nevertheless, GTs have proven to be rather difficult to handle *in vitro* and often require expensive nucleotide-glycoside donors[7]. Therefore, glycoside

[1]Department of Biotechnology and Biomedicine. Technical University of Denmark, Kongens Lyngby, Denmark. [2]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France. [3]Department of Biochemistry and Structural Biology. Lund University, Lund, Sweden. [4]Biotechnology, Department of Chemistry. Lund University, Lund, Sweden. [5]Centre for Analysis and Synthesis, Department of Chemistry. Lund university, Lund, Sweden [6]Department of Chemistry. Technical University of Denmark, Kongens Lyngby, Denmark. [7]Université de Nantes, CNRS, UFIP, France. Correspondence should be addressed to D.T. (datez@dtu.dk) or B.S. (bis@bio.dtu.dk).

hydrolases (GHs) offer an alternative for glycosynthesis. These enzymes are particularly abundant and their diversity in terms of bond breaking ability matches carbohydrate complexity itself. GHs catalyse hydrolysis using a variety of mechanisms, but the majority perform two subsequent displacements at the anomeric carbon, thus yielding a product whose anomeric configuration is identical to that of the substrate[8,9]. GHs operating by such a mechanism are termed "retaining" GHs and represent 67% of all GHs (755496 classified modules) in the CAZy database, grouped in 83 out of its current 161 GH families (as of February 24th, 2020). Allowing for rare exceptions[10], all other GHs are inverting. A result of the double displacement mechanism is that retaining GHs possess the intrinsic potential to catalyse transglycosylation, thus to synthesize glycosidic bonds. Even though most retaining GHs have strong hydrolytic activity and weak, often undetectable transglycosylation activity, some display significant levels of transglycosylation. This reaction is under kinetic control and modulated by a number of factors related to the reaction conditions[11], including acceptor concentration, water activity, substrate activation, temperature, pH, but particularly enzyme properties. Accordingly, engineering of GHs has proven to be a potent way to obtain transglycosylases[12].

The most generic GH engineering approach described to date is the so-called glycosynthase strategy[13–16]. This involves the creation of a crippled enzyme, in which a catalytic carboxylate is replaced by a catalytically impotent moiety. The resulting mutant enzyme is fed with a strongly activated substrate that mimics the reaction intermediate (e.g. an α-glycosyl fluoride for a β-active wild-type (WT) enzyme), and turns the enzyme into an inverting glycosynthase. Albeit powerful, this method relies on the availability of a suitably reactive, but sufficiently stable substrate. Moreover, glycosynthases are intrinsically impotent biocatalysts that display extremely low activity, thus requiring large quantities of enzyme[13–17].

An alternative strategy to convert GHs into efficient transglycosylases is to increase the transglycosylation/hydrolysis (T/H) ratio while conserving the retaining mechanism. To achieve this, a considerable number of studies have employed rational or random protein engineering methodologies. However, such approaches require either in-depth structural and biochemical knowledge and, in some cases, use of sophisticated computational methods (rational design); or the creation of large libraries, introducing location-agnostic modifications, coupled with powerful phenotypic screens (random mutagenesis)[18]. Nevertheless, by simply targeting a small number of conserved active-site residues in several retaining GHs[19–21], we previously demonstrated that transglycosylation capability can be improved without extensive screening using structural information coupled with sequence conservation analysis. Moreover, others have transposed our obtained beneficial mutations to related enzymes, leading to significant transglycosylation yields[22].

Herein we demonstrate how a refined sequenced-based approach can be used to improve the T/H ratio (i.e. enhance transglycosylation capability) in a variety of predominantly hydrolytic GHs. This strategy involves replacing an enzyme's most conserved residues by structural analogues (e.g., Tyr into Phe, Asp into Asn). The approach does not require structural or mechanistic knowledge of the GH. However, for enzymes where catalytic residues and residues distant (>10 Å) from the −1 subsite[23] are known or can be predicted, those are not mutated. Moreover, any conserved glycines and prolines are excluded[20]. Using this strategy, we demonstrate successful application to GHs from families 2, 10, 20, 29 and 51. This approach is i) fast, requiring the generation and analysis of ≤ 10 variants per enzyme; ii) generic, being applicable to a variety of glycosidic bond-forming reactions; iii) procures highly efficient transglycosylases; and iv) allows for mutation transfer.

Results

Multiple sequence alignment and residue conservation. The method described herein relies on identifying conserved residues. While the concept of amino acid conservation is central to molecular evolution theory, to date there is no agreed, precise definition of what it signifies. To circumvent this obstacle, we have devised a methodology that pinpoints residues that are significantly more conserved compared to others within a given sequence. For this approach to be successful, a ranking method is required that is robust even in extreme cases where the target sequences are either highly diverse, meaning that the overall stringency of conservation is low, or alternatively where they are highly homogenous. The three-step methodology begins with the collection of a large number of sequences, which are then clustered to reduce redundancy[24,25]. After clustering, iterative multiple sequence alignments (MSA) are performed to select sequences that share at least 5% sequence identity. This minimum of sequence identity (termed "ID% threshold") is then iteratively increased, while at each step the residues are ranked by decreasing conservation. Their conservation is then plotted against the logarithm of their rank (Fig. 1). When a significant conservation drop is observed after rank 10–15, the sequences are kept and the ID% threshold is no longer increased. The corresponding MSA is analysed to identify the residues hereafter designated as "conserved". Iterative MSA and conservation analysis are performed using ClustalΩ[26] and an in-house script.
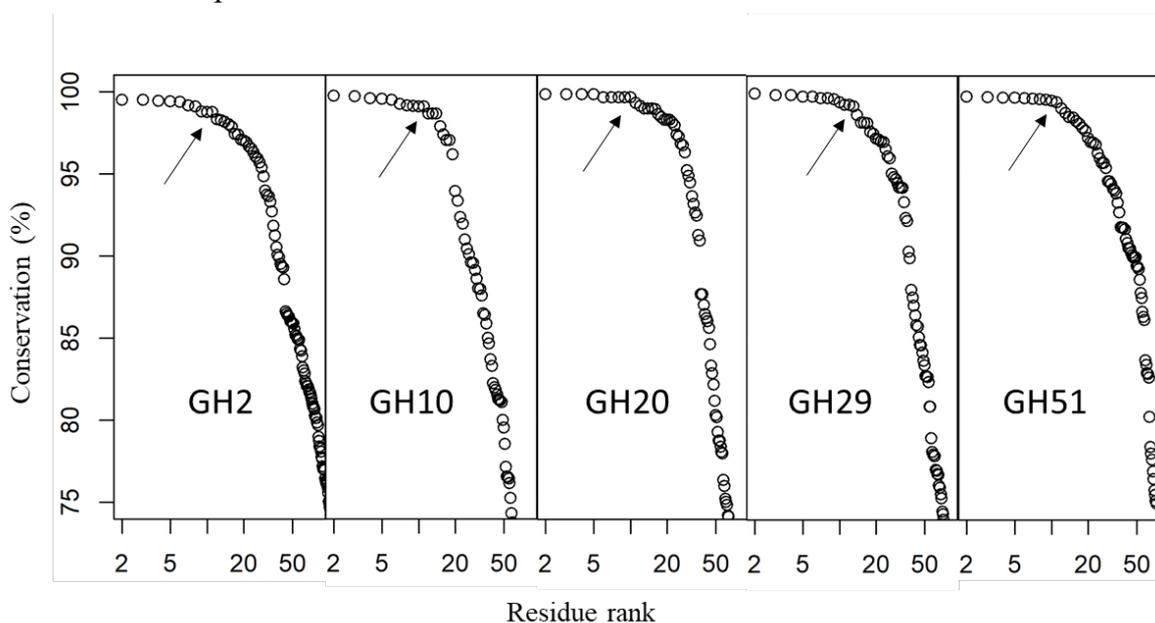
Fig. 1 | Residue conservation analysis. Residues in the MSA are ranked according to decreasing conservation, and their conservation is plotted as function of the logarithm of their rank. From left to right, results obtained from the analysis of the final MSA on GH families 2, 10, 20, 29 and 51 respectively. The arrows indicate the last rank considered as "significantly more conserved" than the rest of the residues in one sequence.

To test the methodology, GHs from families 2, 10, 20, 29 and 51 were selected. These cover a wide range of glycosidic bond features, each representing a distinctive synthetic challenge. The outcomes of the sequence analysis are summarized in Table 1. Comparing GH10 and GH51 shows that similar sequence pool sizes and identical stringency criteria generate different heterogeneity levels (21 vs 35% average identity within sequences kept – termed "mean ID%"), while the cases of GH2 and GH20 reveal that similar heterogeneity can be obtained from a variable number of selected sequences (9301 vs 585 sequences retained from 20 000 retrieved). Nevertheless, in all cases the methodology yields 6–12 candidates for experimental evaluation.

3

GH family	Enzyme target	Species	Retrieved sequences	ID%[a] threshold	Sequences kept	Mean ID%[a,b]	Conserved residues	Candidates evaluated[c]
GH2	*Cf*Man2A	*C. fimi*	20 000	12	9301	22.6	14	12
GH10	Xyn10A	*R. marinus*	9 288	15	4431	21.1	11	9
GH20	*Bb*HI	*B. bifidum*	20 000	10	585	22.5	9	6
GH29	AlfB	*L. casei*	15 328	20	1194	24.0	13	7
GH51	*Tx*Abf	*T. xylanilyticus*	11 553	15	3272	34.6	11	8

125 **Table. 1 | MSA and residue conservation analysis.** [a]ID%: percentage of identity between sequences. [b]Aver-
126 age identity with the reference sequence. [c]Not evaluated Gly, Pro and identified catalytic residues.

127

128 **GH2, β-mannosidase *Cf*Man2A.** Compared to the synthesis of α-D-mannosidic (or to β-
129 D-glucosidic and β-D-galactosidic) bonds, the synthesis of β-D-mannosyl-containing com-
130 pounds is complicated by the axial 2-OH of D-mannosyl moieties. In the case of β-manno-
131 sidase-mediated synthesis, this constraint leads to a different conformational itinerary ($^1S_5$
132 $\rightarrow [B_{2,5}]^{\ddagger} \rightarrow {}^OS_2$) compared to that adopted by most other β-pyranosidases-catalysed reac-
133 tions ($^1S_3 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^4C_1$)[27]. Overall, the challenge of synthesizing β-mannosides[28] has led
134 to interest in enzymatic synthesis[29–31] and several GH2 mannosidases have been success-
135 fully tested[32–34]. Furthermore, a glycosynthase variant of the *Cellulomonas fimi* GH2 β-
136 mannosidase (*Cf*Man2A) proved to be a proficient synthetic catalyst when fed with donor
137 α-D-mannosyl fluoride and various acceptors[29]. Herein, we investigate whether *Cf*Man2A-
138 mediated transmannosylation can be enhanced without using the radical glycosynthase
139 strategy.
140 The transglycosylation ability of wild-type (WT) *Cf*Man2A and variants (collectively
141 "*Cf*Man2A forms") was evaluated with 5 mM *p*-nitrophenyl-β-D-mannopyranoside (β-D-
142 Man*p*O*p*NP) acting as both donor and acceptor substrate. Hence in the transglycosylation
143 reaction a mannosyl unit was transferred from a donor to an acceptor molecule, elongating
144 β-D-Man*p*O*p*NP with formation of β-D-Man*p*-(1→3)-β-D-Man*p*O*p*NP and β-D-Man*p*-
145 (1→4)-β-D-Man*p*O*p*NP (both referred to as *p*NP-Man$_2$) from 5 mM β-D-Man*p*O*p*NP as
146 monitored using HPLC (Figs. 2 and S1). Nuclear magnetic resonance spectroscopy (NMR)
147 and mass-spectrometry (MS) (Fig. S1, Tables S1 and S2) were used to determine the chem-
148 ical structure of the products. Four variants (R360K, N428T, S474A and W612H) showed
149 markedly improved yields for *p*NP-Man$_2$ synthesis (32–43%) compared to *Cf*Man2A-WT
150 (18%), while W169H, D170N, D386N and H496F were discarded due to low activity (<5
151 nkat·mg^{-1}).

152

CfMan2A variant	Yield (%)	Enzyme (μg·mL⁻¹, nM)
WT	18	9.8 (104)
R360K	37	30.9 (330)
W362H	18	4.7 (50)
N428T	43	34.3 (360)
S474A	32	33.2 (350)
H496Q	15	5.3 (56)
W612H	37	23.9 (253)
Q613N	16	15.3 (162)
W623H	17	2.8 (30)
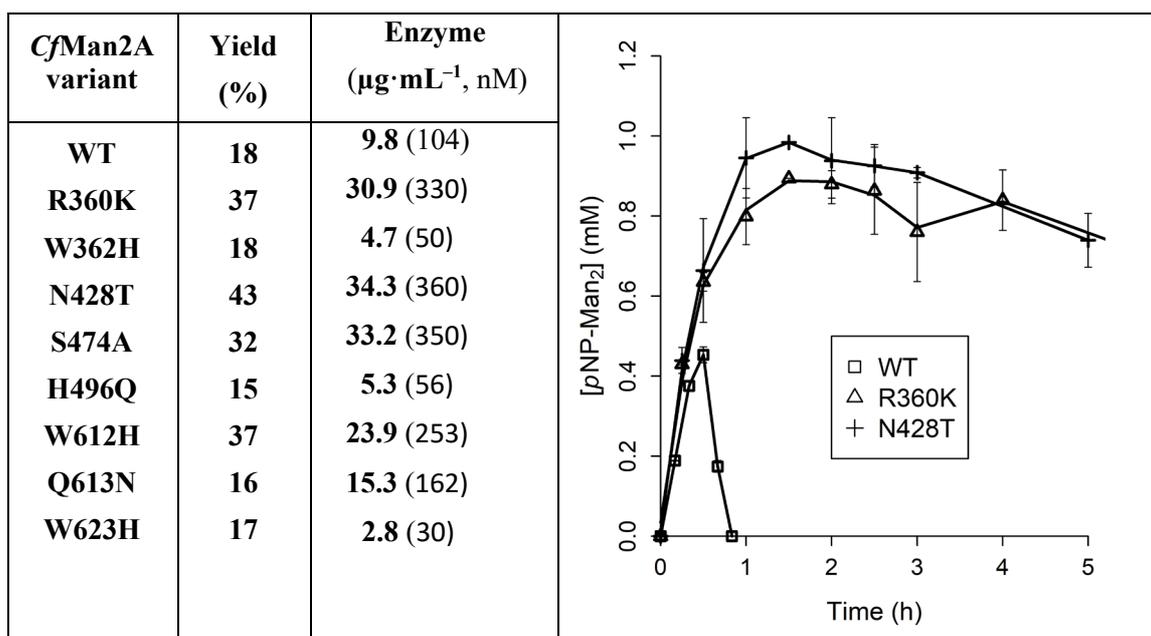
Fig. 2 | GH2 engineering. Left, maximum yields of *p*NP-Man$_2$ synthetized by *Cf*Man2A forms. Protein concentrations used to obtain the yields within 5 h are indicated in μg·mL⁻¹ and nM (in brackets). **Right,** HPLC monitoring of the formation of *p*NP-Man$_2$ by *Cf*Man2A-WT and two of its best mutants.

It is noteworthy that the best four *Cf*Man2A variants described herein were catalytically-active and thus could be applied at loadings that were only moderately greater (2 to 3.5-fold, Fig. 2) than that used for *Cf*Man2A-WT, and hundred-fold lower than the ones required with the *Cf*Man2A-E519S glycosynthase[29]. Moreover, compared to WT enzyme, monitoring revealed that the variants catalysed less secondary product hydrolysis and thus ensured high product concentrations even over prolonged reaction times (Fig. 2). This result is significant because secondary hydrolysis often impedes enzyme-mediated synthesis of β-D-mannoside conjugates[35,36].

GH10, β-*endo*-xylanase *Rm*Xyn10A_CM. Compared with most hexopyranoses, the pentopyranose configuration of the D-xylosyl unit confers enhanced flexibility and allows two distinct conformational itineraries that are compatible with GH hydrolysis: $^{1}S_3 \rightarrow [^{4}H_3]^{\ddagger} \rightarrow {}^{4}C_1$ and $^{2}S_O \rightarrow [^{2,5}B]^{\ddagger} \rightarrow {}^{5}S_1$ [27,37]. Mechanistically interesting as it may be, the real challenge posed by this enzyme is related to its *endo*-activity. This implies the use of a substrate that does not possess an exceptionally good leaving group, and thus does not provide kinetic control of the reaction, and also the necessity to monitor multiple products in order to assess transglycosylation efficiency. This is because the reaction allows the transfer of oligosaccharide moieties onto an acceptor in a single catalytic step and provides the means to synthesize higher oligosaccharides displaying degrees of polymerisation (DP) > 10. *Endo*-1,4-β-xylanase is the predominant enzyme class in the GH10 family and the subject of numerous studies[38–41]. Both negative and positive subsites[23] in GH10 xylanases have been probed to understand their influence on substrate binding and catalysis[42–45], including *Rm*Xyn10A–CM, the catalytic module of the *endo*-xylanase from *Rhodothermus marinus* used here[46]. However, very few reports concerned the transglycosylation reaction, these being limited to mutational studies of the aglycone subsites in a couple of GH10 xylanases[42,45].

To characterise transglycosylation activity, xylotetraose (X$_4$) was used as both donor and acceptor substrate. All *Rm*Xyn10A_CM forms were able to synthesise xylo-oligosaccharides (XOS) larger than X$_4$, predominantly accumulating X$_8$ during the 4 h time frame of the experiment (Figs. 3 and S3). Notably, mutants H69N, N118T and W284H synthesized

5

186 XOS with higher DP during the disproportionation of 20 mM X_4 (Fig. S3) than the WT
187 enzyme.

RmXyn10A–CM variant	Maximal X_8 area (nC·min)	Enzyme ($\mu g \cdot mL^{-1}$, nM)
WT	4.9	10.2 (21)
H69F	8.4	43.6 (90)
H69N	11.6	37.7 (78)
W73H	8.3	41.3 (86)
N118T	13.4	39.5 (82)
N169T	5.5	10.5 (22)
H204F	8.9	35.6 (74)
H204N	6.9	36.3 (75)
D234N	6.6	39.1 (81)
W284H	10.7	38.5 (80)
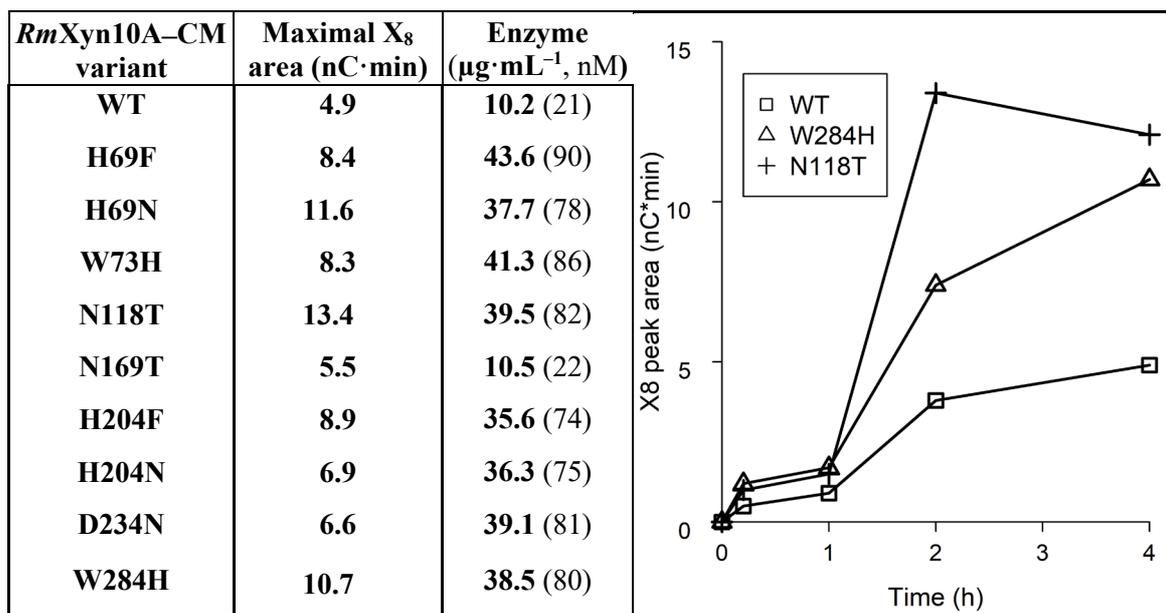
Fig. 3 | GH10 engineering. Left, Xylooctaose (X_8) synthesised by RmXyn10A–CM forms. Protein concentrations used to obtain transglycosylation within 4 h are indicated in $\mu g \cdot mL^{-1}$ and nM (in brackets). **Right,** monitoring of X_8 synthesis by RmXyn10A–CM-WT and two of its best mutants using HPAEC-PAD. Residue numbering refers to the catalytic module.

188 Further analysis revealed that using X_4 as substrate, RmXyn10A–CM variants synthesized
189 XOS of DP 6−11 (Figs. S2 and S3, Table S3). Importantly, compared to the WT enzyme,
190 all variants display greater ability to form XOS exhibiting DP ≥ 8 (Table S3), these being
191 the result of multiple transglycosylation events. The findings reveal that the conserved se-
192 quence approach led to RmXyn10A variants with significantly enhanced transglycosylation
193 abilities compared to the WT enzyme, primarily due to a drastic decrease of hydrolytic ac-
194 tivity (Fig. S4). Available structural data[46] reveal that the three most effective mutational
195 targets (H69, N118 and W284) are all closely interacting with the xylose moiety in the −1
196 subsite (Fig. S5).

197
198 **GH20, β-hexosaminidase BbHI.** During catalysis, most retaining GHs generate a glyco-
199 syl-enzyme intermediate whose anomeric configuration is opposite to that of both the sub-
200 strate (donor) and the products. However, GH families 18, 20, 25, 56, 84, 85 and 123 use a
201 substrate-assisted mechanism, in which an equatorial N-acetyl in position C-2 of the donor
202 acts as the nucleophile[47]. This leads to the formation of a non-covalent oxazoline- or oxa-
203 zolinium ion-enzyme intermediate[48] (oxazolinium ion in the case of GH20[49]). Importantly,
204 this type of mechanism is predominant in GH-catalyzed reactions involving β-D-GlcNAc
205 or β-D-GalNAc, sugars that are highly prevalent in biological systems. Hence, to assess if
206 the methodology presented here is also relevant when a glycosyl-enzyme intermediate is
207 not formed, a GH20 enzyme was targeted.

208 Previously, GH20 hexosaminidases have been extensively used for transglycosyla-
209 tion[50,51]. Therefore, herein we focused on BbHI from *Bifidobacterium bifidum,* sp. *infantis*[52].
210 This enzyme has the ability to catalyse the synthesis of LNT2 (β-D-Glc$p$NAc-(1→3)-β-D-
211 Gal$p$-(1→4)-D-Glc, also known as lacto-N-triose II), a widely sought-after glycomotif that
212 is present in human milk oligosaccharides[7,53]. Moreover, BbHI-D746E – D746 being in-
213 volved in the oxazolinium ion stabilization – has been recently reported as the first glyco-
214 synthase with high enzymatic activity, thus requiring only moderate enzyme loadings[54]. Six

215 variants and *Bb*HI-WT were analysed for their ability to synthesize LNT2 from 10 mM β-
216 D-Glc*p*NAcO*p*NP and 40 mM lactose (Fig. 4). BbHI-H603F displayed the highest T/H ratio
217 (data not schown), but also a drastically reduced activity. With the highest enzyme loading
218 (15 µM), the reaction was incomplete after 12 h, thus further characterization of
219 *Bb*HI-H603F was abandoned.

*Bb*HI variant	Yield (%)	Enzyme (µg·mL⁻¹, µM)	
WT	16	**93** (0.53)	
R577K	36	**229** (1.3)	
H603F	–	–	
D606N	9	**229** (1.3)	
W801H	17	**810** (4.6)	
W882H	66	**862** (4.9)	
D884N	12	**582** (3.3)	
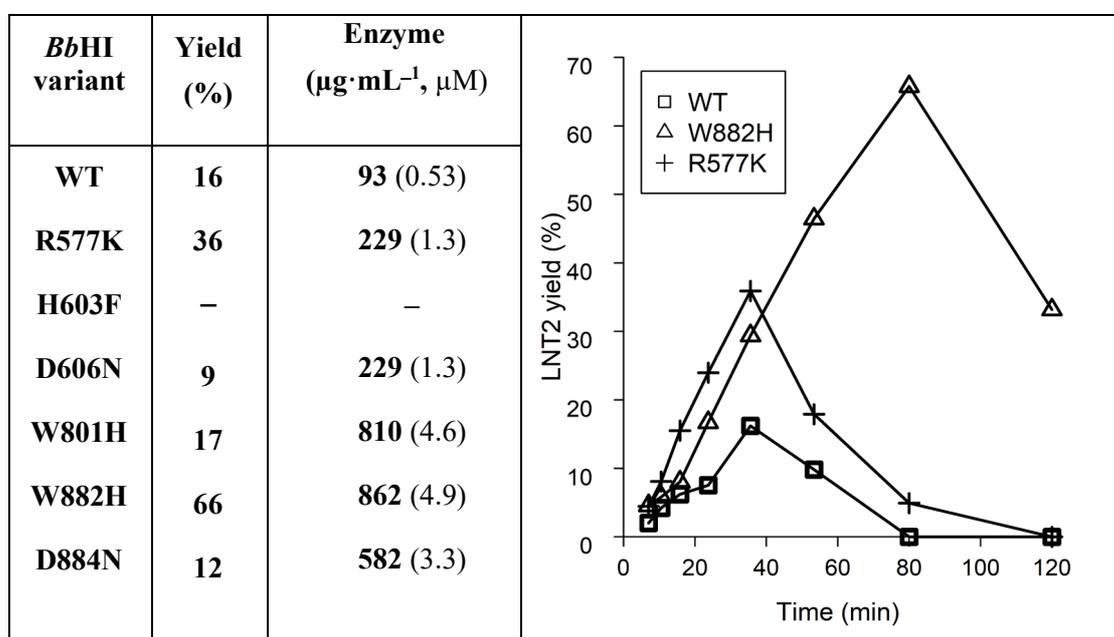
220 **Fig. 4 | GH20 engineering. Left,** maximum yields of LNT2 synthetized by *Bb*HI forms. Protein concentrations
221 used to obtain the yields within 2 h are indicated in µg·mL⁻¹ and µM (in brackets). **Right,** HPAEC-PAD mon-
222 itoring of LNT2 production by *Bb*HI-WT and its two best mutants.

223   The variants R577K and W882H displayed high yields of LNT2 (66% for W882H),
224 although the latter was also poorly active. Nevertheless, *Bb*HI-R577K represents an inter-
225 esting compromise, since compared to *Bb*HI-WT, it generated a markedly higher yield (36
226 vs 16%) of LNT2 at a reasonable enzyme loading (1.3 µM). Importantly, these results
227 demonstrate that our strategy is not restricted to retaining GHs catalysing transglycosylation
228 through a covalent glycosyl-enzyme intermediate. Conversely, the application of our ap-
229 proach to another GH20, from *Ewingella americana* failed to generate variants able to per-
230 form transglycosylation using β-D-Glc*p*NAcO*p*NP as donor and lactose as an acceptor (data
231 not shown). Although this might indicate that the strategy is not fully generic, we believe
232 that this was unsuccessful due to the fact that the WT GH20 from *Ewingella americana*
233 displays no detectable ability to transglycosylate lactose (no other acceptors were tested).
234 This emphasises that the existence of innate, albeit weak, transglycosylation activity is pos-
235 sibly a prerequisite for success when using our approach.

236

237 **GH29, α-L-fucosidases AlfB and AlfC.** L-Fucose is the most common L-sugar in animals[55],
238 invariably connected through an axial glycosidic linkage. Accordingly, a $^1C_4 \rightarrow [^3H_4]^{\ddagger} \rightarrow {}^3S_1$
239 conformational itinerary is followed in enzymatic hydrolysis[47]. A few studies have been
240 undertaken to improve GH29-mediated transfucosylation[56,57], including one in which di-
241 rected evolution was used[58], albeit with varying success. For this study, we focused on AlfB
242 from *Lactobacillus casei*. This enzyme is reported to synthesize α-L-Fuc*p*-(1→3)-D-Glc-
243 NAc[59,60], the α-L-Fuc*p*-(1→3) motif being particularly common with 8 out of 13 known
244 human fucosyltransferases being 3-fucosyltransferases[55]. We assessed AlfB-WT and seven
245 of its mutants for their ability to synthesise α-L-Fuc*p*-(1→3)-D-GlcNAc from 20 mM 2-
246 chloro-4-nitrophenyl α-L-fucopyranoside (α-L-Fuc*p*OCNP) and 20 mM D-GlcNAc (Figs. 5
247 and S7).

7

Variant	Yield (%)	Enzyme ($\mu g \cdot mL^{-1}$, μM)
AlfB-WT	18	65 (1.4)
H21F	–	–
Y45F	21	20 (0.4)
H80F	57	250 (5.2)
H81F	1.3	465 (10)
Y124F	15	740 (16)
D129N	22	117 (2.5)
W130H	31	750 (16)
AlfC-WT	17	150 (1.0)
H87F	26	4.500 (30)
W137H	22	500 (3.3)
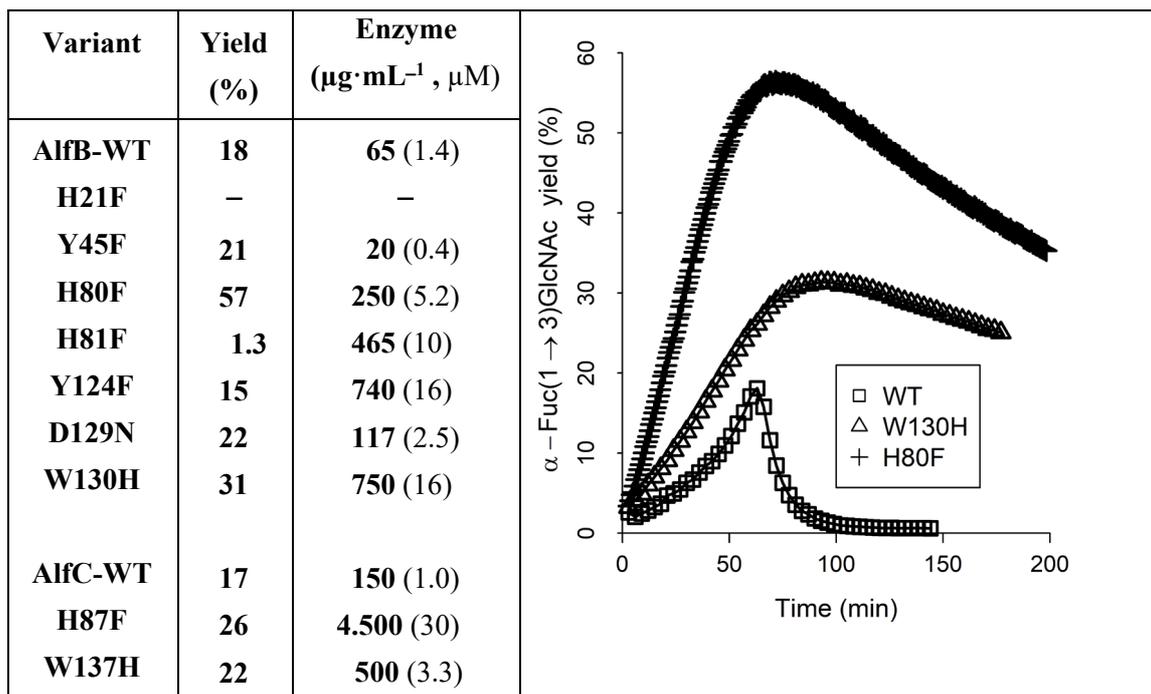
Fig. 5 | GH29 engineering. Left, maximum yields of α-L-Fuc*p*-(1→3)-D-GlcNAc and α-L-Fuc*p*-(1→6)-D-GlcNAc synthetized by AlfB and AlfC forms, respectively. Protein concentrations used to obtain maximum yields within 3 h are given in $\mu g \cdot mL^{-1}$ and μM (in brackets). Right, NMR monitoring of the product formation with AlfB-WT and its two best mutants.

The mutant H21F presented barely detectable activity and was eliminated. Out of the six remaining variants, H80F and W130H presented markedly higher yields than AlfB-WT. AlfB-H80F procured very high transfucosylation yield (57%), while displaying high transglycosylation rate (> 11 mM α-L-Fuc*p*-(1→3)-D-GlcNAc was obtained within 1 h with 5 µM enzyme) and complete regioselectivity. It is noteworthy that mutation of the neighbouring residue (H81F) almost eliminated transglycosylation. Therefore, these two adjacent, conserved histidines possess opposite but determinant roles in regulating the T/H ratio. In reactions catalysed by AlfB-WT the transglycosylation product is rapidly hydrolysed after complete consumption of the donor substrate (Figs. S6 and S7). However, secondary hydrolysis was significantly decreased particularly for AlfB-W130H.

The two successful H80F and W130H mutations were transposed to another fucosidase from *Lactobacillus casei*, AlfC, which is reported to synthesize α-L-Fuc*p*-(1→6)-D-GlcNAc[59,60]. Although AlfB and AlfC are only distantly related (below 30% identity), analysis of AlfC-H87F and AlfC-W137H revealed that these also procured increased disaccharide yields in reactions containing equimolar (8 mM) amounts of α-L-Fuc*p*OCNP and D-GlcNAc (Figs. 5, S8 and S9). Significantly, this result illustrates how successful mutations generated in one enzyme can be transposed to other GHs without the need to perform further analyses. In addition to the reduction of the sequence space and increased relevance, mutations transfer is a key advantage of targeting conserved residues.

GH51, α-L-arabinofuranosidase *Tx*Abf. Five-carbon furanose rings are notoriously more flexible and thermodynamically less stable than their six-carbon counterparts, meaning that unlike pyranoses, furanoses can interconvert between different ring conformations. This reflects similar energy states for the different conformations[61,62], a point that complicates the organic synthesis of furanosides. As catalysis mediated by any given GH involves a

specific conformational itinerary for the donor substrate, the lowered energy barriers be-
tween furanoside conformations[61] implies that furanosidase-catalysed reactions display al-
tered mechanisms when compared to those catalysed by pyranosidases. Therefore, it is of
interest to investigate to which extent the GH engineering approach described herein ap-
plies to furanosidases.

The hydrolytic α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*Tx*Abf)
belonging to the GH51 family[63] was used as a model furanosidase. Displaying inherent abil-
ity to perform transfuranosylation[64,65], *Tx*Abf has been the target of several studies aimed at
improving its transglycosylation capability[21,66–68]. All eight mutants generated in this study
showed improved ability to synthesize arabinoxylo-oligosaccharides (AXOS). Compared to
*Tx*Abf-WT (9% yield), AXOS yields for the variants were in the range 19–82% (Fig. 6 and
Table S4), with R69K, N175T, D297N and H240F/N being the best performers (37–82%
overall yields). The transglycosylation activity (specific activity in transglycosylation mode,
SA_T, Table S5) of the mutants was lower (0.1–66%) than that of the WT enzyme. It is note-
worthy that R69K-, H240F- and D297N-catalysed transglycosylation reactions required rel-
atively low amounts of catalyst (1.8−8 µg·mL^{-1}, i.e. 32−140 nM; Fig. 6).

*Tx*Abf variant	Yield[a] (%)	Enzyme[b] (µg·mL^{-1}, nM)
WT	9	0.2 (3.5)
F26H	22	730 (12 800)
E28Q	23	0.8 (14)
R69K	37	2 (35)
N175T	42	200 (3 500)
H240F	82	8 (140)
H240N	75	220 (3 900)
Y242F	19	300 (5 300)
D297N	47	1.8 (32)
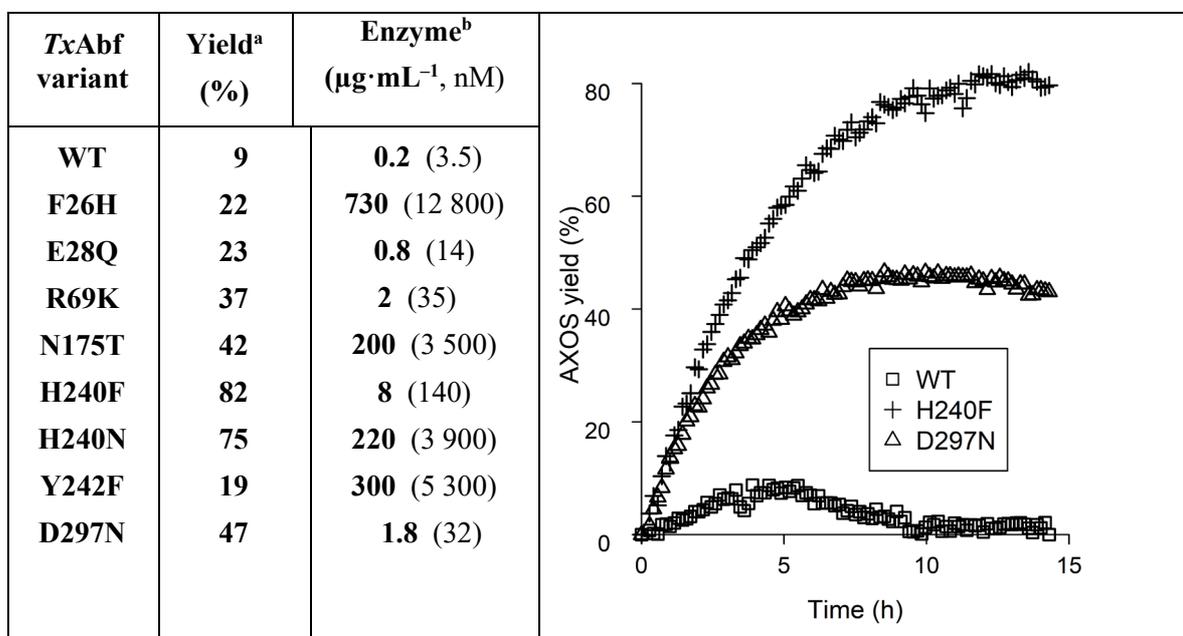
Fig. 6 | GH51 engineering. Left, [a]maximum overall yields of arabinoxylo-tetrasaccharides synthetized by
*Tx*Abf forms using 5 mM α-L-Ara*fOp*NP as donor and 10 mM xylotriose as acceptor. The yields of each of the
regioisomers are indicated in Table S4. [b]Protein concentrations used to obtain the yields within 15 h are given
in µg·mL^{-1} and nM (in brackets). **Right,** NMR monitoring of transglycosylation yield progress for *Tx*Abf-WT
and two of its more significant mutants.

Previous work using random mutagenesis and screening already yielded mutants
F26L and R69H[21] that, compared to F26H and R69K described herein, are slightly better
catalysts for transglycosylation (1.2-fold). However, the current strategy is less labour-in-
tensive, as it circumvents large library screening. In this respect, it is remarkable that the
variants H240N and H240F display different yields and regioselectivities (Table S4), thus
illustrating the value of further probing and fine-tuning of hotspots identified using our strat-
egy.

Significantly, compared to all previously reported single-mutants that enhance the T/H ratio
in *Tx*Abf[21,66–68], H240F/N display the highest transglycosylation yields combined with no-
ticeably greater regioselectivity (Table S4). Thus, we combined this mutation with the

9

N216W[21] one, that favours higher regioselectivity towards the (1→2)-linkage of α-L-arabi-nofuranosyl moieties to the non-reducing terminal D-xylopyranosyl of xylotriose as acceptor (*i.e.* A[2]XX)[21]. While both transglycosylation activity (SA$_T$, Table S5) and yield (62 compared to 82% for H240F) remained relatively high, the reaction was almost completely regioselective (Table S4 and Fig. S10). Conveniently, available structural data and the considerable corpus of knowledge related to *Tx*Abf furnish hypotheses to explain how the different mutations enhance the T/H ratio (Fig. S11 and related discussion).

Discussion

Mutating enzymes from GH families 2, 10, 20, 29, and 51, which represent different clans[5] (A, K and R), led to the successful enhancement of transglycosylation yields (2- to 9-fold compared to WT enzymes) in more than 50% of selected candidates. This powerful demonstration validates our conserved-residue approach and illustrates its applicability to retaining GHs irrespective of the structural fold and the specific mechanism. Moreover, a variety of sugars, D/L-configurations, pyranose/furanose forms and α/β-stereochemistry are tolerated, thus new access to hitherto refractory syntheses is provided. In a rather fast and direct manner, the strategy procured the means to reach transglycosylation yields in the range 50−80%, and thus allowed the high yield synthesis of oligosaccharides such as α-L-Fuc*p*-(1→3)-D-GlcNAc, α-L-Fuc*p*-(1→6)-D-GlcNAc, lacto-*N*-triose II (LNT2), oligomannosides, oligoxylosides and arabinoxylo-oligosaccharides. Remarkably, in each of the five GHs families targeted, while using reasonable enzyme loadings, it proved possible to obtain at least one mutant displaying enhanced transglycosylation yield.

Our approach is unusual in that it requires neither extensive screening, nor in-depth knowledge of the target enzyme. This is possible, because the method systematically targets conserved residues, which are generally omitted in enzyme engineering approaches in order to avoid loss of activity or stability[69]. It is a powerful approach here as it does not aim to improve a defining characteristic of an enzyme, but rather sets out to eliminate a property (*i.e.* the ability to perform hydrolysis). We anticipate that our method can be applied to other enzymes displaying activities or properties that must be suppressed rather than enhanced. One obvious caveat is that some key determinants of the T/H ratio might not be conserved residues and will thus be undetectable using this method. Thus, this strategy is not very likely to lead to the absolute best possible solution, but it does swiftly lead to an improvement over the WT enzyme.

Converting hydrolytic GHs into potent glycosynthetic tools is an attractive approach to extend the synthetic chemist's toolbox, while introducing catalysts that obey green chemistry principles (e.g. use of non-toxic catalysts, aqueous solvents). However, so far, the success of this approach has been hampered by the time and effort necessary to develop appropriate biocatalysts for each target reaction. The strategy presented herein goes a long way to surmounting this obstacle, making it much simpler to obtain tailored biocatalysts that can then be used to operate straightforward, relatively inexpensive synthesis reactions that do not require difficult to obtain sugar donors, lengthy protection/deprotection cycles or exorbitant quantities of enzyme. Moreover, the transferability implies that the reported mutations can be readily transferred to other GHs from the five described families, expanding the portfolio of available evolved transglycosylases and synthetic oligosaccharides.

10

References

1. Laine, R. A. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the isomer barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* **4**, 759–767 (1994).

2. Nielsen, M. M. & Pedersen, C. M. Catalytic glycosylations in oligosaccharide synthesis. *Chem. Rev.* **118**, 8285–8358 (2018).

3. Panza, M., Pistorio, S. G., Stine, K. J. & Demchenko, A. V. Automated chemical oligosaccharide synthesis: novel approach to traditional challenges. *Chem. Rev.* **118**, 8105–8150 (2018).

4. Krasnova, L. & Wong, C. H. Oligosaccharide synthesis and translational innovation. *J. Am. Chem. Soc.* **141**, 3735–3754 (2019).

5. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490-5 (2014).

6. Yonekura-Sakakibara, K. & Hanada, K. An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* **66**, 182–193 (2011).

7. Faijes, M., Castejón-Vilatersana, M., Val-Cid, C. & Planas, A. Enzymatic and cell factory approaches to the production of human milk oligosaccharides. *Biotechnol. Adv.* **37**, 667–697 (2019).

8. Koshland, D. E. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* **28**, 416–436 (1953).

9. Vasella, A., Davies, G. J. & Böhm, M. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **6**, 619–29 (2002).

10. Teze, D. *et al.* The catalytic acid-base in GH109 resides in a conserved GGHGG loop and allows for comparable α-retaining and β-inverting activity in an *N*-acetylgalactosaminidase from *Akkermansia muciniphila*. *ACS Catal.* **10**, 3809–3819 (2020).

11. Lundemo, P., Karlsson, E. N. & Adlercreutz, P. Eliminating hydrolytic activity without affecting the transglycosylation of a GH1 β-glucosidase. *Appl. Microbiol. Biotechnol.* **101**, 1121–1131 (2017).

12. Bissaro, B., Monsan, P., Fauré, R. & O'Donohue, M. J. J. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* **467**, 17–35 (2015).

13. Mackenzie, L. F., Wang, Q., Warren, R. A. J. & Withers, S. G. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.* **120**, 5583–5584 (1998).

14. Hayes, M. R. & Pietruszka, J. Synthesis of glycosides by glycosynthases. *Molecules* **22**, (2017).

15. Malet, C. & Planas, A. From β-glucanase to β-glucansynthase: Glycosyl transfer to α-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS Lett.* **440**, 208–212 (1998).

16. Moracci, M., Trincone, A., Perugino, G., Ciaramella, M. & Rossi, M. Restoration of the activity of active-site mutants of the hyperthermophilic β-glycosidase from *Sulfolobus solfataricus*: Dependence of the mechanism on the action of external nucleophiles. *Biochemistry* **37**, 17262–17270 (1998).

17. Mayer, C., Zechel, D. L., Reid, S. P., Warren, R. A. J. & Withers, S. G. The E358S mutant of *Agrobacterium sp.* β-glucosidase is a greatly improved glycosynthase. *FEBS Lett.* **466**, 40–44 (2000).

18. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev.*

11

409   *Genet.* **16**, 379–394 (2015).

410   19.   Teze, D., Daligault, F., Ferrières, V., Sanejouand, Y.-H. & Tellier, C. Semi-rational
411   approach for converting a GH36 α-glycosidase into an α-transglycosidase.
412   *Glycobiology* **25**, 420–427 (2015).

413   20.   Teze, D. *et al.* Semi-rational approach for converting a GH1 β-glycosidase into a β-
414   transglycosidase. *Protein Eng. Des. Sel.* **27**, 13–19 (2014).

415   21.   Bissaro, B. *et al.* Molecular design of non-Leloir furanose-transferring enzymes
416   from an α-L-arabinofuranosidase: a rationale for the engineering of evolved
417   transglycosylases. *ACS Catal.* **5**, 4598–4611 (2015).

418   22.   Yang, J. *et al.* Engineering T. naphthophila β-glucosidase for enhanced synthesis of
419   galactooligosaccharides by site-directed mutagenesis. *Biochem. Eng. J.* **127**, 1–8
420   (2017).

421   23.   Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding
422   subsites in glycosyl hydrolases. *Biochem. J.* **321**, 557–559 (1997).

423   24.   Li, W., Jaroszewski, L. & Godzik, A. Clustering of highly homologous sequences to
424   reduce the size of large protein databases. *Bioinformatics* **17**, 282–283 (2001).

425   25.   Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT suite: a web server for
426   clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).

427   26.   Sievers, F. & Higgins, D. G. Clustal Omega. *Curr. Protoc. Bioinforma.* **2014**,
428   3.13.1-3.13.16 (2014).

429   27.   Davies, G. J., Planas, A. & Rovira, C. Conformational analyses of the reaction
430   coordinate of glycosidases. *Acc. Chem. Res.* **45**, 308–316 (2011).

431   28.   Gridley, J. J. & Osborn, H. M. I. Recent advances in the construction of β-D-
432   mannose and β-D-mannosamine linkages. *J. Chem. Soc. Perkin Trans. 1* 1471–1491
433   (2000).

434   29.   Nashiru, O. *et al.* β-mannosynthase: Synthesis of β-mannosides with a mutant β-
435   mannosidase. *Angew. Chemie - Int. Ed.* **40**, 417–420 (2001).

436   30.   Jahn, M. *et al.* Expansion of the glycosynthase repertoire to produce defined manno-
437   oligosaccharides. *Chem. Commun.* **3**, 1327–1329 (2003).

438   31.   Sasaki, A., Ishimizu, T., Geyer, R. & Hase, S. Synthesis of β-mannosides using the
439   transglycosylation activity of endo-β-mannosidase from *Lilium longiflorum*. *FEBS*
440   *J.* **272**, 1660–1668 (2005).

441   32.   Zechel, D. L. *et al.* Mechanism, mutagenesis, and chemical rescue of a β-
442   mannosidase from *Cellulomonas fimi*. *Biochemistry* **42**, 7195–7204 (2003).

443   33.   Eneyskaya, E. V. *et al.* Transglycosylating and hydrolytic activities of the β-
444   mannosidase from *Trichoderma reesei*. *Biochimie* **91**, 632–638 (2009).

445   34.   Shi, P. *et al.* Cloning and characterization of a new β-mannosidase from
446   *Streptomyces sp.* S27. *Enzyme Microb. Technol.* **49**, 277–283 (2011).

447   35.   Morrill, J. *et al.* β-Mannanase-catalyzed synthesis of alkyl mannooligosides. *Appl.*
448   *Microbiol. Biotechnol.* **102**, 5149–5163 (2018).

449   36.   Rosengren, A. *et al.* Enzymatic synthesis and polymerisation of β-mannosyl
450   acrylates produced from renewable hemicellulosic glycans. *Green Chem.* **21**, 2104–
451   2118 (2019).

452   37.   Iglesias-Fernández, J., Raich, L., Ardèvol, A. & Rovira, C. The complete
453   conformational free energy landscape of β-xylose reveals a two-fold catalytic
454   itinerary for β-xylanases. *Chem. Sci.* **6**, 1167–1177 (2015).

455   38.   Evangelista, D. E., Kadowaki, M. A. S., Mello, B. L. & Polikarpov, I. Biochemical
456   and biophysical characterization of novel GH10 xylanase prospected from a sugar
457   cane bagasse compost-derived microbial consortia. *Int. J. Biol. Macromol.* **109**,
458   560–568 (2018).

459   39.   Hu, J. & Saddler, J. N. Why does GH10 xylanase have better performance than

460 GH11 xylanase for the deconstruction of pretreated biomass? *Biomass Bioenerg.*
461 **110**, 13–16 (2018).

40. Nordberg Karlsson, E., Schmitz, E., Linares-Pastén, J. A. & Adlercreutz, P. Endo-
462 xylanases as tools for production of substituted xylooligosaccharides with prebiotic
463 properties. *Appl. Microbiol. Biotechnol.* **102**, 9081–9088 (2018).
464

41. Niderhaus, C., Garrido, M., Insani, M., Campos, E. & Wirth, S. Heterologous
465 production and characterization of a thermostable GH10 family endo-xylanase from
466 *Pycnoporus sanguineus* BAFC 2126. *Process Biochem.* **67**, 92–98 (2018).
467

42. Armand, S., Andrews, S. R., Charnock, S. J. & Gilbert, H. J. Influence of the
468 aglycone region of the substrate binding cleft of *Pseudomonas* xylanase 10A on
469 catalysis. *Biochemistry* **40**, 7404–7409 (2001).
470

43. Charnock, S. J. *et al.* Key residues in subsite F play a critical role in the activity of
471 *Pseudomonas fluorescens* subspecies cellulosa xylanase A against
472 xylooligosaccharides but not against highly polymeric substrates such as xylan. *J.*
473 *Biol. Chem.* **272**, 2942–2951 (1997).
474

44. Charnock, S. J. *et al.* The topology of the substrate binding clefts of glycosyl
475 hydrolase family 10 xylanases are not conserved. *J. Biol. Chem.* **273**, 32187–32199
476 (1999).
477

45. Moreau, A., Shareck, F., Kluepfel, D. & Morosoli, R. Alteration of the cleavage
478 mode and of the transglycosylation reactions of the xylanase A of *Streptomyces*
479 *lividans* 1326 by site-directed mutagenesis of the Asn173 residue. *Eur. J. Biochem.*
480 **219**, 261–266 (1994).
481

46. Aronsson, A. *et al.* Structural insights of RmXyn10A – A prebiotic-producing
482 GH10 xylanase with a non-conserved aglycone binding region. *Biochim. Biophys.*
483 *Acta - Proteins Proteomics* **1866**, 292–306 (2018).
484

47. Roth, C. *et al.* Structural and mechanistic insights into a *Bacteroides vulgatus*
485 retaining *N*-acetyl-β-galactosaminidase that uses neighbouring group participation.
486 *Chem. Commun.* **52**, 11096–11099 (2016).
487

48. Coines, J., Alfonso-Prieto, M., Biarnés, X., Planas, A. & Rovira, C. Oxazoline or
488 oxazolinium ion? the protonation state and conformation of the reaction
489 intermediate of chitinase enzymes revisited. *Chem. Eur. J.* **24**, 19258–19265 (2018).
490

49. Teze, D. *et al.* A single point mutation converts GH84 *O*-GlcNAc hydrolases into
491 phosphorylases: experimental and theoretical evidence. *J. Am. Chem. Soc.* **142**,
492 2120–2124 (2020).
493

50. Slámová, K., Bojarová, P., Petrásková, L. & Křen, V. β-*N*-Acetylhexosaminidase:
494 what's in a name...? *Biotechnol. Adv.* **28**, 682–693 (2010).
495

51. Slámová, K. *et al.* Synthesis of derivatized chitooligomers using transglycosidases
496 engineered from the fungal GH20 β-N-acetylhexosaminidase. *Adv. Synth. Catal.*
497 **357**, 1941–1950 (2015).
498

52. Miwa, M. *et al.* Cooperation of β-galactosidase and β-*N*-acetylhexosaminidase from
499 bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure.
500 *Glycobiology* **20**, 1402–1409 (2010).
501

53. Zeuner, B., Teze, D., Muschiol, J. & Meyer, A. S. Synthesis of human milk
502 oligosaccharides : protein engineering strategies for improved enzymatic
503 transglycosylation. *Molecules* **24**, 2033–2055 (2019).
504

54. Weingarten, M., Baldenius, K. & Nidetzky, B. Glycosynthase principle transformed
505 into biocatalytic process technology: Lacto-*N*-triose II production with engineered
506 *exo*-hexosaminidase. *ACS Catal.* **9**, 5503–5514 (2019).
507

55. Schneider, M., Al-Shareffi, E. & Haltiwanger, R. S. Biological functions of fucose
508 in mammals. *Glycobiology* **27**, 601–618 (2017).
509

56. Zeuner, B., Vuillemin, M., Holck, J., Muschiol, J. & Meyer, A. S. Loop engineering
510

13

511 of an α-1,3/4-L-fucosidase for improved synthesis of human milk oligosaccharides.
512 *Enzyme Microb. Technol.* **115**, 37–44 (2018).

513 57. Saumonneau, A. *et al.* Design of an α-L-transfucosidase for the synthesis of
514 fucosylated HMOs. *Glycobiology* **26**, 261–269 (2015).

515 58. Osanjo, G. *et al.* Directed evolution of the α-L-fucosidase from *Thermotoga*
516 *maritima* into an α-L-transfucosidase. *Biochemistry* **46**, 1022–1033 (2007).

517 59. Rodríguez-Díaz, J., Carbajo, R. J., Pineda-Lucena, A., Monedero, V. & Yebra, M. J.
518 Synthesis of fucosyl-*N*-acetylglucosamine disaccharides by transfucosylation using
519 α-L-fucosidases from *Lactobacillus casei. Appl. Environ. Microbiol.* **79**, 3847–3850
520 (2013).

521 60. Rodríguez-Díaz, J., Monedero, V. & Yebra, M. J. Utilization of natural fucosylated
522 oligosaccharides by three novel α-L-fucosidases from a probiotic *lactobacillus casei*
523 strain. *Appl. Environ. Microbiol.* **77**, 703–705 (2011).

524 61. Taha, H. A., Richards, M. R. & Lowary, T. L. Conformational analysis of
525 furanoside-containing mono- and oligosaccharides. *Chem. Rev.* **113**, 1851–1876
526 (2013).

527 62. Wang, X. & Woods, R. J. Insights into furanose solution conformations: beyond the
528 two-state model. *J. Biomol. NMR* **64**, 291–305 (2016).

529 63. Debeche, T., Cummings, N., Connerton, I., Debeire, P. & O'Donohue, M. J. Genetic
530 and biochemical characterization of a highly thermostable α-L- arabinofuranosidase
531 from *Thermobacillus xylanilyticus. Appl. Environ. Microbiol.* **66**, 1734–1736
532 (2000).

533 64. Rémond, C. *et al.* Synthesis of pentose-containing disaccharides using a
534 thermostable α-L-arabinofuranosidase. *Carbohydr. Res.* **339**, 2019–2025 (2004).

535 65. Remond, C., Plantier-Royon, R., Aubrya, N. & O'Donohue, M. J. An original
536 chemoenzymatic route for the synthesis of β-D-galactofuranosides using an α-L-
537 arabinofuranosidase. *Carbohydr. Res.* **340**, 637–644 (2005).

538 66. Bissaro, B. *et al.* Mutation of a pH-modulating residue in a GH51 α-L-
539 arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of
540 transfuranosylation products. *Biochim. Biophys. Acta* **1840**, 626–636 (2014).

541 67. Arab-jaziri, F. *et al.* Enhancing the chemoenzymatic synthesis of arabinosylated
542 xylo-oligosaccharides by GH51 α-L-arabinofuranosidase. *Carbohydr. Res.* **401**, 64–
543 72 (2015).

544 68. Arab-Jaziri, F. *et al.* Engineering transglycosidase activity into a GH51 α-L-
545 arabinofuranosidase. *New Biotechnol.* **30**, 536–544 (2013).

546 69. Bendl, J. *et al.* HotSpot Wizard 2.0 : automated design of site-specific mutations and
547 smart libraries in protein engineering. *Nucleic Acids Res.* **44**, 479–487 (2016).

548 70. Apweiler, R. *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids*
549 *Res.* **32**, D115-9 (2004).

550 71. Covington, A. K., Paabo, M., Robinson, R. A. & Bates, R. G. Use of the glass
551 electrode in deuterium oxide and the relation between the standardized pD (pa$_D$)
552 scale and the operational pH in heavy water. *Anal. Chem.* **40**, 700–706 (1968).

553
554

Acknowledgements

Author contributions

The project was conceived by D.T. and overlooked and coordinated by D.T. and B.S.. Y.-H.S. conceived the conservation analysis approach, and D.T performed the bioinformatics. M.W. and H.S. conceived the GH2 experiments, which were performed by M.W. and G.C. (NMR analysis). K.Z.G.A. and E.N.-K. conceived the GH10 experiments, which were performed by K.Z.G.A.. B.S., D.T. and R.L. conceived the GH20 experiments, which were performed by R.L.. B.S., D.T. and J.D. conceived the GH29 experiments, which were performed by D.T.. GH29 were produced and purified by M.E.R.. J. Z. and R.F. conceived the GH51 experiments, which were performed by J.Z and supervised by M.O.D. and R.F.. D.T. wrote the first draft, and all authors contributed to write the paper.

Competing interests

The authors declare no competing financial interests.

Methods

Materials. Genes and genes variants, all codon-optimized for use in *E. coli* and inserted in pET24a, pET28b(+) or pET28a(+), were ordered from GenScript (Piscataway, USA) or Biomatik (Ontario, Canada), respectively. Gene sequences can be found in the Uniprot database[70] under the codes Q9XCV4 (*Cf*Man2A), P96988 (Xyn10A), D4QAP4 (BbHI), A0A125UD88 (AlfB), A0A422MHI3 (AlfC) and O69262 (*Tx*Abf). Substrates were ordered from either Sigma-Aldrich, Carbosynth or Megazyme.

General procedures

Bioinformatics. Protein BLAST searches were performed on the NCBI server (https://blast.ncbi.nlm.nih.gov/Blast.cgi), from the non-redundant protein database[70], using default options, except in the case of the "Max target sequences" parameter, which was set at 20000. Queries were made between January 31st, 2019 and May 3rd, 2019. Obtained sequences were clustered[24] to limit pairwise sequence identity at 80% by iterative cd-hit runs[25]. Iterative multiple sequence alignments were performed using ClustalΩ[26] to progressively increase minimum pairwise sequence identity to a predefined threshold (10–20%) and reach convergence using make_msa.sh, a homemade bash script, available at https://gitlab.univ-nantes.fr/sanejouand-yh/Sequences/tree/master. The same script was used to analyse sequence conservation.

Protein production and purification. Unless otherwise specified, pET24a, pET28a(+) or pET28b(+) plasmids bearing target genes were used to transform BL21(DE3) *E. coli* cells. Precultures of transformed cells were used to inoculate lysogeny broth media containing 30–50 mg·L^{-1} kanamycin. Cultures (0.5−2 L) were incubated at 37°C with shaking until OD$_{600}$ reached ~0.5–1. Gene expression was induced with 200–500 μM isopropyl-β-D-1-thiogalactopyranoside and continued 3–16 h at 20–37°C. Cultures were subsequently centrifuged,

15

609 pellets resuspended, lysed, and centrifuged. The enzymes were purified from the supernatant
610 by Ni^{2+} (or Co^{2+})-affinity IMAC chromatography. Eluates were analysed by SDS-PAGE,
611 and protein concentration was determined spectrophotometrically using UV$_{280}$ absorption
612 and molar extinction coefficients calculated in ExPASy (www.expasy.org).
613
614 **GH Family-specific procedures**
615
616 **GH2, *Cf*Man2A.** Activity of the *Cf*Man2A and its variants were determined by measuring
617 the release of *p*-nitrophenol (*p*NP) from *p*-nitrophenyl-β-D-mannopyranoside (β-D-
618 Man*pO*p*NP), using a version of the method described by Zechel et al[32] adapted for micro-
619 plate assay. Briefly, reactions mixtures containing 1 or 5 mM β-D-Man*pO*p*NP, 35 mM so-
620 dium phosphate pH 7.0 and appropriately diluted enzyme were incubated at 35°C for 10
621 min. The reaction was stopped by adding 1 M Na$_2$CO$_3$ and released *p*NP was measured on
622 an Epoch Microplate spectrophotometer (BioTek Instruments, USA) at 405 nm. All assays
623 and reactions (below) were in duplicates.
624 To evaluate the transglycosylation ability of the *Cf*Man2A forms, reactions containing 5 mM
625 β-D-Man*pO*p*NP and 3−34 μg·mL^{-1} enzyme in 35 mM sodium phosphate pH 7.0 were incu-
626 bated up to 6 h at 35°C. Aliquots were collected at appropriate time intervals throughout the
627 reaction, heat denatured (95°C, 10 min) and filtered (0.22 μm PTFE). Aliquots were sepa-
628 rated at 40°C on a Luna Omega SUGAR HPLC-column (Phenomenex, USA) using an Ulti-
629 Mate 3000 HPLC system (Thermo Fisher Scientific, USA) with 40:60 water:acetonitrile el-
630 uent (v/v) at 1 mL·min^{-1}. A VWD-3400RS detector was used (Thermo Fisher Scientific,
631 USA), measuring absorbance at 300 nm. The presence of transglycosylation products were
632 determined with mass spectrometry and analysed with NMR spectroscopy as described in
633 Supplementary Information, section 1.2.
634 Progress curves of *p*NP-Man$_2$ production for *Cf*Man2A forms were generated from HPLC
635 analysis as above for reaction times up to 6 h (7−10 sampling points, see Fig. 2). *p*NP-Man$_2$
636 was quantified using *p*NP-cellobioside as a standard. The yield of *p*NP-Man$_2$ was calculated
637 as the amount of β-D-Man*pO*p*NP used as either acceptor or donor in the production of *p*NP-
638 Man$_2$ divided by the amount of loaded β-D-Man*pO*p*NP.
639
640 **GH10, *Rm*Xyn10A_CM.** GH10 transglycosylation activity was characterized by incubating
641 20 mM X$_4$ with *Rm*Xyn10A_CM forms. Reaction mixtures containing 10–43 μg·mL^{-1}
642 GH10 in 20 mM sodium phosphate pH 7.0 were incubated at 65°C for 4 h. Aliquots (15 μL)
643 were withdrawn at different time points, diluted in 0.5 mM NaOH, filtered, and analysed on
644 high performance anion-exchange chromatography coupled with pulsed amperometric de-
645 tection (HPAEC-PAD) using an ICS-5000 (Dionex) monitored by the software Chromeleon.
646 Separation was carried out at 30°C on a CarboPac PA-200 using 100 mM NaOH at 0.5
647 mL·min^{-1} and a linear gradient of 0–120 mM sodium acetate.
648 The reaction mixtures for MALDI-TOF-MS analysis contained 10–43 μg·mL^{-1} GH10 in 20
649 mM sodium phosphate pH 7.0 and were incubated at 60°C for 4 h. MALDI-TOF-MS spectra
650 were obtained on a Bruker Daltonics Autoflex Speed MALDI-TOF (/TOF) spectrometer in
651 positive ion reflector mode and recorded in the mass range from 200 to 4000 or 5000 Da.
652 Samples were diluted in MilliQ H$_2$O to a total salt concentration <10 mM, and 1 μL of this
653 dilution was mixed with 0.5 μL aqueous 10% dihydroxybenzoic acid (DHB) matrix solution
654 on a stainless steel target and left to dry at room temperature. Sample irradiation was done
655 at 55% laser power by targeting the laser pulses at amorphous crystal regions, regularly
656 shifting to remove heterogeneity in the sample. Calibration was done internally by addition
657 of xylo-oligosaccharides X$_1$−X$_6$. The hydrolysis assay was performed in 0.1 mL reactions
658 with 2 mM β-D-xylotriosideO*p*NP in 20 mM sodium phosphate pH 7.0, at 70°C for 5 min

and the reaction was stopped using one volume of 0.1 M NaOH. The absorbance was measured at 400 nm using a Multiskan spectrophotometer from Thermo Scientific. The extinction coefficient (18250 M·cm^{-1}) of released p-nitrophenol (pNP) was determined using a standard of pNP (Sigma).

GH20, *Bb*HI. GH20 transglycosylation was monitored by HPAEC-PAD. Samples with 0.5–5 µM GH20, 10 mM β-D-Glc$p$NAcOpNP and 40 mM lactose in 50 mM sodium phosphate, 0.1% BSA pH 7.0 were incubated (2 h, 37°C), then heat denatured (2 min, 98°C), centrifuged, and the obtained supernatant was diluted five-fold in milliQ H$_2$O and filtered. Separation was carried out at 30°C on a CarboPac PA-1 (Dionex) using an ICS-5000 (Dionex) monitored by the software Chromeleon (Dionex). NaOH 0.1 M was used as eluent in 20 min separations at 250 µL·min^{-1}.

GH29, AlfB and AlfC. GH29 transfucosylation was assessed from NMR spectra recorded on an 800 MHz Bruker Avance III (799.75 MHz for ^{1}H) equipped with a 5 mm TCI cryoprobe using ^{1}H with presaturation. AlfB-catalysed reactions were carried out with 20 mM α-L-Fuc$p$OCNP, 20 mM D-GlcNAc, and 20–750 µg·mL^{-1} (0.4–16 nM) GH29 in 600 µL 40 mM sodium phosphate pD 7.0 (pD=pH$_{meter\ reading}$+0.4, 6.6 on readings)[71]. Time course experiments were obtained using pseudo-2D kinetics experiments, with ^{1}H NMR spectra recorded every 3 min. Integration of anomeric protons is inaccurate due to the closeness of the presaturated HOD peak, therefore chemical shifts at 4.26 (H-2 of D-GlcNAc in α-L-Fuc$p$-(1→3)-D-GlcNAc) and 4.05 ppm (α-L-Fuc$p$OCNP) were used for integration (Fig. S6). AlfC-catalysed reactions were identical but using 8 mM donor and acceptor, and the use of the fucose methyl group protons to assess concentrations (Figs. S7 and S8).

GH51, *Tx*Abf. The activities of *Tx*Abf and mutants thereof were determined using a discontinuous enzyme assay[66]. Reactions were performed in triplicate at 45°C in 50 mM sodium phosphate pH 7.0 containing 1 mg·mL^{-1} BSA, using 5 mM α-L-Ara$f$OpNP without/with 10 mM xylotriose as acceptor in hydrolysis and transglycosylation modes, respectively. The amount of pNP released, which was measured at 401 nm, was calculated using an appropriate standard curve of pNP. Negative controls containing all of the reactants except the enzyme were used to correct for spontaneous hydrolysis of the donor substrate. To monitor the transglycosylation products profiles from ^{1}H NMR spectra at 500 MHz on a Bruker Avance II spectrometer equipped with a TCI probe, reactions were performed at 45°C in NMR tubes containing 600 µL 5 mM α-L-Ara$f$OpNP, 10 mM xylotriose and *Tx*Abf enzymes (3.5−12 800 nM) in 10 mM sodium phosphate pH 7.0/D$_2$O: 9/1, v/v, with 1 mg·mL^{-1} BSA. The quantity of enzyme (Fig. 6) was adjusted to suit the 13−19 h reaction time frame. Time course of NMR monitoring was obtained using pseudo-2D kinetics experiments based on a phase sensitive NOESY sequence with presaturation, with spectra collected every approximatively 9 min (i.e. twice 32 scans). The transglycosylation yields were determined by relative integration of anomeric proton signals from the α-L-Ara$f$ unit of each AXOS (Table S4)[21,66]. Concentration of α-L-Ara$f$OpNP donor was quantified by integrating its relevant anomeric proton signals at 5.86 ppm. Molar balances, based on initial donor signal as internal reference, were used to convert the transglycosylation product signal integral into concentration.

1

2   # Rational enzyme design without structural
3   # knowledge: a sequence-based approach for
4   # efficient generation of glycosylation catalysts

5   **David Teze[1,*], Jiao Zhao[2], Mathias Wiemann[3], Kazi Zubaida Gulshan Ara[4], Rossana**
6   **Lupo[1], Mette Errebo Rønne[1], Göran Carlström[5], Jens Duus[6], Yves-Henri**
7   **Sanejouand[7], Michael J. O'Donohue[2], Eva Nordberg-Karlsson[4], Régis Fauré[2], Henrik**
8   **Stålbrand[3], Birte Svensson[1,*]**

9

10

11   ## <u>Contents</u>

12

22

23

[1]Department of Biotechnology and Biomedicine. Technical University of Denmark, Kongens Lyngby, Denmark. [2]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France. [3]Department of Biochemistry and Structural Biology. Lund University, Lund, Sweden. [4]Biotechnology, Department of Chemistry. Lund University, Lund, Sweden. [5]Centre for Analysis and Synthesis, Department of Chemistry. Lund university, Lund, Sweden [6]Department of Chemistry. Technical University of Denmark, Kongens Lyngby, Denmark. [7]Université de Nantes, CNRS, UFIP, France. Correspondence should be addressed to D.T. (datez@dtu.dk) or B.S. (bis@bio.dtu.dk).

24 This SI section contains methods and a range of results from analytics (high-performance liquid
25 chromatography, HPLC; mass spectrometry, MS; nuclear magnetic resonance spectroscopy, NMR)
26 used to identify specific oligosaccharides and monitor their production during transglycosylation
27 reactions.

1) GH2, β-mannosidase *Cf*Man2A

1.1 HPLC and MS analysis of *Cf*Man2A transglycosylation products

32 HPLC (LUNA Omega Sugar column) was used to monitor the formation of transglycosylation prod-
33 ucts of *Cf*Man2A-catalyzed reactions containing 5 mM *p*-nitrophenyl β-D-mannopyranoside (β-D-
34 Man*p*O*p*NP) as both donor and acceptor substrate (Figure S1). For MS and NMR analyses, 5 mM β-
35 D-Man*p*O*p*NP and 34 µg·mL^{-1} *Cf*Man2A-N428T, in 35 mM sodium phosphate pH 7.0 reacted for 2
36 h at 35°C. The reaction was stopped by heat denaturation (95°C, 10 min) and 100 µL was filtered
37 and separated using HPLC. Elution volumes for β-D-Man*p*O*p*NP (peak 1) and *p*NP (peak 5) were
38 determined by injection of chemical standards. Fractions eluting between 3.3 and 4.1 mL were col-
39 lected (85 µL), covering peaks 2 through 5 (Fig. S1). The collected fractions were concentrated ap-
40 proximately 10-fold by evaporation in an RVC 2-18 vacuum concentrator (Martin Christ Freeze
41 Dryers, Germany) and analysed using matrix-assisted laser desorption/ionisation time-of-flight MS
42 (MALDI-TOF MS). Fractions (0.5 µL) were combined with 0.5 µL matrix solution (10 mg·mL^{-1}
43 2,5-dihydroxybenzoic acid in 5 mM sodium acetate pH 5.3) directly on a stainless steel target and
44 dried under warm air. The fractions were analysed on a 4700 Proteomics Analyzer (Applied Biosys-
45 tems, USA) and data were analysed using DataExplorer software (Applied Biosystems, USA). This
46 revealed that peaks 2 and 3 (eluted between 3.3 and 3.6 mL; Fig. S1) contained compounds with m/z
47 corresponding to *p*-nitrophenyl-β-D-mannobioside (*p*NP-Man$_2$). Similarly, MS revealed that frac-
48 tions collected between 3.8 and 4 mL (i.e. peaks 4 and 5) contained *p*-nitrophenyl-β-D-mannotrioside
49 (*p*NP-Man$_3$).
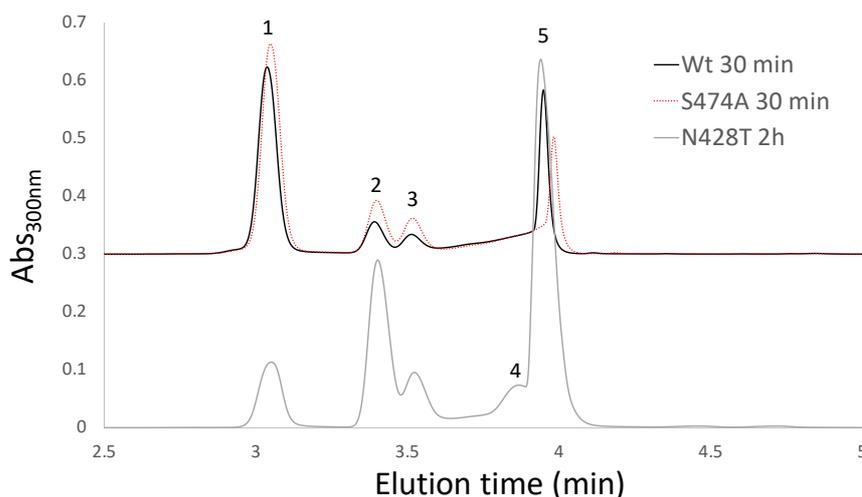
Figure S1 | HPLC analysis of transglycosylation reactions using *Cf*Man2A-WT (black line) and variants S474A (red dotted line; both reaction mixtures diluted four times) and N428T (grey line, undiluted reaction mixture). The N428T reaction mixture was used for NMR analysis and peak identification by MS. Peak 1 is β-D-Man*p*O*p*NP as determined using a standard compound, peaks 2 and 3 were identified as *p*NP-Man$_2$ with MALDI-TOF MS (observed m/z: 486.10, theoretical [M+Na]$^+$ for *p*NP-Man$_2$: 486.12), peak 4 is *p*NP-Man$_3$ (observed m/z: 648.14, theoretical [M+Na]$^+$ for *p*NP-Man$_3$: 648.18) and peak 5 was identified as released *p*NP using a standard compound.

2

1.2 NMR analysis of the *Cf*Man2A-catalyzed transglycosylation reactions

a) NMR methods

*Cf*Man2A-N428T (34 µg·mL^{-1}) was incubated 2 h in 35 mM sodium phosphate pH 7.0 with 5 mM *p*-nitrophenyl-β-D-mannopyranoside (β-D-Man*p*O*p*NP) before removing 4 mL of the reaction mixture for NMR spectroscopic analysis. The sample was twice lyophilized and dissolved in 0.6 mL D$_2$O (99.9 atom % deuterium, Sigma-Aldrich).

NMR spectra were acquired at 318 K using an Agilent Varian VNMRS 500 MHz spectrometer equipped with a 5 mm HCN probe. This temperature was chosen to prevent interference from the resonance of residual HDO. The NMR chemical shifts were referenced with respect to the resonances of the anomeric CH-group (H-1) of α-D-mannose at 5.17 (^{1}H) and 96.8 ppm (^{13}C), these values from the NMR chemical shift database BioMagResBank[1] (www.bmrb.wisc.edu, entry no. bmse000018) were measured using a solution of α-D-mannose (0.5 mM) in 50 mM sodium phosphate buffer, equivalent to pH 7.4, in D$_2$O at 298 K, referenced to sodium trimethylsilylpropane-sulfonate (DSS). Standard 2D experiments (DQF-COSY, z-TOCSY with 120 ms mixing time, NOESY with 500 ms mixing time, ^{13}C-HSQC, and ^{13}C-HMBC, optimized for 8 Hz) were used for chemical shift assignments, using SPARKY[2] software. TOCSY and NOESY experiments were typically run using a spectral width of 4680 Hz in the indirect dimension, 512 increments, 32 scans, and a recycle delay of 2 s, resulting in roughly 22 h acquisition time. The 2D ^{13}C-HSQC experiment was acquired with improved resolution in the indirect dimension using non-uniform sampling (NUS) of 400 (41%) out of a total of 974 increments. The experiment used a recycle delay of 2 s, 32 scans, and an indirect spectral width of 12066 Hz, giving a total acquisition time of 16 h. The NUS sampled ^{13}C-HSQC experiment was reconstructed and processed using NMRPipe[3]. The $^1J_{\text{C-1,H-1}}$ scalar coupling constants were determined from the ^{13}C-satellites in a 1D ^{1}H spectrum.

An additional high resolution 1D ^{1}H spectrum at 318 K was obtained using a Bruker 500 MHz AVANCE III HD spectrometer equipped with a 5 mm BBOF probe. 1024 transients were acquired during 5 s and a recycle delay of 1 s. Using this spectrum, the $^3J_{\text{H-1,H-2}}$ scalar coupling constants were estimated from the doublet splitting of the anomeric H-1 signals in the spectrum, processed with a pure sine window function to obtain highest possible resolution. Estimates of the relative amounts of the different reaction products were obtained from integration of H-1 and resolved H-2 resonances in a 1D ^{1}H NMR spectrum.

b) NMR results

The *Cf*Man2A-N428T reaction mixture contained the substrate β-D-Man*p*O*p*NP, transglycosylation products (Fig. S1) and D-mannose that results from hydrolysis of the substrate. Identification of the transglycosylation products was performed using ^{1}H- and ^{13}C-NMR spectroscopy. The 1D ^{1}H spectrum of the sample revealed the presence of ten anomeric resonances, corresponding to ten D-Man*p* moieties (labelled A-I). Three of the D-Man units could be identified as α-D-Man*p*, β-D-Man*p*, and β-D-Man*p*^{B}O*p*NP, i.e. D-Man*p* unit B, from comparisons to ^{1}H- and ^{13}C- NMR spectra of the substrate β-D-Man*p*O*p*NP (NMR data acquired *in-house*), or using the NMR chemical shift database BioMagResBank[1], for α-, and β-D-Man*p* residues. Each of the spin systems with the remaining anomeric resonances were identified, and virtually complete ^{1}H and ^{13}C resonance assignments of the corresponding D-Man*p* residues were obtained (Table S1). The chemical shifts of the ^{1}H anomeric (H-1) resonances are consistent with expected chemical shifts for substituted β-D-Man*p* units[4,5].

3

96   Three of the anomeric protons (from D-Man*p* moieties A–C) show correlations in the ^{13}C-HMBC
97   spectrum to the *p*-nitrophenyl group, identifying substituted D-Man*p* residues. Information on the
98   connectivity between the different D-Man*p* residues were then obtained from elevated ^{13}C chemical
99   shift of the substituted carbon[6], direct observation of heteronuclear three-bond correlations over the
100   glycosidic linkage, and observation of NOE between protons close in space. The *p*NP-linked D-Man*p*
101   units A and C, together with the remaining five D-Man*p* units (labelled D–G & I) could be shown to
102   belong to D-Man*p* units of four reaction products, i.e. two dimannosides and two trimannosides,
103   identified as β-D-Man*p*[I]-(1→3)-β-D-Man*p*[C]O*p*NP (18%), β-D-Man*p*[E]-(1→4)-β-D-Man*p*[A]O*p*NP
104   (56%), β-D-Man*p*[F]-(1→4)-β-D-Man*p*[G]-(1→3)-β-D-Man*p*[C]O*p*NP (13%), and β-D-Man*p*[F]-(1→4)-β-
105   D-Man*p*[D]-(1→4)-β-D-Man*p*[A]O*p*NP (13%), with the approximate relative yields given in parenthesis.
106   Table S2 lists the specific NMR observations used for the identifications of these transglycosylation
107   products. D-Man*p* residues termed either A, C, or F above have virtually identical chemical shifts as
108   they have similar chemical environments, although they are each present in two different reaction
109   products (Table S1). Many of the 2D resonances from these D-Man*p* groups show a slight defor-
110   mation or skewness, indicating the presence of two signals. Distinct cross peaks and/or NOEs were
111   observed from D-Man*p* residues A, C, and F to their respective adjacent D-Man*p* residues in the
112   different reaction products (Table S2).

113   The analysis of the two known dimannosides β-D-Man*p*-(1→3)-β-D-Man*p*O*p*NP and β-D-Man*p*-
114   (1→4)-β-D-Man*p*O*p*NP is in good agreement with the literature (Table S1)[4], taking into account the
115   different experimental temperatures and referencing procedures. However, it is noteworthy that com-
116   paring the previously reported[4] values with each other, ^{13}C-shifts for β-D-Man*p*-(1→4)-β-D-
117   Man*p*O*p*NP are unusually small compared to their reported values for β-D-Man*p*-(1→3)-β-D-
118   Man*p*O*p*NP, with a difference of about 3 ppm, possibly due to different referencing within the same
119   study[4].

120   The anomeric configurations were in all cases determined to be of the form β-D-Man*p*, with an
121   axial H-1. The value of the $^1J_{C\text{-}1,H\text{-}1}$ scalar coupling constant has been reported to be a reliable indi-
122   cator of the anomeric configuration[7], and for D-Man*p* from groups A, C, and E the measured values
123   for $^1J_{C\text{-}1,H\text{-}1}$ were all ~162 Hz, in contrast to a value of ~170 Hz for an α-configuration (i.e. with an
124   equatorial H-1). The $^1J_{C\text{-}1,H\text{-}1}$ coupling constant could not be determined for the other D-mannosyl
125   groups due to low sensitivity or spectral overlap. The $^3J_{H\text{-}1,H\text{-}2}$ scalar coupling constants could be
126   determined for several of the D-Man*p* units, and were all ~1 Hz as expected for β-D-Man*p*-containing
127   oligosaccharides[8]. Spectral overlap prevented determination of the coupling constants for units F and
128   I. Additional confirmations for the β-configuration were obtained from strong NOE observed from
129   H-1 to H-2, H-3, and H-5 protons[9], for all D-Man*p* moieties. D-Man*p* units D and G, which both are
130   the internal unit in a trimannoside, have weaker H-1 to H-3 NOE.

131

132

133

134 **Table S1 |** ^{1}H and ^{13}C NMR chemical shifts (ppm) for identified dimannosides and trimannosides in D$_2$O at
135 318 K.

Carbon no.	1	2	3	4	5
β-D-Man*p*^{I}-(1→3)-β-D-Man*p*^{C}O*p*NP					
D-Man*p*^{C} – ^{1}H	5.47	4.43	4.06	3.84	3.64
D-Man*p*^{C} – ^{13}C	99.8	70.1	81.5^a	67.7	79.0
D-Man*p*^{I} – ^{1}H	4.88	4.10	3.68	3.61	3.40
D-Man*p*^{I} – ^{13}C	99.7	73.5	75.7	69.6	79.2
β-D-Man*p*^{E}-(1→4)-β-D-Man*p*^{A}O*p*NP					
D-Man*p*^{A} – ^{1}H	5.50	4.29	3.92	3.95	3.73
D-Man*p*^{A} – ^{13}C	99.9	72.5	74.0	79.0^a	77.9
D-Man*p*^{E} – ^{1}H	4.76	4.09	3.66	3.60	3.45
D-Man*p*^{E} – ^{13}C	102.9	73.3	75.6	69.4	79.2
β-D-Man*p*^{F}-(1→4)-β-D-Man*p*^{G}-(1→3)-β-D-Man*p*^{C}O*p*NP					
D-Man*p*^{C} – ^{1}H	5.47	4.43	4.06	3.84	3.64
D-Man*p*^{C} – ^{13}C	99.8	70.1	81.5^a	67.7	79.0
D-Man*p*^{G} – ^{1}H	4.90	4.15	3.84	3.85	3.53
D-Man*p*^{G} – ^{13}C	99.6	73.0	74.3	79.4^a	77.8
D-Man*p*^{F} – ^{1}H	4.73	4.06	3.65	3.58	3.44
D-Man*p*^{F} – ^{13}C	102.9	73.2	75.6	69.4	79.2
β-D-Man*p*^{F}-(1→4)-β-D-Man*p*^{D}-(1→4)-β-D-Man*p*^{A}O*p*NP					
D-Man*p*^{A} – ^{1}H	5.50	4.29	3.92	3.95	3.73
D-Man*p*^{A} – ^{13}C	99.9	72.5	74.0	79.0^a	77.9
D-Man*p*^{D} – ^{1}H	4.78	4.14	3.82	3.84	3.57
D-Man*p*^{D} – ^{13}C	102.9	72.7	74.2	79.2^a	77.8
D-Man*p*^{F} – ^{1}H	4.73	4.06	3.65	3.58	3.44
D-Man*p*^{F} – ^{13}C	102.9	73.2	75.6	69.4	79.2

136 ^{a}Elevated ^{13}C chemical shifts revealing site of substitutions are marked in red.

137

138

139

Table S2 | NMR information used for the identifications of transglycosylation products.

β-D-Manp^I-(1→3)-β-D-Manp^{C}Op$NP:

- ^{13}C chemical shift (δ = 81.5 ppm) for β-D-Manp^C C-3,
- Direct cross peak in ^{13}C-HMBC between β-D-Manp^I H-1 and β-D-Manp^C C-3,
- NOE between β-D-Manp^I H-1 and β-D-Manp^C H-2,
- NOE between β-D-Manp^I H-1 and β-D-Manp^C H-3.

β-D-Manp^E-(1→4)-β-D-Manp^{A}Op$NP:

- ^{13}C chemical shift (δ = 79.0 ppm) for β-D-Manp^A C-4,
- NOE between β-D-Manp^E H-1 and β-D-Manp^A H-4,
- NOE between β-D-Manp^E H-1 and β-D-Manp^A H-3,
- NOE between β-D-Manp^E H-1 and β-D-Manp^A H-6,

Partial overlap of β-D-Manp^A C-4 and β-D-Manp^E C-5 prevents 3-bond correlation from β-D-Manp^E H-1 to be unambiguously observed.

β-D-Manp^F-(1→4)-β-D-Manp^G-(1→3)-β-D-Manp^{C}Op$NP:

- ^{13}C chemical shift (δ = 81.5 ppm) for β-D-Manp^C C-3,
- ^{13}C chemical shift (δ = 79.4 ppm) for β-D-Manp^G C-4,
- Direct cross peak in ^{13}C-HMBC between β-D-Manp^G H-1 and β-D-Manp^C C-3,
- NOE between β-D-Manp^G H-1 and β-D-Manp^C H-2,
- NOE between β-D-Manp^G H-1 and β-D-Manp^C H-1,
- NOE between β-D-Manp^F H-1 and β-D-Manp^G H-4/3,
- NOE between β-D-Manp^F H-1 and β-D-Manp^D H-4/3,

Partial overlap of β-D-Manp^G C-4 and β-D-Manp^F C-5 prevents 3-bond correlation from β-D-Manp^F H-1 to be unambiguously observed.

β-D-Manp^F-(1→4)-β-D-Manp^D-(1→4)-β-D-Manp^{A}Op$NP:

- ^{13}C chemical shift (δ = 79.0 ppm) for β-D-Manp^A C-4,
- ^{13}C chemical shift (δ = 79.2 ppm) for β-D-Manp^D C-4,
- Direct cross peak in ^{13}C-HMBC between β-D-Manp^D H-1 and β-D-Manp^A C-4,
- NOE between β-D-Manp^F H-1 and β-D-Manp^D H-4/3
- NOE between β-D-Manp^D H-1 and β-D-Manp^A H-4/3

Partial overlap of β-D-Manp^D C-4 and β-D-Manp^F C-5 prevents 3-bond correlation from β-D-Manp^F H-1 to be unambiguously observed.

141

142

143

2) GH10, β-xylanase *Rm*Xyn10A_CM
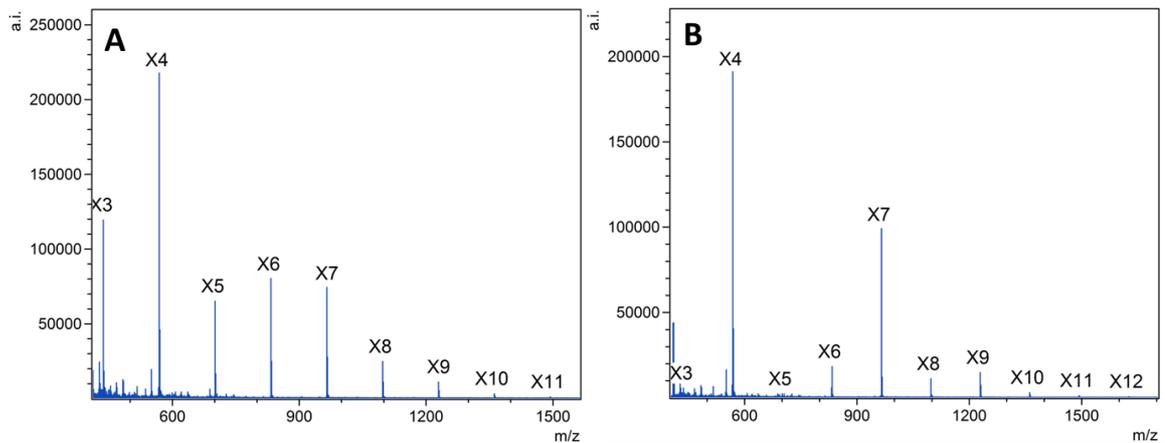
Figure S2 | Xylo-oligosaccharide analysis using mass spectrometry. MALDI-TOF-MS spectra of the sodium adducts of transglycosylation products generated using: **A.** *Rm*Xyn10A_CM-WT; **B.** *Rm*Xyn10A_CM-N118T. Reactions were run at 60°C in 20 mM sodium phosphate pH 7.0. Major transglycosylation product peaks were detected at m/z 833.4827, 965.5205 and 1097.5826 corresponding to the $[M+Na]^+$ ions of D-Xyl*p*-containing hexamer (X_6), heptamer (X_7) and octamer (X_8). Additional peaks were observed at m/z 1229.65, 1361.73 and 1493.83 corresponding to $[M+Na]^+$ ions of X_9, X_{10} and X_{11} (theoretical $[M+Na]^+$: 1229.38, 1361.42 and 1493.47).
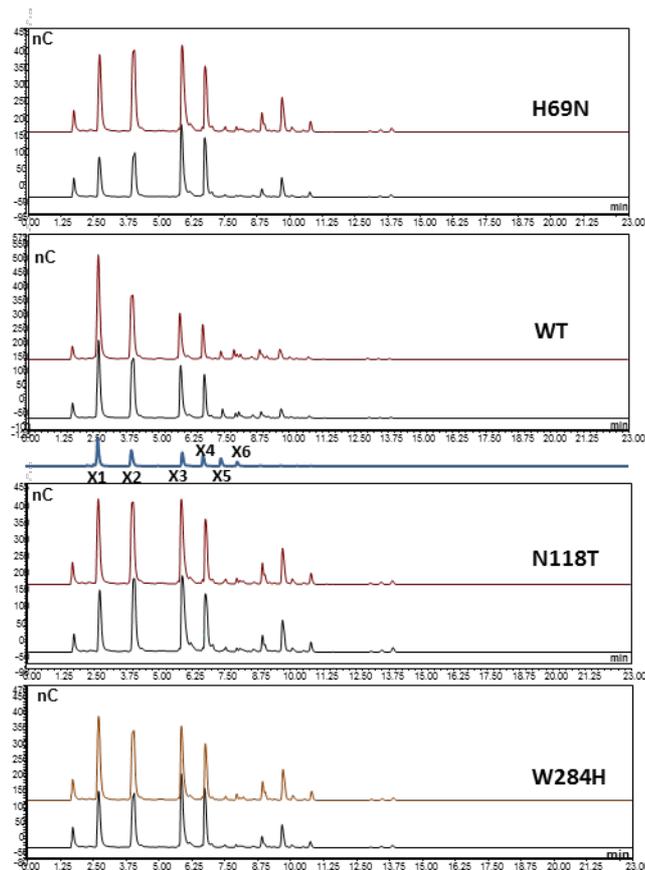
Figure S3 | HPAEC-PAD analysis of xylo-oligosaccharides. Chromatograms showing product profiles of transglycosylation reactions catalysed by *Rm*Xyn10A_CM-WT and its variants H69N, N118T and W284H. For each *Rm*Xyn10A_CM form, chromatograms display reaction profiles after 2 (black) and 4 h (brown), respectively. D-xylose (X_1) and xylo-oligosaccharides up to xylohexaose (X_6) were used as standards (blue) to facilitate product analysis.

159
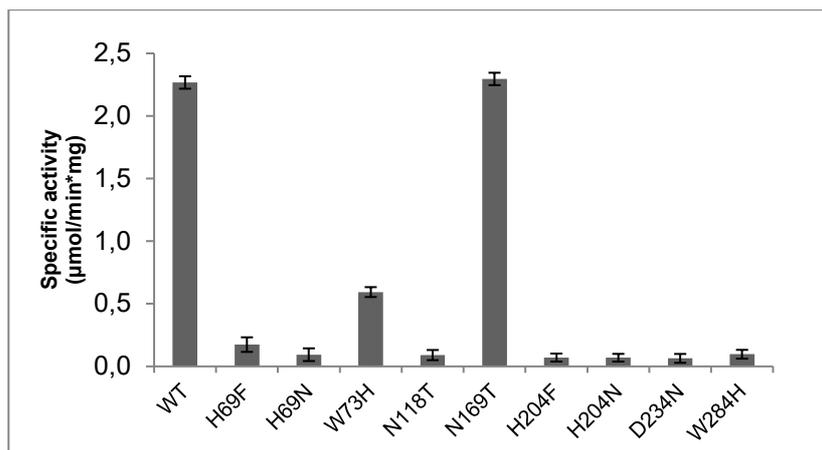
Figure S4 | Hydrolytic activity of *Rm*Xyn10A_CM forms. Hydrolytic activity of the *Rm*Xyn10A_CM
forms on 2 mM β-D-xylotriosideO*p*NP in 20 mM sodium phosphate pH 7.0, at 70℃.

162

163
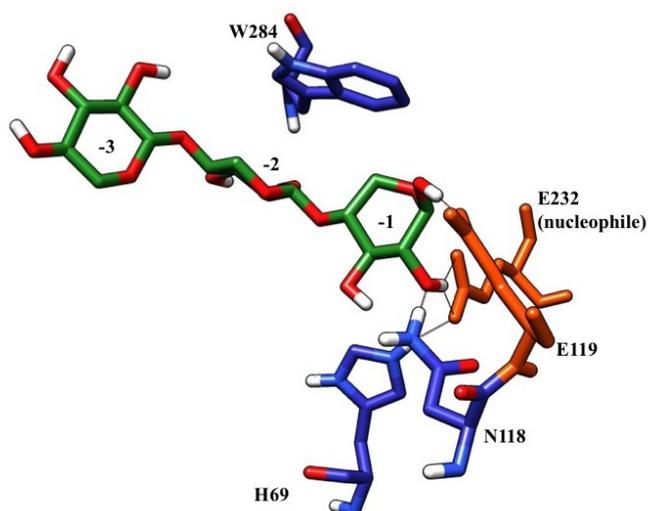
**Figure S5 | Illustration of the binding region of modelled *Rm*Xyn10A_CM showing the positions of
H69, N118 and W284.** Xylotriose (X3) coloured green is docked into subsites −3 to −1. Catalytic residues
are coloured orange and potential hydrogen bonds are shown as dotted black lines. This figure was prepared
by using UCSF Chimera.

168

169 **Table S3 |** Areas (nC·min) of peaks corresponding to products identified in reactions catalysed by
170 *Rm*Xyn10A_CM forms using 20 mM xylotetraose (X_4) as both donor and acceptor, in 20 mM sodium phos-
171 phate pH 7.0 at 65℃.

Time	1 h						2 h						4 h					
Peak	X_5	X_6	X_7	X_8	X_9	X_{10}	X_5	X_6	X_7	X_8	X_9	X_{10}	X_5	X_6	X_7	X_8	X_9	X_{10}
WT	3.0 ±0.7	2.7 ±0.5	0.7 ±1.0	0.9 ±1.3	0.4 ±1.1	0.1 ±1.4	3.0 ±0.8	1.2 ±0.4	2.6 ±1.0	3.8 ±2.0	0.3 ±1.2	0.6 ±1.0	3.1 ±0.4	2.4 ±0.5	4.7 ±0.6	4.9 ±2.1	0.5 ±1.2	0.6 ±1.0
H69F	0.1 ±1.4	0.3 ±2.1	0.7 ±1.7	1.4 ±0.4	0.3 ±1.6	0.2 ±1.1	0.4 ±1.6	0.3 ±1.1	2.7 ±1.2	6.8 ±1.1	0.6 ±1.0	1.5 ±1.2	0.6 ±1.1	0.6 ±1.3	4.3 ±1.1	8.4 ±2.3	0.8 ±1.3	1.6 ±0.5
H69N	0.1 ±1.1	0.3 ±1.1	0.7 ±1.0	1.5 ±0.1	0.3 ±1.2	0.3 ±1.3	0.6 ±1.2	0.4 ±1.3	2.5 ±0.6	6.4 ±2.7	0.6 ±0.6	1.5 ±1.0	1.1 ±1.3	0.8 ±1.3	5.9 ±1.1	11.6 ±1.9	0.9 ±1.0	1.8 ±0.4
W73H	0.4 ±1.6	0.7 ±1.2	0.8 ±1.1	1.7 ±0.5	0.3 ±2.5	0.3 ±1.1	0.9 ±2.5	0.5 ±1.1	2.8 ±0.8	6.6 ±1.2	0.6 ±0.8	1.5 ±1.0	1.2 ±1.1	0.6 ±1.5	4.2 ±1.2	8.3 ±2.6	0.8 ±1.4	1.6 ±0.6
N118T	0.1 ±1.0	0.4 ±2.0	0.6 ±1.0	1.5 ±1.3	0.3 ±1.4	0.3 ±2.9	1.2 ±1.4	0.9 ±2.9	6.5 ±1.1	13.4 ±2.2	0.8 ±0.6	2.0 ±0.9	1.3 ±1.0	1.0 ±1.4	6.8 ±1.5	12.1 ±3.3	1.0 ±1.5	1.7 ±1.4
N169T	2.2 ±1.4	2.6 ±1.1	0.7 ±1.2	1.4 ±3.0	0.2 ±1.0	0.2 ±1.6	3.0 ±1.0	1.9 ±1.6	2.9 ±0.9	5.5 ±1.5	0.4 ±1.1	1.2 ±0.9	2.7 ±0.9	1.1 ±1.0	2.7 ±1.8	4.8 ±1.9	0.6 ±1.2	1.2 ±1.0
H204F	0.1 ±1.3	0.1 ±1.6	0.5 ±1.5	1.3 ±1.1	0.2 ±1.1	0.2 ±1.4	0.3 ±1.1	0.4 ±1.4	2.5 ±1.0	5.8 ±1.8	0.6 ±1.3	1.4 ±1.0	0.6 ±1.3	0.7 ±1.1	4.8 ±0.9	8.9 ±2.4	0.9 ±1.4	1.5 ±1.2
H204N	0.1 ±1.0	0.1 ±1.2	0.7 ±1.2	1.5 ±0.3	0.1 ±1.3	0.2 ±1.2	0.4 ±1.4	0.4 ±1.2	2.4 ±0.9	6.0 ±0.5	0.6 ±1.1	1.3 ±0.9	0.5 ±1.2	0.6 ±1.3	3.7 ±0.9	6.9 ±2.1	0.8 ±1.7	1.4 ±0.6
D234N	0.1 ±1.1	0.4 ±2.5	0.9 ±1.1	1.7 ±1.0	0.2 ±1.3	0.3 ±1.2	0.5 ±1.2	0.3 ±1.6	2.7 ±0.8	5.7 ±0.4	0.6 ±1.1	1.5 ±0.8	0.7 ±1.2	0.6 ±1.1	3.4 ±1.0	6.6 ±1.5	0.8 ±1.5	1.5 ±1.1
W284H	0.1 ±1.3	0.3 ±1.3	0.9 ±1.3	1.7 ±1.7	0.2 ±1.3	0.2 ±1.2	0.6 ±1.1	0.5 ±1.0	3.5 ±1.1	7.4 ±1.1	0.7 ±1.1	1.5 ±1.1	0.9 ±1.2	1.0 ±1.1	6.2 ±0.8	10.7 ±1.5	1.0 ±1.5	1.7 ±1.6

172

173

174

9

3) GH29, α-fucosidase AlfB

176

177 Time course of disaccharide formation catalysed by AlfB-WT reveals a peculiar behaviour (Fig. 5).
178 Disaccharide formation is slow and once maximal disaccharide yield is reached; it appears to plunge
179 rapidly. Plotting *p*NP release rate versus donor substrate concentration ([α-L-Fuc*p*OCNP]) reveals
180 that this untypical behaviour can be partly explained by the fact that at high (>8 mM) concentration
181 the donor inhibits the WT enzyme (Fig. S6) and that at low donor concentration secondary hydrolysis
182 rapidly leads to the decomposition of the disaccharide (Fig. S6). In the case of AlfB-H80F, and most
183 other AlfB mutants, substrate inhibition appears to be abolished, and secondary hydrolysis is re-
184 duced.

Figure S6 | AlfB behavior as a function of α-L-Fuc*p*OCNP concentration. Left, AlfB-WT hydrolysis kinetic profile assessed at 25 °C in 50 mM sodium phosphate pH 7, at 20 ng·mL^{-1}. Right, transglycosylation product evolution as a function of donor conversion.
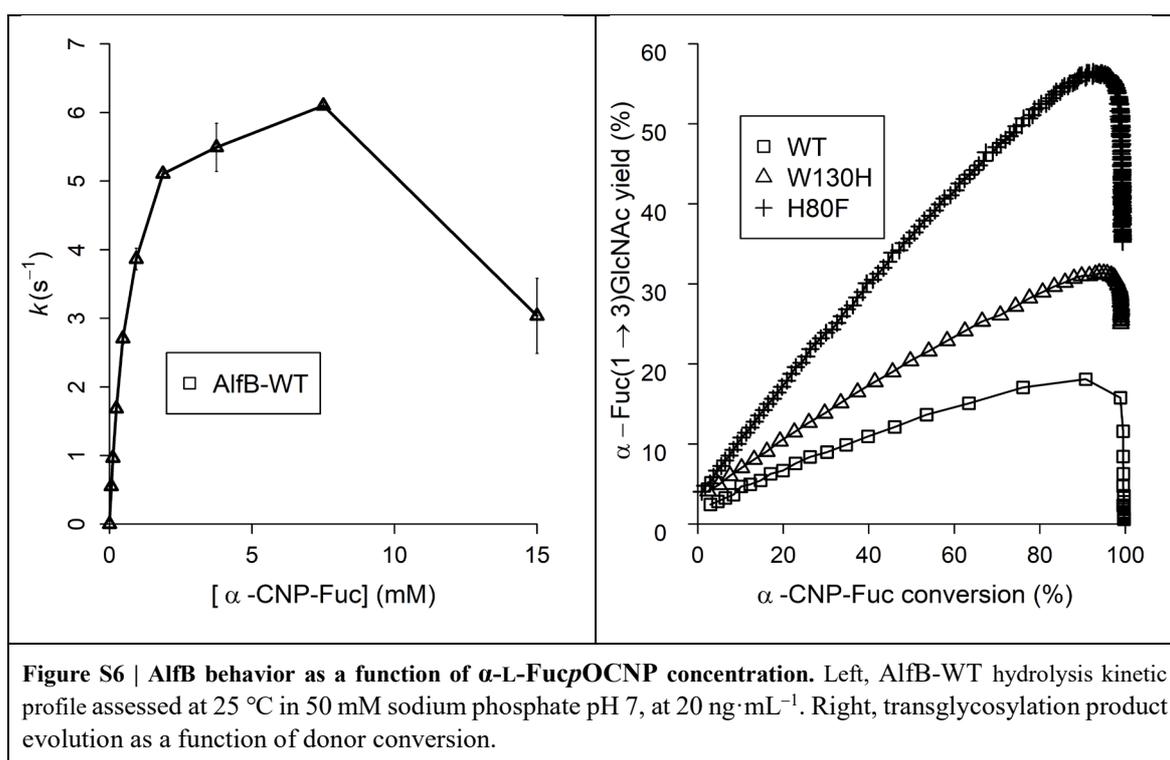
185

186 NMR spectroscopy was used to assess the transfucosylation capabilities of AlfB forms (see Methods
187 section in the main text). Not all anomeric signals could be followed due to the close vicinity with
188 the HOD peak and its suppression, but signals for the donor, the acceptor and the disaccharide
189 product could be monitored. Particularly, the latter was followed with 4 signals, one for both
190 anomers of the disaccharide, two for its α-anomer and one for the β-anomer (δ = 5.06, 4.93, 4.9
191 and 4.06 ppm, respectively; Fig. S6).

192
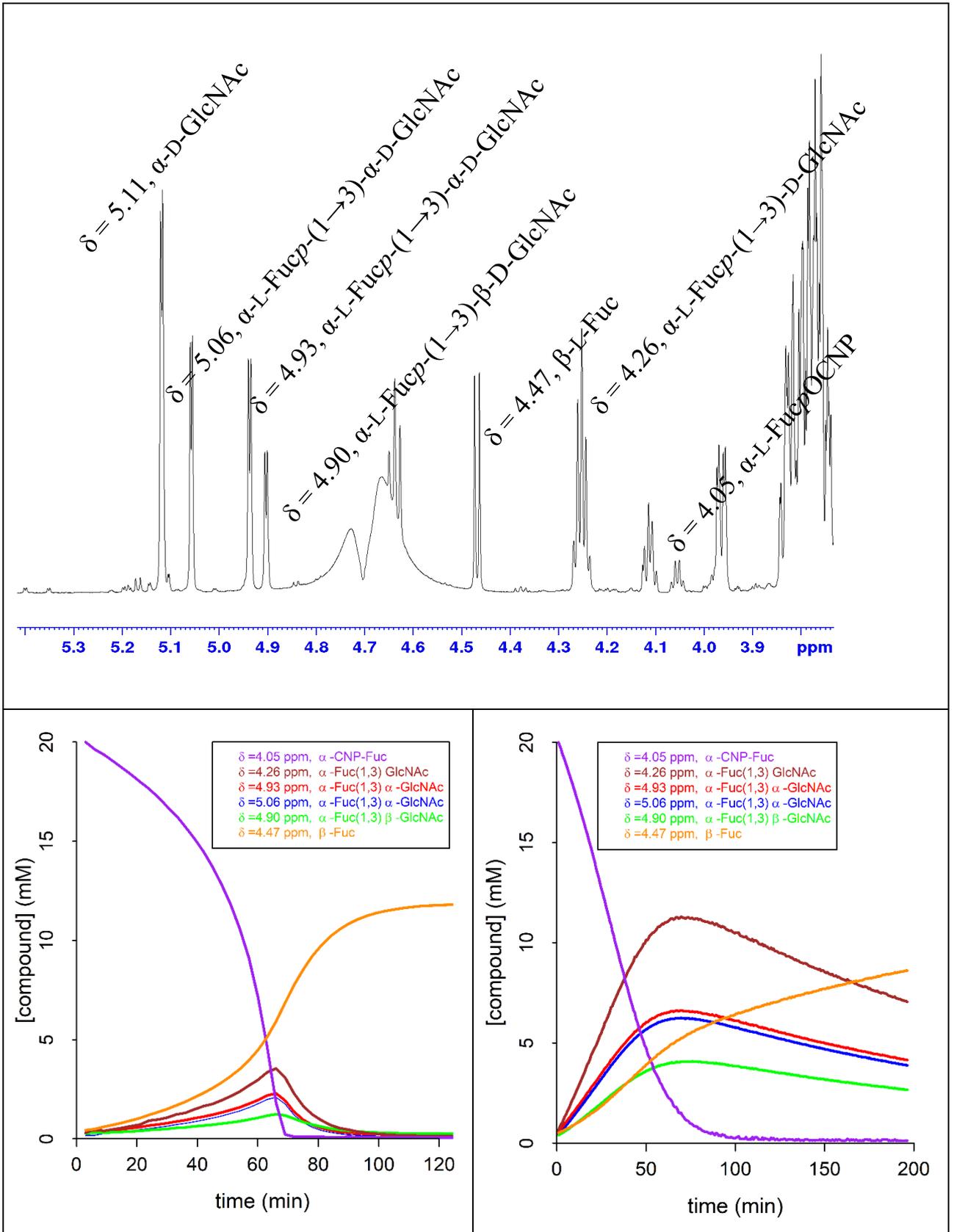
Figure S7 | NMR monitoring of transfucosylation by AlfB forms. Top, NMR signals used to monitor the reaction. Bottom, Time-course for AlfB-WT (left) and AlfB-H80F (right).

4) GH29, α-fucosidase AlfC

The feasibility of transposing a successful mutation to a related enzyme was assessed using the fucosidase AlfC from *Lactobacillus casei*. Signals from α-L-Fucp-(1→6)-D-GlcNAc were insufficiently resolved from those of α-L-FucpOCNP, L-Fuc and D-GlcNAc in the H-2 and anomeric region to allow accurate monitoring of yield (Fig. S7). Thus, compound formation and disappearance were monitored using the methyl group of L-Fuc (Fig. S8). It is important to note that although progress curves established using the anomeric region were less precise, they were nevertheless consistent with those obtained when monitoring the methyl group (< 2% deviation for maximum yield assessment).
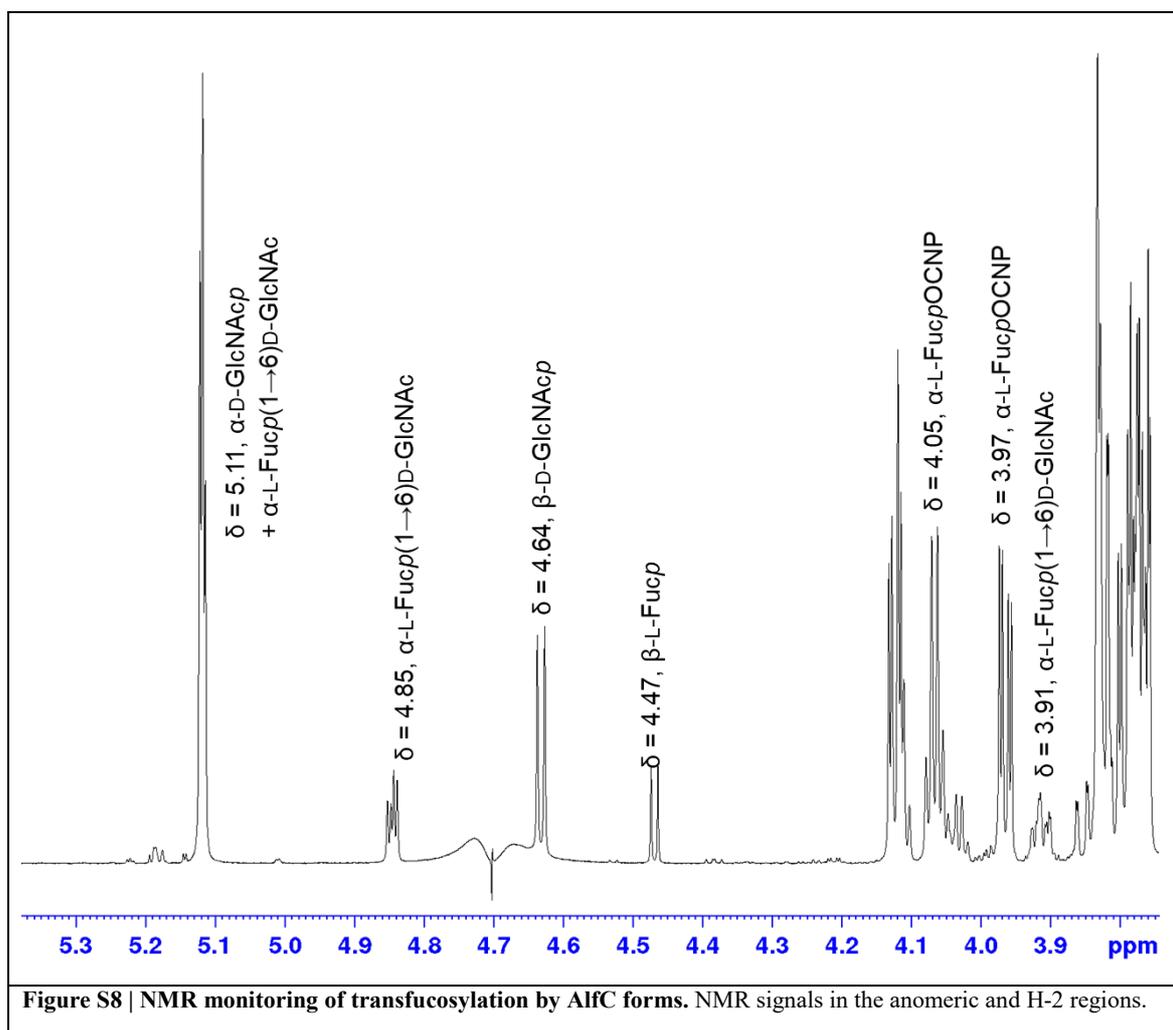
Figure S8 | NMR monitoring of transfucosylation by AlfC forms. NMR signals in the anomeric and H-2 regions.

Figure S9 | NMR monitoring of transfucosylation by AlfC forms. Top left, NMR signals used to monitor the reaction (blue, after 8 min; red, 17 min; green, 53 min). Time-course for AlfC-WT (top right), AlfC-W138H (bottom left) and AlfC-H87F (bottom right).

208

209

13
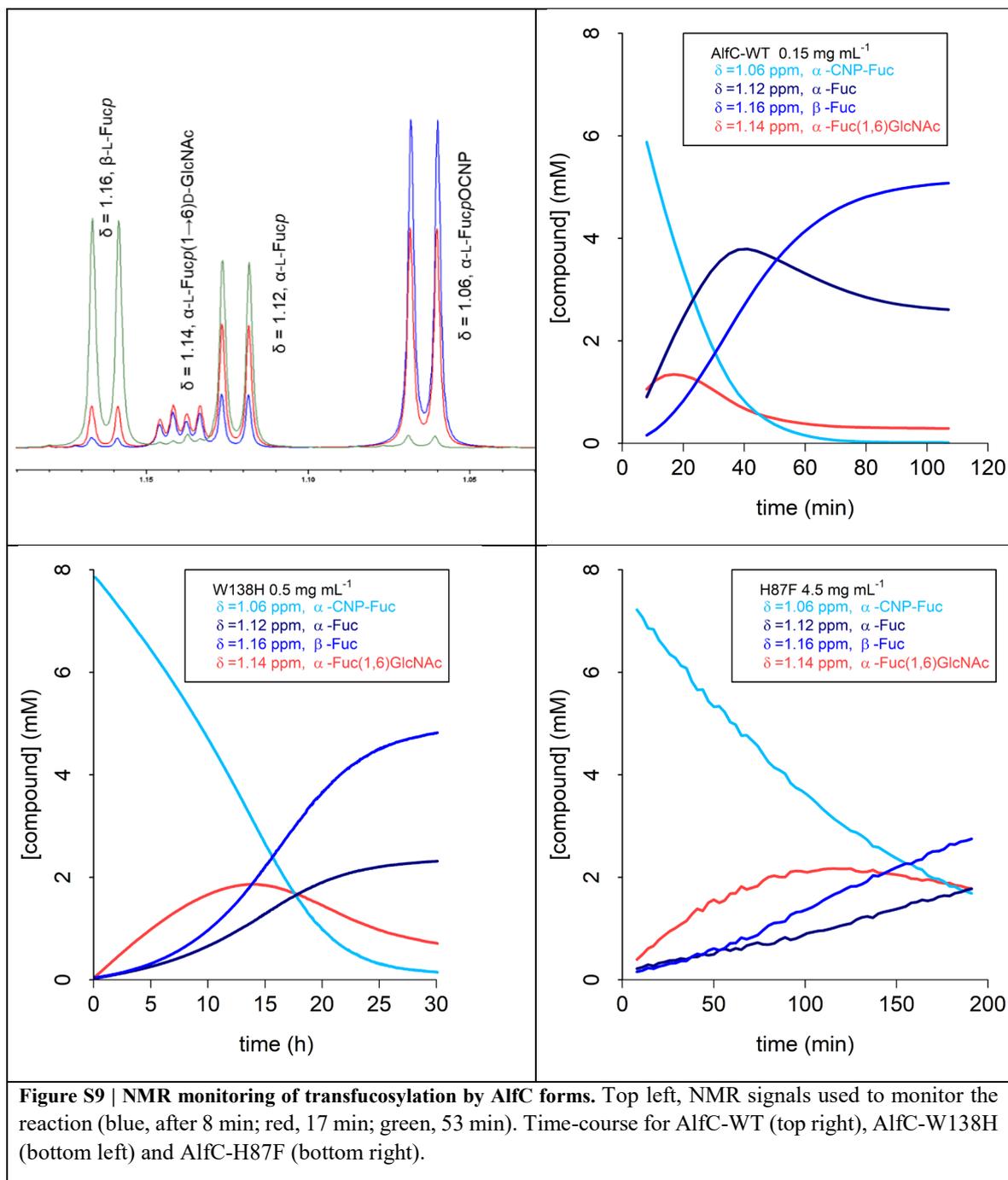
5) GH51, α-L-arabinofuranosidase *Tx*Abf

Table S4 | Maximum yields of arabinoxylo-tetrasaccharides synthetized by *Tx*Abf forms

*Tx*Abf form	Yield (%)			
	XA^{3}X[a]	A^{3}XX[a]	A^{2}XX (+ XA^{2}X)[a]	Total AXOS[b]
	5.40 ppm[c]	5.32 ppm[c]	5.29-5.28 ppm[c]	
WT	2	4	4	9
F26H	8	5	11	22
F26L[d]	6	9	11	26
E28Q	5	8	10	23
R69K	10	13	16	37
R69H[d]	15	12	23	46
N175T	12	12	22	42
H240F	3	3	76	82
H240N	16	14	59	75
Y242F	4	4	12	19
D297N	12	15	23	47
N216W-H240F[d,e]	-	-	62[e]	62[e]
R69H-N216W-L352M[d,e]	-	-	70[e]	70[e]

[a]See Fauré et al. for a comprehensive description of AXOS nomenclature[10]. [b]Maximum yield of each AXOS was reached at different times explaining why total AXOS yield cannot be obtained by summing the individual maximum yields of all products. [c]NMR chemical shift of the anomeric proton of α-L-Ara$f$ unit of AXOS reported in the literature at approximatively 25°C[11]. Displacement of ^{1}H chemical shifts for the α-L-Ara$f$ anomeric signal of each AXOS towards blinded region (Δδ = -0.08 ppm) is observed at 45°C. [d] Single mutants F26L[12] and R69H[11] were randomly generated and the triple mutant R69H-N216W-L352M[11] was obtained by recombination. [e]N216W procures highly regioselective synthesis of the A^{2}XX tetrasaccharide[11].
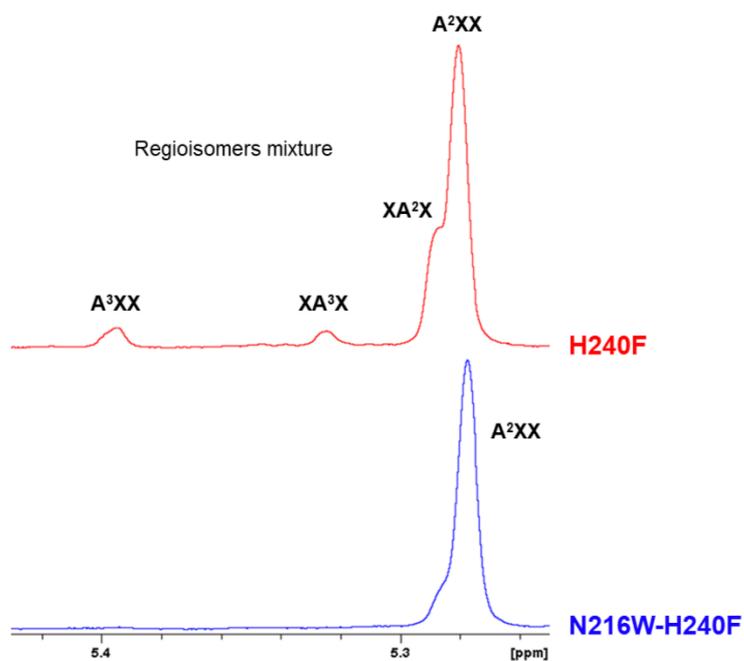
Figure S10 | NMR monitoring of transarabinofuranosylation by *Tx*Abf variants. ^{1}H NMR anomeric signals of α-L-Ara*f*-containing regioisomer(s) for H240F and regioselective N216W-H240F.

226 **Table S5 |** Specific activity (SA, in IU·mg^{-1}) of TxAbf forms determined by pNP release.

Enzyme	SA$_H$[a]	SA$_T$[b]	R = SA$_T$/SA$_H$
WT	261.79 ± 10.72	125.49 ± 1.27	0.5
F26H	0.02 ± 0.001	0.02 ± 0.001	1.0
F26L[c]	105.76 ± 2.72	83.26 ± 2.44	0.8
E28Q	52.95 ± 2.09	32.95 ± 0.44	0.6
R69K	14.12 ± 0.08	9.78 ± 0.36	0.7
R69H[c]	2.84 ± 0.07	7.41 ± 0.23	2.6
N175T	0.15 ± 0.003	0.14 ± 0.01	1.0
H240F	3.21 ± 0.09	3.66 ± 0.09	1.2
H240N	0.05 ± 0.001	0.19 ± 0.01	4.2
Y242F	0.07 ± 0.004	0.07 ± 0.01	1.0
D297N	7.71 ± 0.34	12.61 ± 0.22	1.7
N216W-H240F	5.49 ± 0.10	8.19 ± 0.21	1.5
R69H-N216W-L352M[c]	0.97 ± 0.11	1.82 ± 0.07	1.9

227 [a]SA$_H$ was achieved in hydrolysis mode with 5 mM α-L-Ara$f$OpNP only. [b]SA$_T$ was achieved in transglyco-
228 sylation mode with 5 mM α-L-Ara$f$OpNP and in the presence of 10 mM xylotriose. [c]Single mutants F26L[12]
229 and R69H[11] were randomly generated and the triple mutant R69H-N216W-L352M[11] was obtained through
230 recombination.

231
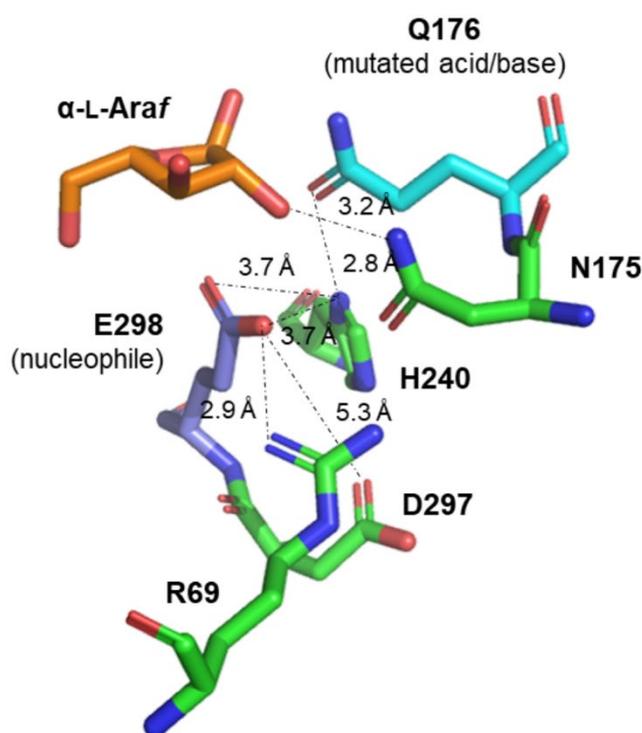
Figure S11 | Location of R69, N175, H240 and D297 within *Tx*Abf-E176Q subsite −1 (PDB ID: 2VRQ). The figure was prepared using PyMol Molecular Graphics System, v0.99 (Schrödinger).

Among the eight positions targeted by the conserved-sequence approach, the four yielding the best variants are spatially (within approximately 5 Å) and/or sequentially close to the catalytic residues 176 (acid/base, in cyan) and E298 (nucleophile, in deep blue). Accordingly, R69K, N175T, H240F/N and D297N (in green) are thought to impact the local H-bonding network and thus the pK_a cycling that occurs during catalysis.[13] Previously, the study of R69H revealed that R69 plays a vital role in the modulation of the ionization state and nucleophilic strength of E298.[11] N175 is thought to be involved in transition state stabilization. Therefore, modification of the interaction N175···OH-2 of α-L-Ara*f* unit (in orange), for example in N175T, might perturb the correct functioning of the two-step catalytic mechanism.[14] Additionally, H240 could be involved in a putative water channel affecting water dynamics.[15–17] Overall, the different mechanistic consequences of mutations at these four positions all translate into reduced water-mediated deglycosylation and/or increased lifetime of the covalent glycosyl-enzyme intermediate, which in turn favours sugar-mediated deglycosylation, hence transglycosylation. Therefore, all five single-mutants achieve the sought after result, namely tipping the T/H balance in favour of transglycosylation.

252 References

253

254   1.   Ulrich, E. L. *et al.* BioMagResBank. *Nucleic Acids Res.* **36**, 402–408 (2008).

255   2.   Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: Enhanced software for
256   biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

257   3.   Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on
258   UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

259   4.   Eneyskaya, E. V. *et al.* Transglycosylating and hydrolytic activities of the β-mannosidase
260   from *Trichoderma reesei*. *Biochimie* **91**, 632–638 (2009).

261   5.   Rosengren, A. *et al.* Enzymatic synthesis and polymerisation of β-mannosyl acrylates
262   produced from renewable hemicellulosic glycans. *Green Chem.* **21**, 2104–2118 (2019).

263   6.   Jansson, P.-E., Kenne, L. & Schweda, E. Nuclear magnetic resonance and conformational
264   studies on monoacetylated methyl D-gluco- and D-galacto-pyranosides. *J. Chem. Soc.,*
265   *Trans.* **53**, 1689–1699 (1987).

266   7.   Bubb, W. A. NMR spectroscopy in the study of carbohydrates: Characterizing the structural
267   complexity. *Concepts Magn. Reson. Part A Bridg. Educ. Res.* **19**, 1–19 (2003).

268   8.   Harjunpää, V., Teleman, A., Siika-Aho, M. & Drakenberg, T. Kinetic and stereochemical
269   studies of manno-oligosaccharide hydrolysis catalysed by β-mannanases from *Trichoderma*
270   *reesei*. *Eur. J. Biochem.* **234**, 278–283 (1995).

271   9.   Agrawal, P. K. NMR Spectroscopy in the structural elucidation of oligosaccharides and
272   glycosides. *Phytochemistry* **31**, 3307–3330 (1992).

273   10.   Fauré, R. *et al.* A brief and informationally rich naming system for oligosaccharide motifs
274   of heteroxylans found in plant cell walls. *Aust. J. Chem.* **62**, 533–537 (2009).

275   11.   Bissaro, B. *et al.* Molecular design of non-Leloir furanose-transferring enzymes from an α-
276   L-arabinofuranosidase: a rationale for the engineering of evolved transglycosylases. *ACS*
277   *Catal.* **5**, 4598–4611 (2015).

278   12.   Arab-Jaziri, F. *et al.* Engineering transglycosidase activity into a GH51 α-L-
279   arabinofuranosidase. *New Biotechnol.* **30**, 536–544 (2013).

280   13.   McIntosh, L. P. *et al.* The pK_a of the general acid/base carboxyl group of a glycosidase
281   cycles during catalysis: A ^{13}C-NMR study of *Bacillus circulans* xylanase. *Biochemistry* **35**,
282   9958–9966 (1996).

283   14.   Raich, L. *et al.* A trapped covalent intermediate of a glycoside hydrolase on the pathway to
284   transglycosylation . insights from experiments and quantum mechanics / molecular

285   mechanics simulations. *J. Am. Chem. Soc.* **138**, 3325–3332 (2016).

286 15. Teze, D. *et al.* Conserved water molecules in family 1 glycosidases: A DXMS and
287   molecular dynamics study. *Biochemistry* **52**, 5900–5910 (2013).

288 16. David, B., Arnaud, P., Tellier, C. & Sanejouand, Y. Towards the design of efficient
289   transglycosidases : the case of the GH1 of *Thermus thermophilus*. *Protein Eng. Des. Sel.* 1–
290   8 (2019).

291 17. David, B. *et al.* Internal water dynamics control the transglycosylation/hydrolysis balance in
292   the agarase (AgaD) of *Zobellia galactanivorans*. *ACS Catal.* **7**, 3357–3367 (2017).

293