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

David Teze, Jiao Zhao, Mathias Wiemann, Kazi Zubaida Gulshan Ara, Rossana Lupo, Birgitte Zeuner, Marlène Vuillemin, Mette Errebo Rønne, Göran Carlström, Jens Duus, et al.

HAL Id: hal-02906907
https://hal.inrae.fr/hal-02906907v2

Submitted on 25 Jun 2021

Chemistry
A European
Journal

Accepted Article

Title: Rational Enzyme Design without Structural Knowledge:
A Sequence-Based Approach for Efficient Generation of
Transglycosylases

Authors: David Teze, Jiao Zhao, Mathias Wiemann, Kazi Z. G. Ara,
Rossana Lupo, Birgitte Zeuner, Marlène Vuillemin, Mette
E. Rønne, Göran Carlström, Jens Ø. Duus, Yves-Henri
Sanejouand, Michael J. O'Donohue, Eva Nordberg Karlsson,
Régis Fauré, Henrik Stålbrand, and Birte Svensson

This manuscript has been accepted after peer review and appears as an
Accepted Article online prior to editing, proofing, and formal publication
of the final Version of Record (VoR). This work is currently citable by
using the Digital Object Identifier (DOI) given below. The VoR will be
published online in Early View as soon as possible and may be different
to this Accepted Article as a result of editing. Readers should obtain
the VoR from the journal website shown below when it is published
to ensure accuracy of information. The authors are responsible for the
content of this Accepted Article.

WILEY-VCH

RESEARCH ARTICLE

Rational Enzyme Design without Structural Knowledge: A Sequence-Based Approach for Efficient Generation of Transglycosylases

David Teze,* Jiao Zhao, Mathias Wiemann, Zubaida G. A. Kazi, Rossana Lupo, Birgitte Zeuner, Marlène Vuillemin, Mette E. Rønne, Göran Carlström, Jens Ø. Duus, Yves-Henri Sanejouand, Michael J. O'Donohue, Eva Nordberg Karlsson, Régis Fauré, Henrik Stålbrand and Birte Svensson*

[*]   Dr. David Teze, Rossana Lupo, Dr. Birgitte Zeuner, Dr. Marlène Vuillemin, Mette E. Rønne, Prof. Birte Svensson
        Department of Biotechnology and Biomedicine
        Technical University of Denmark
        Kongens Lyngby, Denmark
        E-mail: datez@dtu.dk or david.teze@gmail.com and bis@bio.dtu.dk
        Dr. Jiao Zhao, Dr. Michael J. O'Donohue, Dr. Régis Fauré
        Toulouse Biotechnology Institute
        Université de Toulouse, CNRS, INRAE, INSA
        Toulouse, France
        Mathias Wiemann, Prof. Henrik Stålbrand
        Department of Biochemistry and Structural Biology
        Lund University
        Lund, Sweden.
        Dr. Kazi Z G Ara, Dr. Göran Carlström, Prof. Eva Nordberg Karlsson
        Department of Chemistry
        Lund University
        Lund, Sweden.
        Prof. Jens Ø. Duus
        Department of Chemistry
        Technical University of Denmark
        Kongens Lyngby, Denmark
        Dr. Yves-Henri Sanejouand
        UFIP
        Université de Nantes, CNRS
        Nantes, France

        Supporting information for this article is given via a link at the end of the document.

Abstract: 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 is useful to remedy this, but it usually requires prior structural knowledge of the target enzyme, and/or relies on extensive, time-consuming screening and analysis. Here we describe a straightforward strategy that involves rational rapid *in silico* analysis of protein sequences. The method pinpoints 6–12 single mutant candidates to improve transglycosylation yields. Requiring very little prior knowledge of the target enzyme other than its sequence, the method is generic and procures catalysts for the formation of glycosidic bonds involving various D/L-, α/β-pyranosides or furanosides, and *exo-* and *endo-* action. Moreover, mutations validated in one enzyme can be transposed to others, even distantly related enzymes.

Introduction

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 structural complexity. Carbohydrates consist 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 difficult to handle *in vitro* and often require expensive nucleotide-glycoside donors.[7] Consequently, glycoside hydrolases (GHs) are a viable alternative for enzyme-catalysed glycosynthesis. The GHs are particularly abundant and their diversity in terms of bond breaking ability matches carbohydrate complexity itself. GH catalysed hydrolysis of glycosidic bonds uses a variety of mechanisms, but

RESEARCH ARTICLE

the majority of GHs perform two subsequent displacements at the anomeric carbon, thus yielding products whose anomeric configuration is identical to that of the substrate.[8,9] GHs operating by such a mechanism are termed "retaining" GHs and represent at least 66% of all GHs (888 506 classified modules) in the CAZy database, grouped in 83 out of its current 161 GH families (as of January 7th, 2021). 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 of them display significant levels of transglycosylation. This reaction is under kinetic control and modulated by a number of factors related to the reaction conditions,[11,12] including acceptor concentration, water activity, substrate activation, temperature, pH and, most importantly, the enzyme properties. Accordingly, engineering of GHs has proven to be a potent way to obtain transglycosylases.[13]

The most generic GH engineering approach described to date is the so-called glycosynthase strategy,[14–17] in which a catalytic carboxylate is replaced by a catalytically impotent group. 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-glycoside specific wild-type (WT) enzyme) and turns the enzyme into an inverting glycosynthase (Scheme S1). Albeit powerful, this method relies on the availability of a suitably reactive and sufficiently stable substrate. Moreover, glycosynthases are intrinsically impotent biocatalysts that display extremely low activity, thus large quantities of enzymes are needed.[14–18]

An alternative strategy is to convert GHs into efficient transglycosylases, 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 mutant libraries, introducing location-agnostic modifications, coupled with powerful phenotypic screens (random mutagenesis).[19] Nevertheless, by simply targeting a small number of conserved active-site residues in several retaining GHs,[20–22] we previously demonstrated that transglycosylation capability can be improved without extensive screening, using structural information coupled with sequence conservation analysis. Moreover, using information on beneficial mutations, others have successfully transposed our results to related GHs, leading to significant transglycosylation yields.[23]

Herein we demonstrate how a refined sequenced-based protein engineering approach can be used to enhance the transglycosylation capability in a wide variety of predominantly hydrolytic GHs. This strategy involves identifying and replacing a retaining GH's most conserved residues by structural analogues (e.g., Tyr by Phe, Asp by Asn).

Barring rare exceptions,[10] retaining GHs follow a double-displacement mechanism in which two consecutive reactions are catalysed. The first is a nucleophilic attack leading to donor-glycan (or aglycone) bond breakage and the formation of an intermediate (glycosylation) of opposite anomeric configuration compared to substrates and final products. The second step is disruption of the intermediate (deglycosylation) by a water molecule (hydrolysis) or another acceptor (transglycosylation). Thus, transglycolytaing retaining GHs are stabilising the transition states (TS‡) of four reactions: glycosylation using the initial donor (G_D) or the transglycosylation product (G_P), hydrolysis (H), and transglycosylation (T) (Scheme S1). Transglycosylation yields are primarily restricted by two rate ratio: the transglycosylation over hydrolysis ratio (T/H), and the glycosylation rates with the initial donor or the formed transglycosylation products (G_D/G_P ratio, governing secondary hydrolysis). Therefore, reducing the stabilisation of the hydrolysis TS‡ relative to that of transglycosylation, even by as little as half a kcal·mol^{-1}, drastically modifies the T/H ratio and increases synthetic yields. Hence, we hypothesize that replacing residues presumably involved in TS‡ stabilisation is likely to generate enzyme variants displaying modified T/H and G_D/G_P ratios, among which will be improved transglycosylases. Assuming that evolutionary pressure on retaining GHs mainly affects residues involved in either protein folding or stabilisation of the glycosylation TS‡ and/or the hydrolysis TS‡, herein to improve transglycosylation yield we targeted conserved residues, excluding glycines and prolines that are most likely to be essential to enzyme structural integrity. Our hypothesis is that the substitution of these residues will potentially yield mutants that display greater ability to catalyse transglycosylation and that these point mutations will be transferable to other related GHs. Advantageously, the method described herein requires neither structural knowledge nor in-depth mechanistic understanding of the target GH. Nevertheless, when the target GH is characterized, the catalytic residues and those that are distant (> 10 Å) from the −1 subsite[24] can be conveniently excluded. Using this strategy, we demonstrate how the transglycosylation yields of GHs from families 2, 10, 20, 29, 51 and 151 are significantly enhanced using an approach that is relatively straightforward and time-efficient, necessitating the analysis of 6–12 mutants per enzyme, applicable to a broad variety of glycosidic bond-forming reactions and transferrable, allowing the portability of successful mutations to related enzymes.

Results and Discussion

Multiple sequence alignment and residue conservation

The method described herein relies on identifying a small number of conserved candidate amino acid 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 approximately 15 residues that are significantly more conserved compared to others within a query GH sequence. The method is calibrated to avoid the selection of more residues, because increasing the number of targets would increase the likelihood of falling outside of the active site region and thus targeting residues less likely to be involved in TS‡ stabilisation. On the other hand, selecting fewer residues would provide an insufficient number of targets for mutation. However, both the number and the diversity of the retrieved sequences are highly dependent on the query sequence. Therefore, the sample size of the batch of sequences in which approximately 15 residues are significantly more conserved must be adequately large. For this approach to be successful, a ranking method is also required

RESEARCH ARTICLE

that is robust, even in extreme cases where the sequence batch is either highly diverse, meaning that the overall stringency of conservation is low, or alternatively where the sequences are highly homogenous.

Our three-step methodology begins with the collection of a large number of sequences within the GH family of interest. These are then clustered to reduce redundancy.[25,26] After clustering, iterative multiple sequence alignments (MSA) are performed to select sequences that share at least 5% sequence identity. This minimal sequence identity (termed "ID% threshold") is then iteratively increased and the residues are ranked by decreasing conservation. Their conservation is 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 then analysed to identify the residues hereafter designated as "conserved". This iterative MSA and conservation analysis are performed using ClustalΩ[27] and an in-house script (available at https://hal.archives-ouvertes.fr/hal-03076670, see Methods).

To test the robustness of this methodology, GHs from families 2, 10, 20, 29, 51 and 151 were selected, with results being summarised in Table 1. Comparing GH10 and GH51 illustrates that similar sequence pool sizes and identical stringency criteria can generate different heterogeneity levels (21 vs 35% average identity within the kept sequences – termed "mean ID%"), while the examples of GH2 and GH20 reveal that similar heterogeneity can be obtained from a variable number of

selected sequences (9 301 vs 585 retained from 20 000 initially retrieved sequences, Table 1). Overall, when applied to the six GH families the methodology identified 9–16 significantly more conserved residues in the query sequences. Afterwards, mutagenesis was performed, designing experiments to introduce substitutions that preserve as much as possible the residue's size, polarity and aromaticity (Table 1). This is important because more radical changes are likely to engender large perturbations, leading to very strong TS‡ destabilization and loss of activity. Any conserved Gly and Pro residues were omitted from mutagenic analysis, because we assumed that these generally play important roles in protein folding rather than catalysis. Similarly, we avoided to substitute Asn and Gln by Asp and Glu respectively, because the catalytic residues in GHs are very often carboxylates themselves. Therefore, this avoided the inadvertent introduction of additional carboxylic moieties in the immediate vicinity of the catalytic centre. In all cases the methodology yielded 6–12 candidates for mutagenesis and subsequent analysis of transglycosylation capability. Reaction yields were calculated at the time point showing maximal product concentration, and solely for reactions that reached completion, meaning that the donor compound was fully consumed. Analyses were performed using enzyme solutions whose concentrations were different, but which represent similar catalytic activities. Note that in contrast with the glycosynthase approach, transient maximal yields obtained via transglycosylation are largely independent of enzyme loads (Scheme S1).
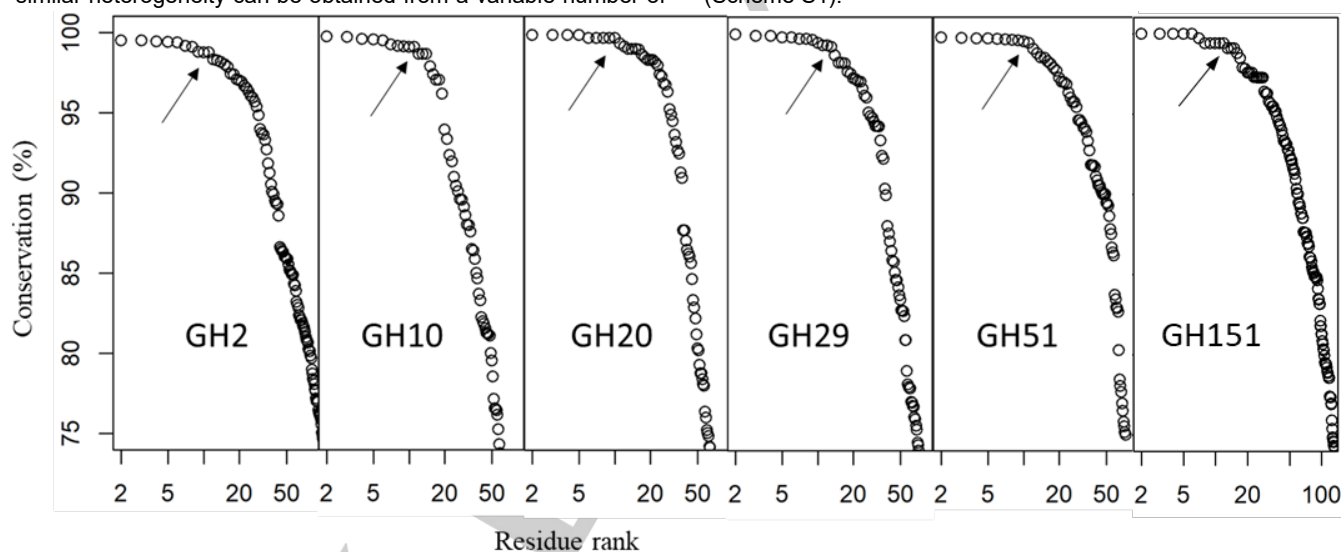
Figure 1. Residue conservation analysis. Residues in the MSA are ranked according to decreasing conservation, and their conservation is plotted as function of the logarithm of their rank. The plots are obtained from the analysis of the final MSA on GH families 2, 10, 20, 29, 51 and 151 respectively (for Mean ID% see Table 1). Arrows indicate the last rank considered as "significantly more conserved" than the rest of the residues in one sequence.

Table 1. Multiple sequence alignment and residue conservation analysis

GH family	Enzyme target	Species	Retrieved sequences	ID%[a] threshold	Sequences kept	Mean ID%[a,b]	Conserved residues	Evaluated candidates[c]
2	CfMan2A	C. fimi	20 000	12	9 301	22.6	14	12
10	Xyn10A	R. marinus	9 288	15	4 431	21.1	11	9
20	BbHI	B. bifidum	20 000	10	585	22.5	9	6
29	AlfB	L. casei	15 328	20	1 194	24.0	13	7
51	TxAbf	T. xylanilyticus	11 553	15	3 272	34.6	11	8
151	PtGH151	P. thiaminolyticus	1 854	20	459	40.6	16	12

Native AA	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
substitution[d]	G	S	N	Q	H	n. a.	F,Q,N	L	Q	I	L	T	n. a.	N	K	A	S	I	H	F

[a] ID%: percentage of identity between sequences. [b] Average identity with the reference sequence (enzyme target). [c] Gly, Pro and identified catalytic residues are not mutated. [d] Substitution table used in this study. n. a.: not applicable.

3

RESEARCH ARTICLE

GH29, α-L-fucosidases AlfB and AlfC

L-Fucose is the most common L-sugar in animals,[28] invariably connected through an axial glycosidic linkage. Accordingly, a $^1C_4 \rightarrow [^3H_4]^{\ddagger} \rightarrow {}^3S_1$ conformational itinerary is followed in enzymatic hydrolysis.[29] A few studies have been undertaken to improve GH29-mediated transfucosylation with varying success.[30–32] For the present study, we focused on AlfB from *Lactobacillus casei*. This enzyme is reported to synthesize α-L-Fucp-(1→3)-D-GlcNAc,[33,34] the α-L-Fucp-(1→3) motif being particularly common with 8 out of 13 known human fucosyltransferases being 3-fucosyltransferases.[28] AlfB-WT and seven of its mutants (collectively "AlfB forms") were evaluated for ability to synthesise α-L-Fucp-(1→3)-D-GlcNAc from equimolar amounts 20 mM 2-chloro-4-nitrophenyl α-L-fucopyranoside (α-L-Fucp OCNP) and 20 mM D-GlcNAc (Table 2, Figs. 2 and S2).

Table 2. GH29 engineering.

GH29s forms	Yield (%)	Ratio T/H	Enzyme (µg·mL⁻¹, µM)
AlfB-WT	18	0.29	65 (1.4)
H21F	–	–	–
Y45F	21	0.36	20 (0.4)
H80F	57	1.89	250 (5.2)
H81F	1.3	0.01	465 (10)
Y124F	15	0.25	740 (16)
D129N	22	0.43	117 (2.5)
W130H	31	0.53	750 (16)
AlfC-WT	17	1.00	150 (1.0)
H87F	26	1.47	4500 (30)
W137H	22	1.93	500 (3.3)

Maximum yields of α-L-Fucp-(1→3)-D-GlcNAc and α-L-Fucp-(1→6)-D-GlcNAc synthetized by AlfB and AlfC forms, respectively. Protein concentrations used to obtain maximum yields within 3 h are given in µg·mL⁻¹ and µM (in brackets).

The time course of disaccharide formation catalysed by AlfB-WT reveals a peculiar behaviour, with a rate increase over time, followed by a sharp decrease of product yield (Fig. 2 top). This results from a combination of substrate inhibition (Fig. S1) and the near absence of secondary hydrolysis in the presence of α-L-Fucp OCNP. The latter is clearly visible from the constant increase in disaccharide formation compared to the donor consumption until > 90% of α-L-Fucp OCNP is consumed (Fig. 2 bottom). This observation supports the postulate that the low yield of transglycosylation products from AlfB-WT is related to a low T/H ratio, rather than secondary hydrolysis. To further investigate this, the T/H ratio was assesed using the initial slope of plot of product formation as a function of donor consumption (Fig. 2 bottom). Clearly, even in the case of AlfB-WT, secondary hydrolysis barely affects the disaccharide yield until the donor is almost fully depleted, thus providing strong evidence showing that maximum yield correlates strongly with the T/H ratio (Table 2). The mutant AlfB-H21F displayed hardly detectable activity and so its study was discontinued. However, among the other six AlfB variants, H80F and W130H generated disaccharide at significantly higher yields than AlfB-WT. Notably, AlfB-H80F reached a transfucosylation yield of 57%. This was accompanied by complete regioselectivity and a high transglycosylation rate (> 11 mM α-L-Fucp-(1→3)-D-GlcNAc were obtained within 1 h with 5 µM enzyme; Fig. S2). It is noteworthy that mutation of the

neighbouring residue (H81F) almost eliminated transglycosylation, with no evidence of secondary hydrolysis. Therefore, the two adjacent H80 and H81 possess opposite but determinant roles in regulating the T/H ratio (Table 2).
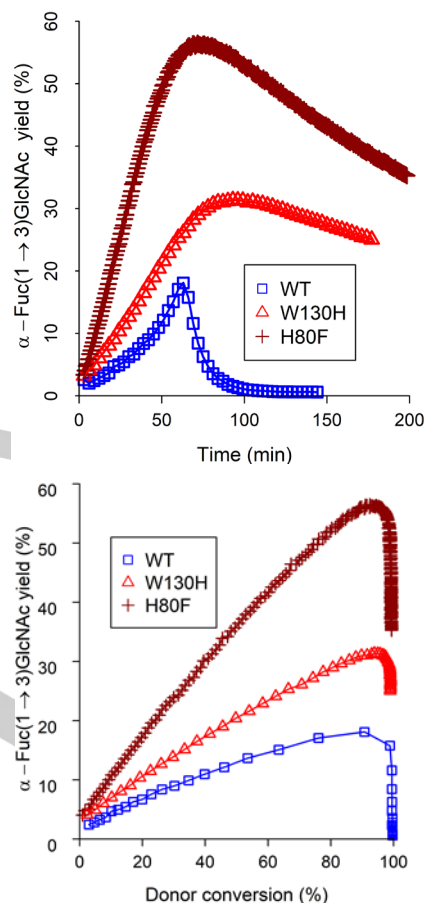

Figure 2. GH29 engineering. Single reactions NMR monitoring of the product formation with AlfB-WT and its two best mutants: transglycosylation product evolution as a function of time (top) and donor conversion (bottom).

Regarding secondary hydrolysis, even though this doesn't play a preponderant role until the donor is consumed, it is clearly responsible for rapid hydrolysis of transglycosylation products at the end of the reaction (Figs. 2 and S2). Therefore, for synthetic purposes, reducing secondary hydrolysis remains important to gain better control over the reaction. From this standpoint, it is noteworthy that the AlfB variants, particularly AlfB-W130H, displayed decreased secondary hydrolysis (Fig. 2 top).

To demonstrate that our results are portable to related enzymes, the best mutations H80F and W130H were transposed to another fucosidase from *Lactobacillus casei*, AlfC, (< 30% identity), that synthesizes α-L-Fucp-(1→6)-D-GlcNAc.[33,34] This yielded two mutants AlfC-H87F and AlfC-W137H that both displayed improved transglycosylation yields of the disaccharide in reactions containing equimolar (8 mM) amounts of α-L-Fucp OCNP and D-GlcNAc (Table 2, Figs. S3 and S4). This result illustrates how successful mutations achieved for one enzyme can be directly transposed to related GHs without the need for further analyses. In addition to the reduction of the sequence space and increased relevance, mutation transfer is a

4

RESEARCH ARTICLE

key advantage of targeting conserved residues. However, regarding secondary hydrolysis, it is noteworthy that unlike AlfB, transglycosylation catalysed by the three AlfC variants was strongly influenced by secondary hydrolysis, with maximum yields being observed at 50–60% of α-L-FucpOCNP consumption (Fig. S4).

GH151, α-L-fucosidase PtGH151

With only three published reports,[35–37] information relating to GH151 family is scarce. Importantly, the identity of the catalytic residues is unknown, a crystal structure has yet to be solved and so far no mechanistic studies have been performed. Nevertheless, a GH151 from *Paenibacillus thiaminolyticus* has been shown to display transglycosylation activity.[36] Therefore, it is plausible that GH151 shows mechanistic similarity with GH29. Overall, GH151 is an ideal case to test whether our methodology is applicable to poorly characterized GH families.

Applying the methodology to PtGH151, the transglycosylating GH151 from *P. thiaminolyticus*,[36] a set of 16 most conserved residues were identified and 11 were selected for mutagenesis, yielding 12 mutants (Table 3). Surprisingly, even though we were unable to exclude the mutation of the catalytic residues, no completely inactive variants were obtained, although mutation of D154 was the most deleterious to the overall activity (50 000-fold decrease; see Table S1). The WT enzyme catalyses self-condensation of 10 mM α-L-FucpOCNP and produces α-L-Fucp-(1→x)-α-L-FucpOCNP disaccharides in 18% yield.

Table 3. GH151 engineering.

PtGH151 forms	Yield (%)	Ratio T/H	Enzyme (µg·mL⁻¹, µM)
WT	18	0.32	1.8 (0.024)
D9N	28	0.49	8300 (108)
H11F	–	–	–
H11N	24	0.45	13 (0.17)
Y86F	19	0.31	9400 (122)
D154N	18	0.34	1000 (13)
N214T	15	0.28	130 (1.7)
E235Q	0.5	0.04	270 (3.6)
L237V	30	0.60	770 (9.9)
Y244F	22	0.44	1080 (14)
F268H	16	0.31	800 (10)
W272H	15	0.29	3500 (45)
D302N	23	0.38	38 (0.5)

Maximum yields of α-L-Fucp-(1→x)-α-L-FucpOCNP by PtGH151 forms from 10 mM α-L-FucpOCNP. Protein concentrations used to obtain maximum yields within 6 h are given in µg·mL⁻¹ and µM (in brackets).

Moderate secondary hydrolysis was observed with PtGH151-WT (Fig. 3B), and initial disaccharide formation increased linearly as a function of donor consumption (Fig. 3), which in self-condensation is strictly equivalent to acceptor consumption. This
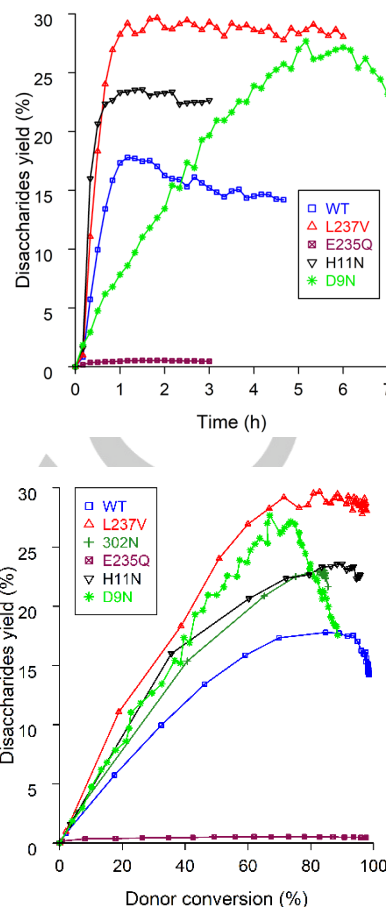
Figure 3. Single reactions reverse phase HPLC monitoring of the product formation with PtGH151 forms: transglycosylation disaccharides evolving as a function of time (top) and donor conversion (bottom).

dual consumption of the substrate led to quite rapid product yield decrease, with maximum yield being observed at 3 mM substrate concentration (Fig. 3 bottom), despite the modest impact of secondary hydrolysis (Fig. 3 top). Overall, 5 out of the 12 mutants procured improved synthetic yields (22–30%), with PtGH151-L237V being the best performer (30% yield) at an enzyme loading of only 10 µM. Significantly, the T/H ratio of the reaction catalysed by this mutant (Table 3) was almost double that of PtGH151-WT (0.60 vs. 0.32), as assessed from the initial slope of product formed as a function of donor consumed (Fig. 3 bottom). Similarly, the other four beneficial mutants showing increased yields also display improved T/H ratios (Fig. 3). Inversely, E235Q almost completely abolished transglycosylation (< 1% synthetic yield). Interestingly, PtGH151-D9N displayed low activity, but a high yield of the transglycosylation product and increased secondary hydrolysis to synthesis ratio. Overall, these results provide a satisfactory demonstration of the relevance of our method for the engineering of higher transglycosylation capabilities in unknown GH families.

5

RESEARCH ARTICLE

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,[38,39] a feature that significantly complicates the organic synthesis of furanosides. Since catalysis mediated by any given GH involves a specific conformational itinerary for the donor substrate, the lowered energy barriers between furanoside conformations[38] implies that furanosidase-catalysed reactions display altered mechanisms when compared to those catalysed by pyranosidases, and GH51 α-L-arabinofuranosidases have been predicted to present $[E_3]^{\ddagger}$ transition states along their conformational itinerary ($^4E \rightarrow [E_3]^{\ddagger} \rightarrow {}^2E \rightarrow [E_3]^{\ddagger} \rightarrow E^3/^4T_3$).[40] Therefore, it is of interest to investigate to what extent the present GH engineering approach applies to furanosidases.

The hydrolytic α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*Tx*Abf) belonging to the GH51 family[41] was used as a model furanosidase. Since *Tx*Abf displays inherent ability to perform transfuranosylation,[42,43] it has been previously subjected to several studies aiming at improving its transglycosylation capability.[22,44–46] Here all eight generated mutants possessed improved ability to synthesize arabinoxylo-oligosaccharides (AXOS) in 19–82% yields compared to 9% yield by *Tx*Abf-WT (Tables 4 and S2). The best variants R69K, N175T, D297N and H240F/N produced AXOS in 37–82% overall yields. The transglycosylation activity (specific activity in transglycosylation mode, SA_T; Table S3) of the mutants was lower (0.1–66%) than that of the WT enzyme. It is noteworthy that R69K-, H240F- and D297N-catalysed transglycosylation required relatively low amounts of catalyst (32−140 nM, Table 4).

Table 4. GH51 engineering.

*Tx*Abf forms	Yield[a] (%)	Ratio T/H	Enzyme[b] (µg·mL⁻¹, µM)
WT	9	0.14	0.2 (0.0035)
F26H	22	0.56	730 (12.8)
E28Q	23	0.69	0.8 (0.014)
R69K	37	1.05	2 (0.035)
N175T	42	3.98	200 (3.5)
H240F	82	15.0	8 (0.14)
H240N	75	>50	220 (3.9)
Y242F	19	0.47	300 (5.3)
D297N	47	3.37	1.8 (0.032)

[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 µM (in brackets).

Although the ability of *Tx*Abf-WT to perform transglycosylation is low, this is not due to secondary hydrolysis because AXOS production increments linearly as a function of α-L-Ara*fOp*NP consumption (Fig. 4 bottom). Therefore, this implies that its poor performance regarding transglycosylation is more related to its low T/H ratio (0.14; Table 4). Significantly, all the *Tx*Abf mutants created in this study displayed an increased (3.5–100 fold) T/H ratio, consistent with improved transglycosylation yields. However, in some mutants, particularly *Tx*Abf-Y242F and *Tx*Abf-E28Q,

secondary hydrolysis was also enhanced. Consequently, *Tx*Abf-H240N, displaying the highest T/H ratio, procured lower yields than *Tx*Abf-H240F due to higher secondary hydrolysis.
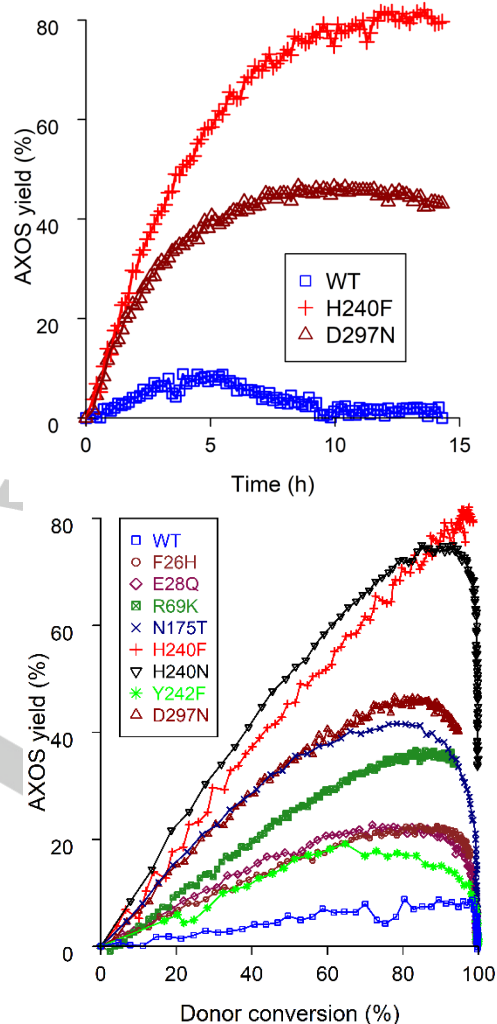

Figure 4. Single reactions NMR monitoring of transglycosylation yield progress for *Tx*Abf forms: transglycosylation AXOS evolution as a function of time (top) or donor conversion (bottom).

Interestingly, two of the targeted amino acid residues were previously identified using random mutagenesis and screening which yielded the mutants F26L and R69H[22] that, compared to F26H and R69K described herein, are slightly better catalysts for transglycosylation (1.2-fold). However, the strategy described is much less labour-intensive, as it circumvents large library screening.

Another remarkable observation is that H240N and H240F display different yields and regioselectivities (Table S2). Apparently, aromaticity is more important to TS‡ stabilisation than polarity and/or eventual hydrogen bond formation. In this respect, it is interesting to note that the exact opposite was observed for position 11 of *Pt*GH151. Overall, these results constitute a strong case for using double substitution (i.e. H→N or F) when targeting His, a residue that displays dual physicochemical properties.

Significantly, compared to all previously reported single-mutants that enhance the T/H ratio in *Tx*Abf,[22,44–46] H240F/N display the highest transglycosylation yields, combined noticeably higher regioselectivity (Table S2). Consequently, we combined this

RESEARCH ARTICLE

mutation with N216W,[22] a mutation that favours higher regioselectivity towards the (1→2)-linkage of α-L-arabinofuranosyl moieties to the non-reducing terminal D-xylopyranosyl of xylotriose as acceptor (i.e. A²XX).[22] Both transglycosylation specific activity (SA$_T$; Table S3) and yield (62 compared to 82% for H240F) remained relatively high and the reaction was almost completely regioselective (Table S2 and Fig. S5).

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. Thereby, R69K, N175T, H240F/N and D297N appear to be involved in the local H-bonding network that impacts catalysis in *Tx*Abf (Fig. S6 and related discussion).

Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) (Fig. S7, Tables S4 and S5) were used to determine the chemical structure of the products. Five variants (R360K, N428T, S474A, W612H and W623H) showed markedly improved yields for *p*NP-Man$_2$ synthesis (32–43%) compared to *Cf*Man2A-WT (18%), while W169H, D170N, D386N and H496F were discarded due to low activity (< 5 nkat·mg⁻¹). The five mutants that showed the highest yields also presented the highest T/H ratios (1.9 to 3.2-fold increase; Table 5).

GH2, β-mannosidase Cf*Man2A*

Compared to the synthesis of α-D-mannosidic (or of β-D-glucosidic and β-D-galactosidic) bonds, the synthesis of β-D-mannosyl-containing compounds is complicated by the axial OH-2 of D-mannosyl moieties. In the case of β-mannosidase-mediated synthesis, this constraint leads to a different conformational itinerary ($^1S_5 \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^0S_2$) compared to that adopted by most other β-pyranosidases-catalysed reactions ($^1S_3 \rightarrow [^4H_3]^\ddagger \rightarrow {}^4C_1$).[47] Overall, the challenge of synthesizing β-D-mannosides[48] has led to interest in enzymatic synthesis[49–51] and several GH2 mannosidases have been successfully tested.[52–54] Furthermore, a glycosynthase variant of the *Cellulomonas fimi* GH2 β-mannosidase (*Cf*Man2A) proved to be a proficient synthetic catalyst when fed with donor α-D-mannosyl fluoride and various acceptors.[49] Herein, we investigate whether *Cf*Man2A-mediated transmannosylation can be enhanced without using the glycosynthase strategy, i.e. keeping the catalytic glutamates intact. The transglycosylation ability of *Cf*Man2A-WT and mutants was evaluated with 5 mM *p*-nitrophenyl β-D-mannopyranoside (β-D-Man*p*O*p*NP) acting as both donor and acceptor substrate (self-condensation reaction). Hence in transglycosylation a D-mannosyl unit was transferred from a donor to an acceptor molecule, thus elongating 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*-(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 monitored using HPLC (Figs. 5 and S7).

Table 5. GH2 engineering.

Cf*Man2A* forms	Yield (%)	Ratio T/H	Enzyme (µg·mL⁻¹, nM)
WT	18	0.18	9.8 (104)
R360K	37	0.38	30.9 (330)
W362H	18	0.11	4.7 (50)
N428T	43	0.47	34.3 (360)
S474A	32	0.50	33.2 (350)
H496Q	15	0.16	5.3 (56)
W612H	37	0.58	23.9 (253)
Q613N	16	0.19	15.3 (162)
W623H	35	0.35	2.8 (30)

Maximum yields of *p*NP-Man$_2$ synthetized by *Cf*Man2A forms from 5 mM β-D-Man*p*O*p*NP. Protein concentrations used to obtain the yields within 5 h are indicated in µg·mL⁻¹ and nM (in brackets).
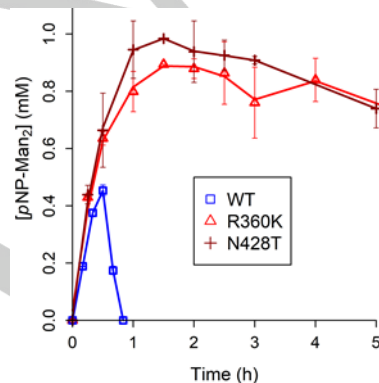

Figure 5. HPLC monitoring of the formation of *p*NP-Man$_2$ by *Cf*Man2A-WT and two of its best mutants, with error bars corresponding to the standard deviation of duplicate incubations.

It is noteworthy that the best four *Cf*Man2A mutants could be applied at loadings that were lower or only moderately greater (0.3 to 3.5-fold; Table 5) than that used for *Cf*Man2A-WT, and hundred-fold lower than used with the *Cf*Man2A-E519S glycosynthase.[49] Moreover, compared to the WT enzyme, HPLC monitoring revealed that the mutants catalysed less secondary product hydrolysis and thus ensured high product concentrations even over prolonged reaction times (Fig. 5). This result is significant because secondary hydrolysis commonly impedes enzyme-mediated synthesis of β-D-mannoside conjugates.[55,56] Finally, it is noteworthy that like in the previous examples, all mutant *Cf*Man2A that increased transglycosylation yield were also characterized by higher T/H ratios, thus providing further evidence to support tight correlation between these two parameters.

7

RESEARCH ARTICLE

GH10, β-endo-xylanase RmXyn10A_CM

Compared with most hexopyranoses, the pentopyranose configuration of the D-xylosyl unit confers enhanced flexibility and allows two distinct conformational itineraries compatible with GH hydrolysis: $^1S_3 \rightarrow [^4H_3]^{\ddagger} \rightarrow {^4C_1}$ and $^2S_0 \rightarrow [^{2,5}B]^{\ddagger} \rightarrow {^5S_1}$.[47,57] Mechanistically interesting as it may be, the real challenge posed by the engineering of RmXyn10A_CM, the catalytic module of the endo-xylanase from Rhodothermus marinus used here,[58] is related to its endo-activity. This implies the use of a substrate that does not possess a good leaving group, thus providing no kinetic control over the reaction and the necessity to monitor multiple products in order to assess transglycosylation efficiency. Indeed, the reaction allows the transfer of oligosaccharide moieties onto an acceptor in a single catalytic step and provides the means to synthesize oligosaccharides of high 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.[59–62] Both negative and positive subsites[24] in GH10 xylanases have been probed to understand their influence on substrate binding and catalysis,[63–66] including those in RmXyn10A_CM.[58] However, only a few reports focused on transglycosylation and these were limited to mutational analysis of the aglycone binding subsites in a couple of GH10 xylanases.[63,66]

Here the transglycosylation activity was characterized using xylotetraose (X_4) as both donor and acceptor substrate. All RmXyn10A_CM forms were able to synthesise xylo-oligosaccharides (XOS) larger than X_4 and predominantly accumulating X_8 during the 4 h time frame of the experiment (Fig. 6 and Tables 6 and S6). Notably, the mutants H69N, N118T and W284H synthesized XOS with higher DP than the WT enzyme during the disproportionation of 20 mM X_4.

Table 6. GH10 engineering.

RmXyn10A_CM forms	Maximal X_8 area (nC·min)	T_E/H_E ratio at 2h	Enzyme (µg·mL⁻¹, nM)
WT	4.9	0.58	10.2 (21)
H69F	8.4	5.42	43.6 (90)
H69N	11.6	4.69	37.7 (78)
W73H	8.3	4.17	41.3 (86)
N118T	13.4	9.78	39.5 (82)
N169T	5.5	2.23	10.5 (22)
H204F	8.9	7.39	35.6 (74)
H204N	6.9	5.29	36.3 (75)
D234N	6.6	4.87	39.1 (81)
W284H	10.7	4.15	38.5 (80)

Xylooctaose (X_8) synthesised by RmXyn10A_CM forms from 20 mM X_4. Protein concentrations used to obtain transglycosylation within 4 h are indicated in µg·mL⁻¹ and nM (in brackets). The ratio of T_E/H_E (see equation) is indicated for t = 2 h, full evolution of the T_E/H_E ratio is provided in SI. All reactions have been performed and analysed in duplicates.

Further analysis revealed that when using X_4 as substrate, RmXyn10A_CM mutants synthesized XOS of DP 6−11 (Figs. 6 and S8, Tables 6 and S6). Importantly, compared to the WT enzyme, all mutants display greater ability to form XOS of DP ≥ 8 (Table S6), this being the result of multiple transglycosylation events. The conserved sequence approach procured RmXyn10A mutants with significantly enhanced transglycosylation abilities compared to the WT enzyme, primarily due to a drastic decrease of hydrolytic activity (Fig. S9) and thus significantly increased T/H

ratios. Available structural data[58] reveal that the three most effective mutational targets (H69, N118 and W284) are all closely interacting with the D-xylosyl residue in the −1 subsite (Fig. S10).
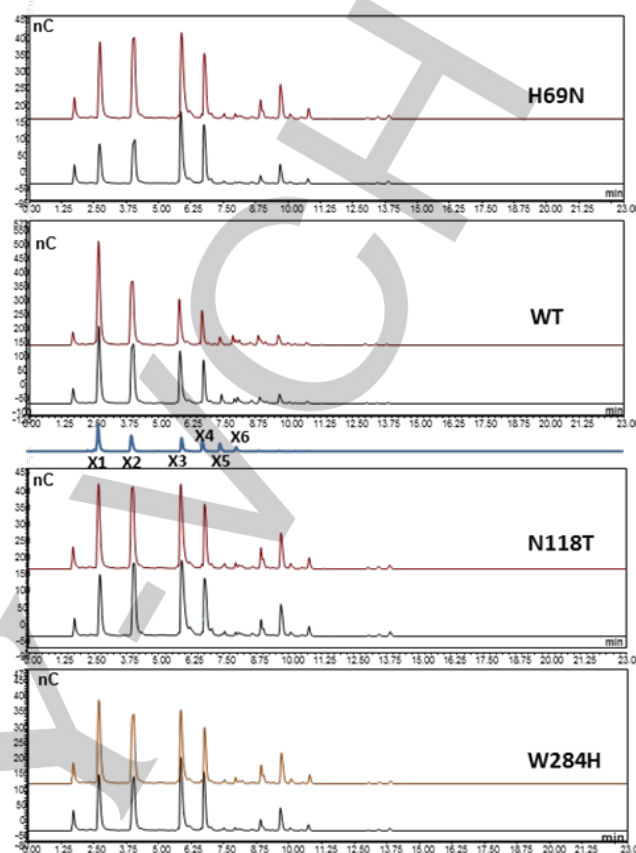
Figure 6. HPAEC-PAD analysis of xylo-oligosaccharides produced by RmXyn10A_CM forms. Chromatograms showing product profiles of transglycosylation reactions catalysed by RmXyn10A_CM-WT and its mutants H69N, N118T and W284H. For each RmXyn10A_CM form, chromatograms display reaction profiles after 2 (black) 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.

Assessing secondary hydrolysis in this endo-reaction is quite difficult, because the transglycosylation products are necessarily both donors and acceptors of further transglycosylation and hydrolysis reactions. Therefore, the measured time progression of the ratio T_E (transglycosylation events)/H_E (hydrolysis events) is underestimated:

$$T_E = \sum_{n=5}^{n=10} A(X_n) + \sum_{n=8}^{n=10} A(X_n)$$

where $A(X_n)$ is the area of the peak corresponding to a XOS of DP n. XOS displaying DP 8–10 are listed twice as a minimum of two transglycosylation events are needed to yield these oligosaccharides.

$$H_E = -T_E + \sum_{n=1}^{n=3} A(X_n)$$

Note that T_E is subtracted from the sum of the areas of peaks of XOS of DP 1–3 as each transglycosylation event yields a XOS of DP 1–3 as well. The final ratio is necessarily underestimated as

8

RESEARCH ARTICLE

XOS of DP > 10 are not accounted for. The full-time course of the T_E/H_E ratio is displayed in Table S6. Remarkably, this ratio only undergoes limited variation throughout the period from 2 h to 4 h, despite significant changes in product profile (i.e. in all reactions the monosaccharide concentration increases and that of di-and trisaccharides decreases; Table S6). For all mutants the T_E/H_E ratio increased by 7–17 folds.

GH20, β-hexosaminidase Bb*HI*

During catalysis, most retaining GHs generate a glycosyl-enzyme intermediate whose anomeric configuration is opposite to that of both the substrate (donor) and the products. However, GH families 18, 20, 25, 56, 84, 85 and 123 use a substrate-assisted mechanism, in which an equatorial *N*-acetyl in position C-2 of the donor acts as the nucleophile.[29] This leads to the formation of a non-covalent oxazoline- or oxazolinium ion-enzyme intermediate[67] (most likely an oxazolinium ion in the case of GH20[68]). Importantly, this type of mechanism predominates in GH-catalysed reactions involving β-D-Glc*p*NAc or β-D-Gal*p*NAc, which are highly prevalent in biological systems. Hence, to assess if the conserved sequence methodology presented here is also applicable when a glycosyl-enzyme intermediate is not formed, a GH20 enzyme was investigated.

Previously, GH20 hexosaminidases have been extensively used for transglycosylation.[69–71] Herein we focus on *Bb*HI from *Bifidobacterium bifidum*.[72,73] This enzyme has the ability to catalyse the synthesis of LNT2 (β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-D-Glc, also known as lacto-*N*-triose II), a sought-after glycomotif present in human milk oligosaccharides.[7,74] Recently, *Bb*HI-D746E – D746 being involved in the oxazolinium ion stabilisation – was reported as the first glycosynthase with high enzymatic activity, thus requiring only moderate enzyme loading.[75] Six mutants and *Bb*HI-WT were analysed for their ability to synthesize LNT2 from 10 mM β-D-Glc*p*NAc*Op*NP and 40 mM lactose (Table 7 and Fig. 7). With no detectable D-GlcNAc release, *Bb*HI-H603F seemed to display the highest T/H ratio, as LNT2 formation was detected without noticeable GlcNAc formation, however combined with drastically reduced activity. At high enzyme loading (15 µM), the reaction was incomplete after 12 h, hence further characterization of *Bb*HI-H603F was abandoned. Note that T/H ratio determination was not pursued as the pulsed amperometric detection does not allow for β-D-Glc*p*NAc*Op*NP concentration determination.

The variants R577K and W882H displayed high yields of LNT2 (66% for W882H), although the latter was poorly active. *Bb*HI-R577K represents an interesting compromise, as it generated a markedly higher yield of LNT2 (36 vs 16% for *Bb*HI-WT) at a reasonable enzyme loading (1.3 µM).

These results demonstrate that the conserved sequence approach is not restricted to retaining GHs catalysing transglycosylation through a covalent glycosyl-enzyme intermediate. Conversely, application of the approach to another GH20, from *Ewingella americana* failed to generate variants able to perform transglycosylation using β-D-Glc*p*NAc*Op*NP as donor and lactose as an acceptor (data not shown). Although this indicates that the strategy is not fully generic, we believe that it was unsuccessful because the *Ewingella americana* WT GH20 did not possess detectable innate ability to use lactose as a glycosyl acceptor. This supports our hypothesis that an innate, even weak, transglycosylation activity is likely a prerequisite for success using our approach.
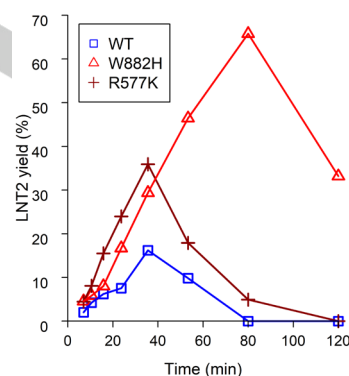
Figure 7. Single reactions HPAEC-PAD monitoring of LNT2 production by *Bb*HI-WT and its two best mutants from 10 mM β-D-Glc*p*NAc*Op*NP and 40 mM lactose.

Table 7. GH20 engineering, maximum yields of LNT2 synthetized by *Bb*HI forms from 10 mM β-D-Glc*p*NAc*Op*NP and 40 mM lactose. Protein concentrations used to obtain the yields within 2 h are indicated in µg·mL⁻¹ and µM (in brackets).

*Bb*HI forms	LNT2 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)

9

RESEARCH ARTICLE

Conclusion

Mutating enzymes from GH families 2, 10, 20, 29, 51 and 151 (belonging to clans A, K, R, or no clan),[5] led to the successful enhancement of transglycosylation yields for most of the selected candidates (Table 8). This validates our conserved-residue approach and illustrates its applicability to retaining GHs irrespective of the structural fold, the specific catalytic mechanism or *exo- vs. endo*-action. Moreover, a variety of sugars, D/L-configurations, pyranose/furanose forms and α/β-stereochemistry are allowed, 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 40−80%, and thus allowed the 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, at least two mutants displaying enhanced transglycosylation yield with reasonable enzyme loadings were obtained for each of the six GHs families targeted. Finally, it is remarkable that irrespective of the target GH, increased T/H ratio (no data available for GH20) correlated with improved yields (Table 8). On the other hand, even when secondary hydrolysis is high (e.g. AlfB-WT), this reaction had little effect on the maximal yield of transglycosylation products. Therefore, one can conclude that even if secondary hydrolysis is undesirable from a practical point of view, it does not need to be a priority target when improving transglycosylation yields.

Table 8. Summary of the transglycosylases improvement via an elaborated sequence conservation-based approach of enzyme engineering.

GH family	Enzyme target	Sequences retrieved	Evaluated candidates (conserved residues)	WT		Best mutant[a]		Improvement rate (%)[b]
				Yield (%)	Ratio T/H	Yield (%)	Ratio T/H	
2	*Cf*Man2A	20 000	12 (14)	18	0.18	43	0.58	42
10	*Rm*Xyn10A	9 288	9 (11)	n.d.[c]	0.58	n.d.[c]	9.78	100
20	*Bb*HI	20 000	6 (9)	16	n.d.[d]	66	n.d.[d]	33
29	AlfB	15 328	7 (13)	18	0.29	57	1.89	57
	AlfC	n. d.[d]	2 (n. d.)	17	1.00	26	1.93	100
51	*Tx*Abf	11 553	8 (11)	9	0.14	82	> 50	100
151	*Pt*GH151	1 854	12 (16)	18	0.32	30	0.60	58

[a] Yield and T/H ratio come from either the same or different mutants. [b] Percentage of evolved enzymes displaying better transglycosylation capability compared to its corresponding WT enzyme, i.e. yield and/or T/H ratio. [c] n.d.: not determined since *Rm*Xyl10A catalysed disproportionation reaction. Proportion of formed X_8 from X_4 was used to evaluate this criterion. [d] n.d.: not determined (not available).

Converting hydrolytic GHs into potent glycosynthetic tools is an attractive approach to extend the chemist's toolbox, introducing catalysts that are compatible with green chemistry principles. To some extent this has already been achieved, yet the time cost and effort to develop such tools is generally high. In this regard, the strategy presented herein constitutes a considerable step forward, because it simplifies and thus democratizes access to improved enzymes for glycosynthesis. For chemists, this implies access to straightforward synthesis reactions, and obviate the need for difficult-to-obtain sugar donors, cumbersome protection/deprotection cycles or vast quantities of enzyme. Another attractive feature of our strategy is its portability. This means that once a beneficial mutation is found in one GH family member, it should be transferable to related enzymes (e.g. in members of the same family). This is precious because it offers the ability to expand the portfolio of available evolved transglycosylases and synthetic oligosaccharides. Nevertheless,

one caveat is that some key determinants of the T/H ratio might not be conserved residues and will thus be inaccessible using this method, although this will not be the case for all T/H ratio determinants. Therefore, the conclusion is that the method is highly likely to generate improved mutants, although the theoretically best mutant may not be among them.

In summary, the conserved sequence approach is novel because it procures engineered transglycosylases without the need for extensive screening or prior in-depth knowledge of the target enzyme. Moreover, the method systematically targets conserved residues, which are generally omitted in enzyme engineering to avoid loss of activity or stability.[76] The power of the approach residues in the fact that its aim is not to improve a defining characteristic of an enzyme, but rather to eliminate a property (i.e. the ability to perform hydrolysis). We anticipate that this method can be applied to other enzymes displaying activities or characteristics that must be suppressed rather than enhanced.

10

10.1002/chem.202100110

RESEARCH ARTICLE

Conflict of Interest

The authors declare no conflict of interest.

[1]   R. A. Laine, *Glycobiology* **1994**, *4*, 759–767.

[2]   M. M. Nielsen, C. M. Pedersen, *Chem. Rev.* **2018**, *118*, 8285–8358.

[3]   M. Panza, S. G. Pistorio, K. J. Stine, A. V. Demchenko, *Chem. Rev.* **2018**, *118*, 8105–8150.

[4]   L. Krasnova, C. H. Wong, *J. Am. Chem. Soc.* **2019**, *141*, 3735–3754.

[5]   V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490-5.

[6]   K. Yonekura-Sakakibara, K. Hanada, *Plant J.* **2011**, *66*, 182–193.

[7]   M. Faijes, M. Castejón-Vilatersana, C. Val-Cid, A. Planas, *Biotechnol. Adv.* **2019**, *37*, 1–31.

[8]   D. E. Koshland, *Biol. Rev.* **1953**, *28*, 416–436.

[9]   A. Vasella, G. J. Davies, M. Böhm, *Curr. Opin. Chem. Biol.* **2002**, *6*, 619–29.

[10]   D. Teze, B. Shuoker, E. K. Chaberski, S. Kunstmann, F. Fredslund, E. N. Karlsson, T. S. Nielsen, E. G. P. Stender, G. H. J. Peters, E. N. Karlsson, et al., *ACS Catal.* **2020**, *10*, 3809–3819.

[11]   P. Lundemo, E. N. Karlsson, P. Adlercreutz, *Appl. Microbiol. Biotechnol.* **2017**, *101*, 1121–1131.

[12]   B. Zeuner, C. Jers, J. D. Mikkelsen, A. S. Meyer, *J. Agric. Food Chem.* **2014**, *62*, 9615–9631.

[13]   B. Bissaro, P. Monsan, R. Fauré, M. J. O'Donohue, *Biochem. J.* **2015**, *467*, 17–35.

[14]   L. F. Mackenzie, Q. Wang, R. A. J. Warren, S. G. Withers, *J. Am. Chem. Soc.* **1998**, *120*, 5583–5584.

[15]   M. R. Hayes, J. Pietruszka, *Molecules* **2017**, *22*, 1434.

[16]   C. Malet, A. Planas, *FEBS Lett.* **1998**, *440*, 208–212.

[17]   M. Moracci, A. Trincone, G. Perugino, M. Ciaramella, M. Rossi, *Biochemistry* **1998**, *37*, 17262–17270.

[18]   C. Mayer, D. L. Zechel, S. P. Reid, R. A. J. Warren, S. G. Withers, *FEBS Lett.* **2000**, *466*, 40–44.

[19]   M. S. Packer, D. R. Liu, *Nat. Rev. Genet.* **2015**, *16*, 379–394.

[20]   D. Teze, F. Daligault, V. Ferrières, Y.-H. Sanejouand, C. Tellier, *Glycobiology* **2015**, *25*, 420–427.

[21]   D. Teze, J. Hendrickx, M. Czjzek, D. Ropartz, Y.-H. H. Sanejouand, V. Tran, C. Tellier, M. Dion, *Protein Eng. Des. Sel.* **2014**, *27*, 13–19.

[22]   B. Bissaro, J. Durand, X. Biarnés, A. Planas, P. Monsan, M. J. O'Donohue, R. Fauré, *ACS Catal.* **2015**, *5*, 4598–4611.

[23]   J. Yang, Q. Wang, Y. Zhou, J. Li, R. Gao, Z. Guo, *Biochem. Eng. J.* **2017**, *127*, 1–8.

[24]   G. J. Davies, K. S. Wilson, B. Henrissat, *Biochem. J.* **1997**, *321*, 557–559.

[25]   W. Li, L. Jaroszewski, A. Godzik, *Bioinformatics* **2001**, *17*, 282–283.

[26]   Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, *Bioinformatics* **2010**, *26*, 680–682.

[27]   F. Sievers, D. G. Higgins, *Curr. Protoc. Bioinforma.* **2014**, *2014*, 3.13.1-3.13.16.

[28]   M. Schneider, E. Al-Shareffi, R. S. Haltiwanger, *Glycobiology* **2017**, *27*, 601–618.

[29]   C. Roth, M. Petricevic, A. John, E. D. Goddard-Borger, G. J. Davies, S. J. Williams, *Chem. Commun.* **2016**, *52*, 11096–11099.

[30]   B. Zeuner, M. Vuillemin, J. Holck, J. Muschiol, A. S. Meyer, *Enzyme Microb. Technol.* **2018**, *115*, 37–44.

[31]   A. Saumonneau, E. Champion, P. Peltier-Pain, D. Molnar-Gabor, J. Hendrickx, V. Tran, M. Hederos, G. Dekany, C. Tellier, *Glycobiology* **2015**, *26*, 261–269.

[32]   G. Osanjo, M. Dion, J. Drone, C. Solleux, V. Tran, C. Rabiller, C. Tellier, *Biochemistry* **2007**, *46*, 1022–1033.

[33]   J. Rodríguez-Díaz, R. J. Carbajo, A. Pineda-Lucena, V. Monedero, M. J. Yebra, *Appl. Environ. Microbiol.* **2013**, *79*, 3847–3850.

[34]   J. Rodríguez-Díaz, V. Monedero, M. J. Yebra, *Appl. Environ. Microbiol.* **2011**, *77*, 703–705.

[35]   D. A. Sela, D. Garrido, L. Lerno, S. Wu, K. Tan, A. Eom, A. Joachimiak, C. B. Lebrilla, D. A. Mills, *Appl. Environ. Microbiol.* **2012**, *12*, 795–803.

[36]   E. Benesová, P. Lipovová, J. Krejzová, T. Kovalová, P. Buchtová, V. Spiwok, B. Králová, *BMC Biotechnol.* **2015**, *15*, 1–7.

[37]   M. Lezyk, C. Jers, L. Kjaerulff, C. H. Gotfredsen, M. D. Mikkelsen, J. D. Mikkelsen, *PLoS One* **2016**, *11*, 1–18.

[38]   H. A. Taha, M. R. Richards, T. L. Lowary, *Chem. Rev.* **2013**, *113*, 1851–1876.

[39]   X. Wang, R. J. Woods, *J. Biomol. NMR* **2016**, *64*, 291–305.

[40]   N. G. S. McGregor, M. Artola, A. Nin-hill, D. Linzel, M. Haon, J. Reijngoud, A. Ram, M.-N. Rosso, G. A. van der Marel, J. D. C. Codée, et al., *J. Am. Chem. Soc.* **2020**, *142*, 4648–4662.

[41]   T. Debeche, N. Cummings, I. Connerton, P. Debeire, M. J. O'Donohue, *Appl. Environ. Microbiol.* **2000**, *66*, 1734–1736.

[42]   C. Rémond, R. Plantier-Royon, N. Aubry, E. Maes, C. Bliard, M. J. O'Donohue, *Carbohydr. Res.* **2004**, *339*, 2019–2025.

[43]   C. Remond, R. Plantier-Royon, N. Aubrya, M. J. O'Donohue, *Carbohydr. Res.* **2005**, *340*, 637–644.

[44]   B. Bissaro, O. Saurel, F. Arab-jaziri, L. Saulnier, A. Milon, M. Tenkanen, *Biochim. Biophys. Acta* **2014**, *1840*, 626–636.

[45]   F. Arab-jaziri, B. Bissaro, C. Tellier, M. Dion, R. Fauré, M. J. O. Donohue, *Carbohydr. Res.* **2015**, *401*, 64–72.

[46]   F. Arab-Jaziri, B. Bissaro, M. Dion, O. Saurel, D. Harrison, F. Ferreira, A. Milon, C. Tellier, R. Fauré, M. J. O'Donohue, *New Biotechnol.* **2013**, *30*, 536–544.

[47]   G. J. Davies, A. Planas, C. Rovira, *Acc. Chem. Res.* **2011**, *45*, 308–316.

[48]   J. J. Gridley, H. M. I. Osborn, *J. Chem. Soc. Perkin Trans. 1* **2000**, 1471–1491.

[49]   O. Nashiru, D. L. Zechel, D. Stoll, T. Mohammadzadeh, R. Antony J Warren, S. G. Withers, *Angew. Chemie - Int. Ed.* **2001**, *40*, 417–420.

[50]   M. Jahn, D. Stoll, R. A. J. Warren, L. Szabó, P. Singh, H. J. Gilbert, V. M. A. Ducros, G. J. Davies, S. G. Withers, *Chem. Commun.* **2003**, *3*, 1327–1329.

[51]   A. Sasaki, T. Ishimizu, R. Geyer, S. Hase, *FEBS J.* **2005**, *272*, 1660–1668.

[52]   D. L. Zechel, S. P. Reid, D. Stoll, O. Nashiru, R. A. J. Warren, S. G. Withers, *Biochemistry* **2003**, *42*, 7195–7204.

[53]   E. V. Eneyskaya, G. Sundqvist, A. M. Golubev, F. M. Ibatullin, D. R.

11

Ivanen, K. A. Shabalin, H. Brumer, A. A. Kulminskaya, *Biochimie* **2009**, *91*, 632–638.

[54] P. Shi, G. Yao, Y. Cao, P. Yang, T. Yuan, H. Huang, Y. Bai, B. Yao, *Enzyme Microb. Technol.* **2011**, *49*, 277–283.

[55] J. Morrill, A. Månberger, A. Rosengren, P. Naidjonoka, P. Von Freiesleben, K. B. R. M. Krogh, K. Bergquist, T. Nylander, E. N. Karlsson, P. Adlercreutz, et al., *Appl. Microbiol. Biotechnol.* **2018**, *102*, 5149–5163.

[56] A. Rosengren, S. J. Butler, M. Arcos-Hernandez, K. E. Bergquist, P. Jannasch, H. Stålbrand, *Green Chem.* **2019**, *21*, 2104–2118.

[57] J. Iglesias-Fernández, L. Raich, A. Ardèvol, C. Rovira, *Chem. Sci.* **2015**, *6*, 1167–1177.

[58] A. Aronsson, F. Güler, M. V. Petoukhov, S. J. Crennell, D. I. Svergun, J. A. Linares-Pastén, E. Nordberg Karlsson, *Biochim. Biophys. Acta - Proteins Proteomics* **2018**, *1866*, 292–306.

[59] D. E. Evangelista, M. A. S. Kadowaki, B. L. Mello, I. Polikarpov, *Int. J. Biol. Macromol.* **2018**, *109*, 560–568.

[60] J. Hu, J. N. Saddler, *Biomass Bioenerg.* **2018**, *110*, 13–16.

[61] E. Nordberg-Karlsson, E. Schmitz, J. A. Linares-Pastén, P. Adlercreutz, *Appl. Microbiol. Biotechnol.* **2018**, *102*, 9081–9088.

[62] C. Niderhaus, M. Garrido, M. Insani, E. Campos, S. Wirth, *Process Biochem.* **2018**, *67*, 92–98.

[63] S. Armand, S. R. Andrews, S. J. Charnock, H. J. Gilbert, *Biochemistry* **2001**, *40*, 7404–7409.

[64] S. J. Charnock, J. H. Lakey, R. Virden, N. Hughes, M. L. Sinnott, G. P. Hazlewood, R. Pickersgill, H. J. Gilbert, *J. Biol. Chem.* **1997**, *272*, 2942–2951.

[65] S. J. Charnock, T. D. Spurway, H. Xie, M. H. Beylott, R. Virden, R.

A. J. Warren, G. P. Hazlewood, H. J. Gilbert, *J. Biol. Chem.* **1999**, *273*, 32187–32199.

[66] A. Moreau, F. Shareck, D. Kluepfel, R. Morosoli, *Eur. J. Biochem.* **1994**, *219*, 261–266.

[67] J. Coines, M. Alfonso-Prieto, X. Biarnés, A. Planas, C. Rovira, *Chem. Eur. J.* **2018**, *24*, 19258–19265.

[68] D. Teze, J. Coines, L. Raich, V. Kalichuk, C. Solleux, C. Tellier, C. André-Miral, B. Svensson, C. Rovira, *J. Am. Chem. Soc.* **2020**, *142*, 2120–2124.

[69] K. Slámová, P. Bojarová, L. Petrásková, V. Křen, *Biotechnol. Adv.* **2010**, *28*, 682–693.

[70] K. Slámová, J. Krejzová, P. Marhol, L. Kalachova, N. Kulik, H. Pelantová, J. Cvačka, V. Křen, *Adv. Synth. Catal.* **2015**, *357*, 1941–1950.

[71] J. Muschiol, M. Vuillemin, A. S. Meyer, B. Zeuner, *catalysts* **2020**, *20*, 1–37.

[72] M. Miwa, T. Horimoto, M. Kiyohara, T. Katayama, M. Kitaoka, H. Ashida, K. Yamamoto, *Glycobiology* **2010**, *20*, 1402–1409.

[73] X. Chen, L. Xu, L. Jin, B. Sun, G. Gu, L. Lu, M. Xiao, *Appl. Environ. Microbiol.* **2016**, *82*, 5642–5652.

[74] B. Zeuner, D. Teze, J. Muschiol, A. S. Meyer, *Molecules* **2019**, *24*, 2033–2055.

[75] M. Weingarten, K. Baldenius, B. Nidetzky, *ACS Catal.* **2019**, *9*, 5503–5514.

[76] J. Bendl, J. Stourac, E. Sebestova, O. Vavra, M. Musil, J. Brezovsky, J. Damborsky, *Nucleic Acids Res.* **2016**, *44*, 479–487.

RESEARCH ARTICLE

Entry for the Table of Contents

Insert graphic for Table of Contents here.
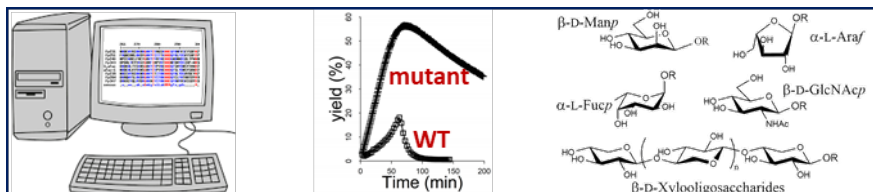
Although glycosylation is common in Nature, it remains a challenge for organic chemistry. Herein, a method to obtain transglycosylases from retaining glycoside hydrolases is described. The approach pinpoints 6–12 single-point mutants to assay. Invariably across six GH families, this generates enzyme variants that display increased transglycosylation yields, an improvement that is mostly due to a relative decrease of hydrolysis compared to transglycosylation. The method only requires sequence information, is fast and applicable to a broad diversity of glycosides.