La partition Hydrolyse / Transglycosylation chez les Glycoside Hydrolases : Proposition d'une hypothèse de synthèse à travers l'évolution moléculaire d'une α-L-arabinofuranosidase de la famille GH51 vers les premières transarabinofuranosylases de type non-Leloir

Bastien Bissaro

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par Institut National des Sciences Appliquées (INSA) de Toulouse

---

Présentée et soutenue par :

Bastien BISSARO

le lundi 15 Septembre 2014

Titre:

ON **H**YDROLYSIS/**T**RANSGLYCOSYLATION **M**ODULATION IN **G**LYCOSIDE **H**YDROLASES:

"Lessons learnt from the molecular design of the first non-Leloir transarabinofuranosylases"

Directeur de Thèse

Pierre Monsan

Jury

Antoni Planas	Professor, IQS (Barcelona, Spain)	Président
Leila Lo Leggio	Ass. Professor, Copenhagen University (Denmark)	Rapportrice
Marco Moracci	Professor, IBP (Naples, Italy)	Rapporteur
Monique Axelos	Head of CEPIA division (INRA, Nantes)	Examinatrice
Morten Sørlie	Professor, NMBU (Ås, Norway)	Examinateur
Pierre Monsan	Emeritus Professor (INSA)	Directeur de thèse
Michael O'Donohue	Research Director (INRA)	Co-encadrant
Régis Fauré	Senior Scientist (CNRS)	Co-encadrant

École doctorale et discipline ou spécialité :
ED SEVAB : Ingénieries microbienne et enzymatique

Unité de Recherche :
Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (UMR 5504)

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par

L'Institut National des Sciences Appliquées de Toulouse

ON **H**YDROLYSIS/**T**RANSGLYCOSYLATION **M**ODULATION

IN **G**LYCOSIDE **H**YDROLASES:

"Lessons learnt from the molecular design of the first non-Leloir transarabinofuranosylases"

Présentée et soutenue par :

Bastien BISSARO

le lundi 15 Septembre 2014

Directeur de Thèse

Pierre Monsan

Jury

Antoni Planas	Professor, IQS (Barcelona, Spain)	Président
Leila Lo Leggio	Ass. Professor, Copenhagen University (Denmark)	Rapportrice
Marco Moracci	Professor, IBP (Naples, Italy)	Rapporteur
Monique Axelos	Head of CEPIA division (INRA, Nantes)	Examinatrice
Morten Sørlie	Professor, NMBU (Ås, Norway)	Examinateur
Pierre Monsan	Emeritus Professor (INSA)	Directeur de thèse
Michael O'Donohue	Research Director (INRA)	Co-encadrant
Régis Fauré	Senior Scientist (CNRS)	Co-encadrant

École doctorale et discipline ou spécialité :
ED SEVAB : Ingénieries microbienne et enzymatique

Unité de Recherche :
Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (UMR 5504)

“The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality.”

Albert Einstein

“ Je ne cherche pas à connaître les réponses,
je cherche à comprendre les questions ”

Confucius

“ Regarder ‘loin’, c’est regarder ‘tôt’ ”

Hubert Reeves

NOM : BISSARO **Prénom :** Bastien Pierre André

Titre: La partition Hydrolyse/Transglycosylation chez les Glycoside Hydrolases : « *Proposition d'une hypothèse de synthèse à travers l'évolution moléculaire d'une α-L-arabinofuranosidase de la famille GH51 vers les premières transarabinofuranosylases de type non-Leloir*».

Specialité: Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries

Filière: Ingénieries microbienne et enzymatique

Année: 2014 **Lieu:** INSA de Toulouse

Jury

Antoni Planas, Professor (IQS, Barcelona, Spain)	• Président
Leila Lo Leggio, Ass. Professor (Copenhagen University, Denmark)	• Rapportrice
Marco Moracci, Professor (IBP, Naples, Italy)	• Rapporteur
Monique Axelos, Head of CEPIA division (INRA, Nantes)	• Examinatrice
Morten Sørlie, Professor (NMBU, Ås, Norway)	• Examinateur
Pierre Monsan, Emeritus Professor (INSA)	• Directeur de thèse
Michael O'Donohue, Research Director (INRA)	• Co-encadrant
Régis Fauré, Senior Scientist (CNRS)	• Co-encadrant

Référent Agreenium

Stéphane Guilbert, Professeur	• SupAgro, Montpellier

Equipe de Recherche: Catalyse et Ingénierie Moléculaire Enzymatiques (CIMEs), Axe Biocatalyse, LISBP de Toulouse

Encadrants de thèse: Pierre Monsan, Michael O'Donohue et Régis Fauré

LAST NAME : BISSARO **First name :** Bastien Pierre André

Title: On **H**ydrolysis/**T**ransglycosylation **M**odulation in **G**lycoside **H**ydrolases:

"Lessons learnt from the molecular design of the first non-Leloir transarabinofuranosylases"

Specialty: Ecological, Veterinary, Agronomic Sciences and Bioengineering

Field: Enzymatic and Microbial engineering

Year: 2014 **Place:** INSA of Toulouse

Jury

Antoni Planas, Professor (IQS, Barcelona, Spain)	• Président
Leila Lo Leggio, Ass. Professor (Copenhagen University, Denmark)	• Rapportrice
Marco Moracci, Professor (IBP, Naples, Italy)	• Rapporteur
Monique Axelos, Head of CEPIA division (INRA, Nantes)	• Examinatrice
Morten Sørlie, Professor (NMBU, Ås, Norway)	• Examinateur
Pierre Monsan, Emeritus Professor (INSA)	• Directeur de thèse
Michael O'Donohue, Research Director (INRA)	• Co-encadrant
Régis Fauré, Senior Scientist (CNRS)	• Co-encadrant

Agreenium adviser

Stéphane Guilbert, Professor	• SupAgro, Montpellier

Research team: Enzymatic Catalysis and Molecular Engineering (CIMEs) team, Biocatalysis axis, LISBP

Thesis Advisors: Pierre Monsan, Michael O'Donohue and Régis Fauré

RÉSUMÉ

Élargir le répertoire de composés accessibles dans le domaine des Glycosciences est d'un intérêt majeur pour la communauté des biologistes du fait que ces composés, oligosaccharides et glyco-conjugués, sont impliqués dans diverses fonctions biologiques, aussi bien au niveau structurel, qu'énergétique voire même signalétique jouant un rôle primordial dans les interactions inter- ou intracellulaires. L'assemblage, la modification ou la déconstruction de ces glyco-structures complexes est possible grâce à l'action d'enzymes, parmi lesquelles l'on retrouve les CAZymes (Carbohydrate Active enZymes). Ces enzymes font partie du répertoire de la base de données CAZy, incluant les Glycoside Hydrolases (GHs) qui représentent le groupe le plus important et ayant pour fonction biologique principale l'hydrolyse des liens glycosidiques. Cependant, un certain nombre de GHs possède aussi la capacité de catalyser des réactions de synthèse (transglycosylation) en tant qu'activité secondaire mineure, voire en tant qu'activité principale pour un nombre restreint d'entre elles, qui sont alors appelées transglycosylases. Sachant que ces deux types de comportements peuvent se retrouver au sein d'une même famille de GH (donc étroitement liés sur le plan évolutif), la **découverte et la compréhension des déterminants moléculaires** qui ont été développés par les GHs au cours de leur **évolution** pour permettre cette **partition** d'activité, **entre hydrolyse et transglycosylation**, est d'une importance capitale pour le domaine de la synthèse chimio-enzymatique et des Glycosciences de manière plus générale.

Ce travail de thèse décrit une proposition de synthèse pour apporter une réponse à cette question fondamentale via une revue critique de la littérature sur le sujet. Sur le plan expérimental, a été réalisée l'évolution moléculaire d'une enzyme spécifique des pentoses, l'α-L-arabinofuranosidase de *Thermobacillus xylanilyticus* (*Tx*Abf) de la famille GH51, vers les premières transarabinofuranosylases de type 'non-Leloir'. Cette évolution itérative a été développée en utilisant un panel d'outils d'ingénierie enzymatique combinant des approches aléatoire, semi-rationnelle, de prédiction *in silico* suivie de recombinaison dans un processus d'évolution dirigée global. Une analyse fine des mutants générés sur le plan mécanistique en lien avec la partition hydrolyse/transglycosylation mène à des conclusions en accord avec la proposition de synthèse issue de la revue de la littérature sur le sujet. Sur un plan plus appliqué, ces nouveaux biocatalyseurs ont ensuite été mis en oeuvre dans des voies de synthèse chimio-enzymatiques pour la préparation de composés furanosylés de structure contrôlée. Le transfert d'L-arabinofuranosyles permet la génération d'arabinoxylo-oligosaccharides (AXOS) ainsi que la conception d'oligosaccharides non naturels, tel que des galactofuranoxylo-oligosaccharides ou des arabinofuranogluco-oligosaccharides. Dans son ensemble, ce travail de recherche constitue les premières étapes clés du développement de méthodes de synthèse chimio-enzymatique plus élaborées pour la conception d'arabinoxylanes artificiels.

Dans le contexte actuel de transition vers une bio-économie, reposant sur des concepts tels que ceux de la bioraffinerie ou de la chimie verte, nous espérons que les outils de glycosynthèse développés au cours de ces travaux trouveront leur application dans la valorisation des pentoses issus de la biomasse. La synthèse à-façon d'**arabinoxylooligo- et polysaccharides** présente nombre de valorisations possibles allant de la préparation de prébiotiques à la conception de matériaux bio-inspirés en passant par la synthèse de modèles de parois végétales.

Mots Clés: Transglycosylation, Arabinofuranosidase, Evolution Dirigée, Ingénierie Rationnelle, Mécanisme enzymatique, Pentoses, Furanoses

SUMMARY

Widening the spectrum of available compounds in the field of Glycosciences is of utmost importance for the entire biology community, because carbohydrates are determinants of a myriad of life-sustaining or threatening processes. The assembly, modification or deconstruction of complex carbohydrate-based structures mainly involves the action of enzymes, among which one can identify Carbohydrate Active enZymes (CAZymes). These enzymes form part of the CAZy database repertoire and include Glycoside Hydrolases (GHs), which are the biggest group of CAZymes, whose main role is to hydrolyze glycosidic linkages. However, some GHs also display the ability to perform synthesis (transglycosylation), an activity that mostly manifests itself as a minor one alongside hydrolysis, but which is the only activity displayed by a rather select group of GHs that are often called transglycosylases. Understanding **how transglycosylases have resulted from the process of evolution is both intriguing and crucial**, because it holds the key to the creation of tailored glycosynthetic enzymes that will revolutionize the field of glycosciences.

In this thesis, an extensive review of relevant scientific literature that treats the different aspects of GH-catalyzed transglycosylation and glycosynthesis is presented, along with experimental results of work that has been performed on a family GH-51 α-L-arabinofuranosidase, a pentose-acting enzyme from *Thermobacillus xylanilyticus* (*Tx*Abf). The conclusions of the literature are presented in the form of a hypothesis, which describes the molecular basis of the hydrolysis/transglycosylation partition and thus provides a proposal on how to engineer dominant transglycosylation activity in a GH. Afterwards, using a directed evolution approach, including random mutagenesis, semi-rational approaches, *in silico* predictions and recombination it has been experimentally possible to create the very first 'non-Leloir' transarabinofuranosylases. The mechanistic analysis of the resultant *Tx*Abf mutants notably focusing on the hydrolysis/transglycosylation partition reveals that the results obtained are consistent with the initial hypothesis that was formulated on the basis of the literature review.

To demonstrate the applicative value of the experimental work performed in this study, the *Tx*Abf mutants were used to develop a chemo-enzymatic methodology that has procured a panel of well-defined furanosylated compounds. Enzyme-catalyzed transfer of arabinofuranosyl moities can be used to generate **arabinoxylo-oligosaccharides** (AXOS), but the design of non-natural oligosaccharides, such as galactofuranoxylo-oligosaccharides or arabinofuranogluco-oligosaccharides is also possible. Overall, the work presented constitutes the first steps towards the development of more sophiscated methodologies that will procure the means to synthesize artificial arabinoxylans, with a first proof of concept being presented at the very end of this manuscript.

In the present context of the bioeconomy transition, which relies on technologies such as biorefining and green chemistry, it is expected that the glycosynthetic tools that have been developed in this work will be useful for the conversion of **pentose sugars** obtained from biomass. The synthesis of tailor-made arabinoxylo-oligo- and polysaccharides may lead to a variety of potential applications including the production of prebiotics, surfactants or bio-inspired materials and, more fundamentally, the synthesis of artificial models of plant cell wall.

Key words: Transglycosylation, Arabinofuranosidase, Directed Evolution, Rational Engineering, Enzymatic Mechanism, Pentoses, Furanoses

SCIENTIFIC CONTRIBUTION

PUBLICATIONS

[1] Faten Arab-Jaziri*, **Bastien Bissaro***, Sophie Barbe, Olivier Saurel, Hélène Débat, Claire Dumon, Virginie Gervais, Alain Milon, Isabelle André, Régis Fauré and Michael J. O'Donohue (2012). *FEBS Journal,* **279** (19), 3598-3611. The functional roles of H98 and W99 and β2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. [online version]

[2] Faten Arab-Jaziri*, **Bastien Bissaro***, Michel Dion, Olivier Saurel, David Harrison, Fernando Ferreira, Alain Milon, Charles Tellier, Régis Fauré and Michael J. O'Donohue (2013). *New Biotechnology,* **30** (5), 536-544. Engineering transglycosidase activity into a GH51 α-L-arabinofuranosidase. [online version]

[3] **Bastien Bissaro**, Olivier Saurel, Faten Arab-Jaziri, Luc Saulnier, Alain Milon, Maija Tenkanen, Pierre Monsan, Michael J. O'Donohue and Régis Fauré (2013). *Biochim. Biophys. Acta - Gen. Subjects*, **1840** (1), 626-636. Mutation of a pH-modulating residue in a GH51 α-L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products. [online version]

[4] Faten Arab-Jaziri, **Bastien Bissaro**, Charles Tellier, Michel Dion, Régis Fauré and Michael J. O'Donohue. Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α-L-arabinofuranosidase. *Submitted*

[5] **Bastien Bissaro**, Julien Durand, Xevi Biarnés, Antoni Planas, Pierre Monsan, Régis Fauré and Michael J. O'Donohue. Molecular design of a 'non-Leloir' furanose-transferring enzyme from an α-L-arabinofuranosidase. *Ready to be submitted.*

[6] **Bastien Bissaro**, Pierre Monsan, Régis Fauré and Michael J. O'Donohue. Glycosynthesis in a Waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Critical review invited to be submitted in Biochemical Journal*

[7] **Bastien Bissaro**, Hélène Débat, Olivier Saurel, Faten Arab-Jaziri, Claire Dumon, Alain Milon, Régis Fauré and Michael J. O'Donohue. Acceptor anomeric control of the transfuranosylation regioselectivity in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *In preparation*.

[8] **Bastien Bissaro**, Xevi Biarnés, Pierre Monsan, Régis Fauré, Michael J. O'Donohue and Antoni Planas. Predicting Mutational "Hot-spots" for Hydrolysis/Transglycosylation Modulation in *Tx*Abf using *BindScan* Algorithm. *In preparation*.

[9] **Bastien Bissaro**, Yuval Shoham, Pierre Monsan, Régis Fauré and Michael J. O'Donohue. Toward Artificial Arabinoxylans. *In preparation*.

* Co-first author.

POSTER AND ORAL COMMUNICATIONS

[I] **Study of hydrolysis/transglycosylation partition of the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus***

Bissaro B., Arab F., Saurel O., Barbe S., Débat H., Gervais V., André I., Milon A., Dumon C., Fauré R., O'Donohue M.

Congrès de la Société Française de Biochimie et Biologie Moléculaire (SFBBM), Ax-les-Thermes, France, 12-14 Octobre 2011 – Flash communication (short talk and poster)

SEVAB Doctoral School Conferences - October 2011

Poster Award (SEVAB d'Or)

[II] **New chemo-enzymatic routes for furanoses-based compounds synthesis**

Bissaro B., Arab-Jaziri F., Fauré R., O'Donohue M.

Renewable Resources and Biorefineries (RRB8), Toulouse, France, 04-06 June 2012

Poster

[III] **Furanosidase directed evolution for pentose-based compounds synthesis**

Bissaro B., Arab-Jaziri F., Monsan P., Fauré R. and O'Donohue M.

Plant Seaweed and Polysaccharides Symposium (PSP), Nantes, France, 17-20 July 2012

Oral communication

[IV] **Investigating the transglycosylation mechanism of the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus***

Bissaro B., Saurel O., Arab-Jaziri F., Saulnier L.; Tenkannen M., Milon A., Fauré R., Monsan P., O'Donohue M.

10th Carbohydrate Bioengineering Meeting (CBM), Prague, Czech Republic, 21-24 April 2013

Oral Flash communication and poster

Poster Award

[V] **Creation of the first non-Leloir transarabinofuranosylases using a molecular evolution strategy applied to a GH51**

Bissaro B., Durand J., Biarnés X., Planas A., Monsan P., Fauré R. and O'Donohue M.

27th International Carbohydrate Symposium (ICS), Bangalore, India, 12-17 January 2014

Oral communication

PENSÉES

Cette partie de ma thèse est peut-être la plus « importante », libre de toute convention, ou presque… Elle est le reflet de cette aventure, dure, quelques fois ingrate, parfois désespérante, souvent excitante et plus rarement jubilatoire. Ce sont ces derniers moments qui justifient certainement la nature initiatique de ce rite qu'est la thèse.

Je tiens à remercier en premier lieu Mike et Régis pour m'avoir donné « carte blanche » tout au long de cette thèse. Une liberté que je n'ai peut-être pas toujours su apprécier à sa juste valeur mais qui a eu son importance dans la manière de mener ma barque. Plus personnellement, Mike tu m'avais dit il y a plus de 3 ans de ça « la thèse est une aventure, la question scientifique doit rester le phare qui te guidera ». Si je devais donner un conseil précieux à ce qui liront ces quelques lignes et seront sur le point de se lancer dans cette aventure de Science c'est précisément celui-là. Pour sortir du brouillard, cherchez la question avant la réponse. Mike, tu impressionnes les étudiants, et les autres, par ta culture scientifique et est souvent de très bon conseil, tant dans les moments d'allégresse scientifique que dans ceux de doutes plus personnels. Nous avons vraiment eu de la chance de t'avoir comme « guide ». Un grand merci du fond du cœur pour tout ça.

Régis, malgré certains désaccords et prises de tête, je pense que nous aurons beaucoup appris, autant l'un que l'autre. Tu étais la première digue à supporter mon humeur parfois assassine, mais tu es toujours resté jovial, positif et disponible face à une tête à claques comme moi. Des qualités essentielles dont j'essaierai de m'inspirer à l'avenir. Je sais que mon insolence (je te laisse continuer la liste de mes « qualités » ;) !) laissera probablement une marque indélébile mais sache que tes qualités humaines (pas si communes que cela) de communication, d'écoute, de souci tout simplement pour tes « subordonnés », j'ai su les apprécier même si ça ne s'est pas toujours vu ! Merci Régis.

Pierre, ce fut un réel honneur d'être ton dernier thésard. Je le suis devenu un peu par la force des choses au début, mais tu as pleinement joué ton rôle et a répondu présent à chaque étape clés malgré tes multiples fonctions et occupations. Merci pour tes encouragements permanents, tes conseils aussi bien scientifiques que professionnels. C'était un vrai plaisir de t'avoir à nos côtés au cours de cette aventure.

I would like also to thank the members of the jury, gathered for my thesis defense, for reading and evaluating my manuscript. I want to give a big thank to Profs Leila Lo Leggio and Marco Moracci, as main opponents, for having spent time on my work and advised some possible improvements. Thank you to Prof. Morten Sørlie for having come to my thesis defense and for the cautious reading of my manuscript as highlighted by the relevant remarks done on it. Merci également au Dr. Monique Axelos qui a accepté d'examiner mon travail et qui m'a soutenu pour la construction de mon projet post-thèse. Enfin, merci à Toni, pour avoir pu être présent lors de ma soutenance et en avoir présidé le jury. Un grand merci pour le travail fourni lors de notre collaboration, avec des remarques et suggestions toujours très pertinentes ou éclairantes. Au passage, je remercie également Xevi pour avoir investi du temps dans la *Tx*Abf et permis de révéler des positions secrètes! C'était vraiment super de travailler avec vous.

Dans la joie, la bonne humeur, les coups de gueules, les moments de déprime, les journées rivés à la paillasse en mode « karaoké », c'est avant tout en équipe, entre collègues qui deviennent amis, que l'on traverse tous ces instants. C'est donc sans ordre prédéfini, soumis à une distribution totalement stochastique (comme l'agacement de mes pensées…) que je tiens du fond du cœur à

remercier les personnes suivantes :

Cyrielle, Vinciane, Capri c'est vraiment fini [;(]. Notre périple Islandais, mais aussi nos sorties toulousaines, le TaeKwonDo, et j'en passe, laisseront des souvenirs impérissables.

Los pablos, je sais le 15 septembre je suis « mort » (i.e. en espagnol : coma éthylique programmé pour festivités non contrôlées). La postérité en jugera. Quoi qu'il en soit vous aurez été de très bons compères, des « puto jefes » je dirais même plus !

Greg, ces quelques lignes ne suffiront pas pour exprimer le bonheur que j'ai eu de travailler/râler/déprimer/boire/faire la fête à tes côtés. Putin de bordel de merde ça va vraiment me manquer tout ça ! Et puis, après plusieurs années d'efforts la « kikiki » attitude portera peut-être ses fruits… ! On aura commencé et terminé cette aventure quasiment en même temps, tout bien réfléchi, et tu auras soutenu 3 jours avant moi, donc tout est en ordre !

Thomas, « je te tiens au courant » pour les remerciements… ;). Je vais devoir être gentil, tu détiens trop d'éléments compromettants à mon encontre. T'es vraiment un pote, j'espère qu'on se reverra dans nos futures vies.

Julien, le statut de stagiaire perpétuel est révolu ! Le tyran s'en va ;). On aura bien appris tous les deux, c'était génial de travailler avec toi, probablement la période la plus enrichissante de ma thèse. Je reviendrai voir toute la petite famille un de ces jours ! (pour ta soutenance…?! j'entends bien être informé, n'est-ce pas ?)

Seydou, sur l'algorithme de l'amitié je pense que nous avons passé les étapes clés, mais il faudrait demander à Sheldon pour en être bien sûr ! On aura partagé cette dernière année de thèse, la rédaction de ces quelques pages de science, au labo et en colloc', en bon compagnons de galère !

Alizée, l'INSA c'est fini (pour l'instant du moins… !), on se sera auto-supportés pendant 7 ans, un vrai chapitre de vie qui se termine, avec en conclusion une amitié que le temps ne saura éroder. A très vite !

Vous allez trop me manquer les copains ;(. J'espère que l'on se reverra bientôt !

Laurie, tout un roman, ou un poème plutôt, qui ne se lit pas mais se vit. Je suis presque sûr que tu as failli aimer la transglycosylation à la fin !

Les stagiaires de l'équipe « Lignos » qui m'ont plus vu en rédaction qu'en manips, Louise et Guillaume, qui ont supporté ma gentillesse légendaire, mes petits mots agréables. Il faut bien respecter les traditions.

La « RU team » (Marjo, Paupau, Vince, Lannig, Marie and co. précédemment cités), ces repas vont vraiment me manquer (dans son acception sociale et rituelle j'entends, mais n'y voyez aucunement une critique négative de la gastronomie RUesque).

Marie-Laure, Hélène, Susana et Faten, la team des « anciennes » (no offense), qui ont été là pour me former alors que je n'étais qu'un stagiaire « en carton » (merci Susana ;) pour la dénomination).

Sophie, Claire, Cédric, le cœur des Lignos, chacun à votre façon vous avez su m'apporter de votre expérience, même si je n'en fait souvent qu'à ma tête.

Mes profs de l'INSA, Pierre M., Mag R.-S., Didier C., Claude M., Alain M., Sandrine M., vous avez participé en grande partie à ma formation et mon orientation au cours de ces 8 dernières années, merci pour ça.

Toute l'EAD1, la bonne ambiance qui y règne, et notamment Mag pour ta bienveillance tant appréciée au sein de notre équipe, on ne le dit pas assez alors autant l'écrire.

Mes collaborateurs, notamment Olivier Saurel et Alain Milon, merci de m'avoir initié à la STD

RMN et pour nous avoir aidé au tout début des cinétiques par RMN. Merci pour avoir partagé un peu de vos connaissances et de votre expertise, nos conversations étaient toujours très enrichissantes.

Les membres de mon comité de thèse, Alain Milon, Cécile Albenne et Sylvain Cottaz, je m'excuse de vous avoir assommés avec mes rapports de comité, c'était pour la bonne cause. Je vous remercie d'avoir pris part à cette aventure à travers vos conseils.

*Tx*Abf, illustre membre des 51, chiffre ô combien providentiel, malgré moult tortures tu ne nous a pas encore livrée tous tes secrets mais tu n'auras pas été trop ingrate au cours de ces dernières années. J'attends de tes nouvelles…

Je remercie aussi les urgences ophtalmiques de Purpan (l'histoire d'un mec qui prenait en photo des UVs…oh le con penseront certains), Peco pour m'avoir initié au branchement de colonnes (malgré une explosion/désintrégation mémorable, je savais bien qu'il fallait arrêter le débit avant que le piston n'arrive au bout…), Vinciane pour ne pas m'en avoir voulu pour le bouchon d'ampoule à brome (expérimentalement vérifié, le caoutchouc fond à 250 degrés…).

Je remercie le monsieur des « Flots » pour nous avoir alimentés en sandwich (bon il faut vraiment faire quelque chose sur la taille des sandwichs par contre…)

Je remercie le bain à sec pour ses bons et loyaux services. Le TECAN aussi.

Je ne remercie pas le petit lutin du labo qui vide les tiroirs, qui planque les ciseaux, qui se siffle les pissettes d'éthanol (d'eau parfois à l'occasion), qui finit (on ne sait pourquoi) tous les stocks d'ITPG et d'antibiotique, qui saupoudre les balances de poudres, qui vole les marqueurs.

Je ne remercie pas Visilog.

Merci à mes amis de toujours, Le noble, Canard, Nicoco, PiouPiou, Néné, mais aussi Laure, Sab, Mazec et Adou, nos «obligations» respectives nous séparent mais l'esprit y est. Il y a très certainement un peu du votre entre les lignes de ce manuscrit.

Ma sœur, ma mère, mon père, mes grands-parents. Ma famille. Cette thèse est un peu la votre aussi. Le fruit d'un soutien indéfectible.

J'ai probablement oublié des personnes dans le tumulte de ces derniers jours, j'espère que vous ne m'en tiendrez pas rigueur. Bonne lecture au plus courageux d'entre vous et, si un néo-thésard tu es, que la force soit avec toi.

TABLE OF CONTENTS

Table of Abbreviations and Definitions

Abbreviations

Abfs : α-L-Arabinofuranosidases
AraT : Arabinosyl transferase activity
AXs : ArabinoXylans
AXH : Arabinoxylan ArabinofuranoHydrolase
AXOS : ArabinoXylo-Oligosaccharides
Bn α-L-Ara*f*-(1,2)α-D-Xyl*p* : Benzyl α-L-arabinofuranosyl-(1→2)-α-D-xylopyranoside
Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p* : Benzyl β-D-galactofuranosyl-(1→2)-α-D-xylopyranoside
Bn-α-D-Xyl*p* : Benzyl α-D-xylopyranoside
CAZy : Carbohydrate Active Enzymes
DP : Degree of Polymerization
ENGase : *Endo*-β-*N*-acetylglucosaminidase
α-D-Gal*p* : α-D-Galactopyranose
β-D-Gal*f* : β-D-Galactofuranose
β-D-Gal*p* : β-D-Galactopyranose
Fe : Feruloyl moiety
FHs : Furanoside Hydrolases
FOS : Fructo-Oligosaccharides
GAXs : Glucurono-ArabinoXylans
G-Enz : Covalent Glycosyl-Enzyme intermediate
GHs : Glycoside Hydrolases
Glc : D-Glucose
GlcNAc : *N*-Acetyl D-glucosamine
GOS : Galacto-Oligosaccharides
GlcAT : Glucuronic acid transfer activity
H_I : Primary Hydrolysis
H_{II} : Secondary Hydrolysis
GTs : Glycosyltransferases
HTS : High-Throughput Screening
ITC : Isothermal Titration Calorimetry
IUBMB : International Union of Biochemistry and Molecular Biology
k_{cat} : Catalytic constant (s^{-1})
K_M : Michaelis-Menten constant (mM)
LG : Leaving group (e.g., *p*NP)
MU : 4-methylumbelliferyl
NDOs : Non Digestible Oligosaccharides
(NR)-unit : Non Reducing end unit
***p*NP** : *para*-nitrophenol
p*NP-α-L-Ara*f : *para*-nitrophenyl α-L-arabinofuranoside

p*NP-β-D-Gal*f	: *para*-nitrophenyl β-D-galactofuranoside
p*NP-β-D-Xyl*p	: *para*-nitrophenyl β-D-xylopyranoside
(Red)-unit	: Reducing end unit
Rho value	: Synthesis/Hydrolysis rate of transglycosylation product
RPM	: Rotation Per Minute
R_T	: Transfer rate (% or µmoles of product for 100 µmoles of consumed donor)
S.A.	: Specific Activity (µmoles of *p*NP released.min^{-1}.mg of enzyme^{-1})
STD NMR	: Saturation Transfer Difference Nuclear Magnetic Resonance
***Tx*Abf**	: α-L-Arabinofuranosidase from *Thermobacillus xylanilyticus*
T/H	: Transglycosylation/Hydrolysis ratio
IU	: International Unit (µmoles of *p*NP released.min^{-1})
V_{max}	: maximum hydrolytic rate (µM.min^{-1})
ν_T	: Apparent transglycosylation rate (%/min)
ν_{HII}	: Apparent secondary hydrolysis rate (%/min)
wt	: wild-type enzyme
X	: Donor conversion rate (%)
X_2	: xylobiose
X_3	: xylotriose
XA^3XX	: β-D-Xylopyranosyl-(1→4)-[α-L-arabinofuranosyl-(1→3)]-β-D-xylopyranosyl-(1→4)-β-D-xylopyranosyl-(1→4)-D-xylopyranose
X-α-L-Ara*f*	: 5-bromo-4-chloro-3-indolyl α-L-arabinofuranoside
XET	: Xyloglucan *endo*-transglycosylase
XOS	: Xylo-OligoSaccharides
XTH	: Xyloglucan *endo*-transglygosylases/hydrolase
XylT	: Xylosyl transferase activity
Y	: Yield (%)

Definitions

H_I and H_{II}, or primary and secondary hydrolysis: According to the canonical double-displacement mechanism of retaining glycoside hydrolases (GHs), after the glycosylation step, the glycosyl-enzyme intermediate can be deglycosylated by an acceptor. When the latter is water, then hydrolysis occurs. Whether hydrolysis is defined as primary or secondary depends on the origin of the donor glycosyl moiety. When the glycosyl moiety is that of the original donor substrate, the reaction is qualified as primary hydrolysis (H_I). However, when a transglycosylation product acts as a donor compound, then it is appropriate to use the term secondary hydrolysis (H_{II}).

Apparent rates of transglycosylation (ν_T) and secondary hydrolysis (ν_{HII}): When the evolution of the transglycosylation product yield is plotted as a function of time, one can estimate the apparent transglycosylation rate by measuring the slope (expressed in % of $[donor]_0$ or mM of product/min) at

the beginning of the reaction. This rate is denoted as “apparent” since the accumulation of transglycosylation product is meant to reflect two undistinguishable phenomena, transfer and secondary hydrolysis, although it is usually assumed that at the beginning of the reaction transglycosylation is the dominant activity. Similarly, measuring the slope of the time-dependent transglycosylation product yield plot at the end of the reaction, once its maximum yield has been reached, allows to evaluate its decay rate, called apparent secondary hydrolysis rate or ν_{HII}. At that stage it is also possible that transglycosylation still occurs.

Rho value: Using the two aforementioned parameters, ν_T and ν_{HII} derived from a single graph, it is possible to calculate an apparent synthesis/hydrolysis rate for a given transglycosylation product (provided that regions with linear time-dependent yield profiles can be obtained). The Rho value, combined with absolute transglycosylation yields, provides the means to define the ability of a mutant enzyme to perform synthetic reactions. This is because mutants that are useful as glycosynthetic enzymes display good transfer rates and low secondary hydrolysis.

R_T: In the course of this thesis, the comparison of the behavior of enzyme mutants in a standardized time-independent way became a necessity, because it was often necessary to use different quantities of enzymes (between wild-type and mutants). To achieve this, transglycosylation product yield was plotted as a function of donor substrate conversion (ranging between 0 to 100%). Therefore, deriving the slope from such a plot provides the so-called transfer rate, or R_T (expressed as a percentage). In other words, for 100 molecules of donor substrate consumed, this rate tells us how many donor glycosyl moities were present in transglycosylation products. In chemistry, R_T is known as a selectivity factor (e.g., between a carbohydrate acceptor *versus* water).

T/H : Throughout this manuscript, the transglycosylation/hydrolysis ratio is denoted as T/H or H/T, depending on the natural function of the wild-type enzyme under investigation. T/H is preferentially employed in the case of wild-type transglycosylases, whereas H/T was deemed more appropriate to describe hydrolases. The exception to this is when transglycosylation is improved in a wild-type hydrolytic GH since in that case it seems more convenient and intuitive to talk about a T/H increase rather than a H/T decrease.

INTRODUCTION

Nature is polyglot and Glycochemistry is one of the most beautiful and sophisticated languages that it practices. The diversity of the “sugar alphabet” and the way each letter can be associated makes the “Glyctionnary” vocabulary almost infinite, which reflects the enormous number of interactions that are mediated by carbohydrates. If one can liken classical glycochemistry to the way in which people express themselves in a foreign tongue, then glyco-enzymology can be likened to native speakers, since carbohydrate-active enzymes exquisitely master Nature’s idiom. In order to better understand the language of sugars and to decipher Nature’s codes and subtleties, it is essential to be able to make the ‘words’ and rather like in living languages, invent new ones. This is the quest of glycochemists who have struggled with the elementaries of this language for many decades, but are now acquiring greater mastery through the use of enzymes.

Several milestones have marked the essor of enzymology, from the cell-free observation of reactions catalysed by cell-free extracts (Payen, 1833, Kühne, 1877, Buchner, 1897) and the discovery of the proteinaceous nature of enzymes (Sumner, 1926, Northrop and Stanley, 1930), through to the elucidation of three-dimensional structures (Kendrew, 1963, Perutz, 1963 and Phillips, 1965) and catalytic mechanisms (Sinnott, 1990), and finally with their exploitation in glycoscience (Mackenzie et al., 1998; Malet et al., 1998) and the emerging ability to engineer enzymes using predictive computational approaches (Lutz, 2010).

In the field of synthetic glycochemistry, Glycosides Hydrolases (GHs) are proving to be useful tools to build glycomotifs. This is because these enzymes are very widespread, covering an enormous range of substrate and linkage specificties, and have often proved to be amenable to heterologous production. Moreover, GHs often use quite simple, affordable substrates and operate in aqueous conditions, low to moderate temperatures and in moderate pH range. Even more importantly from a synthetic standpoint, certain GHs that operate via a double displacement mechanism display the innate ability to catalyse both hydrolysis and transglycosylation, which represent two mechanistic outcomes of glycosidic bond breaking. Generally, transglycosylation is a minor occurrence in GHs and is often considered as a parasitic side reaction. However, in glycosynthesis it is a sought after property of GHs. Therefore, when GHs are used in glycosynthetic strategies the main challenge is to enhance and/or modulate transglycosylation, whilst simultaneously supressing hydrolysis, the latter being no easy task. The quest to diminish or eliminate hydrolysis in GHs has profited from the introduction of ever more powerful enzyme engineering techniques. However, unfortunately the rules that govern the hydrolysis/transglycosylation (H/T) partition in GHs are still unclear and thus the creation of glycosynthetic enzymes from GHs is not a straightforward matter, often relying on a hit and miss strategy. This is all the more frustrating, because Nature has furnished some intriguing examples of GHs that are natural transglycosylases, displaying little or no ability to catalyse

hydrolysis, while being highly related to essentially 'hydrolytic' GHs.

In the course of this scientific adventure, we present a study in which we have attempted to distill the messages from the considerable wealth of information that has been provided by predecessors and peers in order to understand how highly-related GHs can display quite different H/T equilibria. One key element in the process of tackling this challenge is to recall that hydrolysis is the inevitable outcome of reactions that occur in aqueous environments in which water is overwhelmingly present. Therefore, to understand how the outcome can be otherwise (i.e. a sugar moiety is transferred to an acceptor other than water) it is pertinent to ask **how evolution has provided certain GHs displaying the ability to efficiently perform transglycosylation in the presence of water**. Unfortunately, the study of these exceptional GHs has not provided the answer to this question. Therefore, another way to provide the answer is to evolve a hydrolytic GH, selecting for acquired ability to perform transglycosylation and monitor the molecular changes as they occur along the evolutionary pathway. This is the strategy that has been adopted in this work.

In this work, we have used the family GH51 α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*Tx*Abf) as the enzyme model to tackle the aforementioned challenge. In addition to its many practical attributes (i.e. its 3D structure is known, it is thermostable and is readily produced in *E. coli*), this enzyme is exciting because it is a furanoside hydrolase (FH) and as such can be considered as a potential tool for glycosynthetic strategies involving furanoses, which are the trickiest of sugars to handle using classical glycochemistry. In this respect, very little work has been performed on **GH-catalyzed transfuranosylation**, apart from the pionnering studies that have been performed in our laboratory. Therefore, the description of the different engineering steps that have led us to create in this work the first 'non-Leloir' transarabinofuranosylase is of major importance.

In order to describe how our work has provided new insight into how hydrolytic GHs can be converted into transglycosylases, this manuscript is divided into three main chapters, which are dedicated to (I) a state-of-the-art review followed by a description of the main goals and challenges of the study, (II) the results and discussion, presented as research articles that are supplemented with complementary information and (III) a summary and conclusions, which aim to provide the reader with a clear overview. Regarding chapter I, this is divided into three parts, the first providing a review of the accumulated knowledge on GH-catalyzed transglycosylation, including a ready-to-publish review on how Nature has provided GHs that can perform glycosynthesis in aqueous medium. The second part focuses on the specific and rather understudied area of glycosynthesis using pentose, and creates a connection with the challenges inherent to the bioeconomy. Finally, the third part provides an update on what is known about *Tx*Abf, reviewing the work of the O'Donohue group over the past 14 years. Within this section, work that I performed during my Master's degree

(January-July 2011), focused on the functional role of the β2α2 mobile loop and on the creation of mutants displaying increased transglycosylation are presented. Although the latter results do not constitute the core of the work that I performed at this time, they are included because they provide a link to the work that is presented in the rest of the thesis and in the related publications presented as supporting information (Supporting A1, 2 and 3). Chapter II of this thesis is divided into five main parts, which together present the corpus of results obtained during my doctoral studies. Part I presents a study of the determinants of regioselectivity in *Tx*Abf-catalyzed transglycosylation reactions. In this study a determinant of regiospecificity was identified. Building on some preliminary work that was described in the PhD thesis of Dr. Faten Arab-Jaziri, Part II describes the first steps towards the creation of a *Tx*Abf-based transfuranosylases, while Part III describes the use of a predictive *in silico* approach that was perfomed in collaboration with Prof. Antoni Planas and Dr. Xevi Biarnés (IQS, Barcelona). This study was designed to predict mutations that would increase acceptor recognition and thereby favor transglycosylation. In the meantime a new random mutagenesis round, coupled to purpose-made screening strategy, led to the identification of a key H/T residue modulator, as revealed in Part IV. The latter describes how the strategy outlined in Part II was followed up, using recombination to create new generations of mutants, containing mutations that had been previously identified by random mutagenesis and others identified by semi-rational engineering (Part I) and *in silico* predictions (Part III). This work led to the rollout of an artificial molecular evolution pathway that procured first-of-a-kind 'non-Leloir' transarabinofuranosylases. These were thoroughly analyzed, especially from a mechanistic point of view. Finally, Part V describes the more applicative facet of this PhD thesis. Indeed, it describes how the enzyme tools created were used to synthesize hemicellulose-like, complex structures via original chemo-enzymatic strategies, which demonstrate some of the opportunities that are offered by the results of this work.

References

Lutz, S. (2010) Beyond directed evolution--semi-rational protein engineering and design. *Curr. Opin. Biotechnol.*, 21 (6), pp.734–743.

Mackenzie, L.F., Wang, Q., Warren, R.A.J. & Withers, S.G. (1998) Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.*, 120 (22), pp.5583–5584.

Malet, C. & Planas, A. (1998) From β-glucanase to β-glucansynthase: glycosyl transfer to α-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS lett.*, 440 (1-2), pp.208–212.

Sinnott, M.L. (1990) Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.*, 90 (7), pp.1171–1202.

First chapter:

State-of-the-Art

Part I

TO TRANSFER OR NOT TO TRANSFER, THIS IS THE QUESTION…

Carbohydrates are ubiquitous in Nature and fulfill vital biological functions. For this reason, it is of prime importance to be able to synthesize *in vitro* complex carbohydrates and glyco-conjugates to satisfy a wide variety of needs. Unfortunately, because of the complex nature of carbohydrates, sugar-based compounds are often difficult to access via conventional organic chemistry and the preparation of structurally more elaborate carbohydrate-based compounds can still be a tricky and tedious task, despite the outstanding progress in the field (*c.f*. section I). Nevertheless, coupling chemistry and enzymology is a promising avenue for development and indeed several successes using glycoside hydrolases (GHs) have already been reported. The main features of GHs and the way in which these enzymes can be exploited for glycosynthesis are described in section II, while section III is dedicated to a review of accumulated knowledge on the underlying (common) origins of hydrolytic and transglycosylation activities in different GHs families.

I-1. Glycosynthesis: Stakes and Challenges

Although nucleic acids and protein are the primary targets of scientific investigation in the life sciences field, carbohydrates and glycosynthesis have nevertheless attracted interest for a long time with the first glycosylation reaction being described in 1879 by the American organic chemist Arthur Michael. Since then, other renowned chemists, such as Emil Fischer (acid catalysis, thioglycosides), Wilhelm Koenigs and Edward Knorr (halides glycoside donor) have provided valuable contributions to the field of glycosynthesis using classical organic chemistry. Being widespread, vital components of living organisms, carbohydrates are extremely complex, since unlike DNA, carbohydrates can be formed from a multitude of different monosaccharides that can be present in the form of either D- or L- stereoisomers. As an illustration of the structural diversity that can be reached, by playing on anomeric stereochemistry, ring cyclization and linkage regioselectivity, more than 100 combinations can be obtained from two simple carbohydrates! Moreover, the assembly of monosaccharides into larger molecules containing two or more sugar moieties is achieved through the formation of glycosidic bonds of diverse nature (e.g. *O*-, *S*- or *N*-linkage). Such chemical complexity gives rise to the existence of an infinite number of carbohydrate-containing macromolecular structures, which intervene in a tremendously varied number of biological contexts where they participate in reactions, interactions, energetics and structures (Varki et al., 2009). Mastering complex sugar-based assemblies represents a field of research in itself, which is of great importance for the development of, for example, chemotherapeutics, bio-inspired materials, detergents etc. Recently, great advances have been made in the development of methods for the synthesis of complex oligosaccharides and glycoconjugates of biomedical relevance by research groups headed by Geert Jan Boons, Chi-Huey Wong, Alexei V. Demchenko or Stephen G. Withers to cite but a few. However, in spite of intensive efforts, many questions still remain unanswered, notably regarding the underlying mechanisms that control the formation of glycosidic bonds (e.g. stereo- and regio-selectivity), although many determinants are now well known (c.f. the section hereafter).

1. On the Complexity of Chemical Glycosylation

Generally speaking chemical glycosylation (Fig. 1) designates the linkage of a glycosyl donor to an acceptor, leading to the formation of a glycoside product, or an oligosaccharide when the acceptor is a sugar. Generally, this reaction can be described by four main steps: (1) The formation of a donor-promoter complex, which occurs when the glycosyl donor bears a nucleophilic leaving group (LG) and the promoter displays electrophilic character; (2) ionization of the glycosyl donor, forming an oxocarbenium intermediate that displays sp^2 hybridization at *C*-1; (3) nucleophilic attack on the reactive anomeric carbon by an exogenous acceptor; and (4) the transfer of a proton, leading to the

formation of the neutral glycoside end-product.

Within each of these steps one can identify factors that will influence the outcome of glycosylation. The step leading to the oxocarbenium intermediate is governed by the mode of activation of the LG by a promoter, which can be denoted as direct, remote or bidentate (mix of both) (Ranade et al., 2013). This step is also strongly influenced by the nature of protecting groups (PO) present on the donor sugar ring and the conformation of the latter. Protecting groups can have "arming" (e.g. ethers) or "disarming" (e.g. esters) effects with respect to the reactivity of the LG, these being determined by the electron withdrawing character of the PO. The exact nature of the effect exerced by the PO on the LG will influence the outcome of glycosylation (Mydock et al., 2010). The impact of PO number and type on the oxocarbenium intermediate stabilization are also well known, dictating to some extent the stereoselectivity of the nucleophilic attack by the acceptor molecule. For instance, neighboring group participation with an acetyl PO at *C*-2 position leads to the formation of an acetoxonium intermediate blocking the bottom face of the anomeric carbon and leading to a 1,2-trans glycoside product. The stereoelectronic effect of sugar ring substituents is also a feature that glycochemists are aware of (Jensen et al., 2006), with for example axial *C*-3 or *C*-4 substituents increasing stabilization.

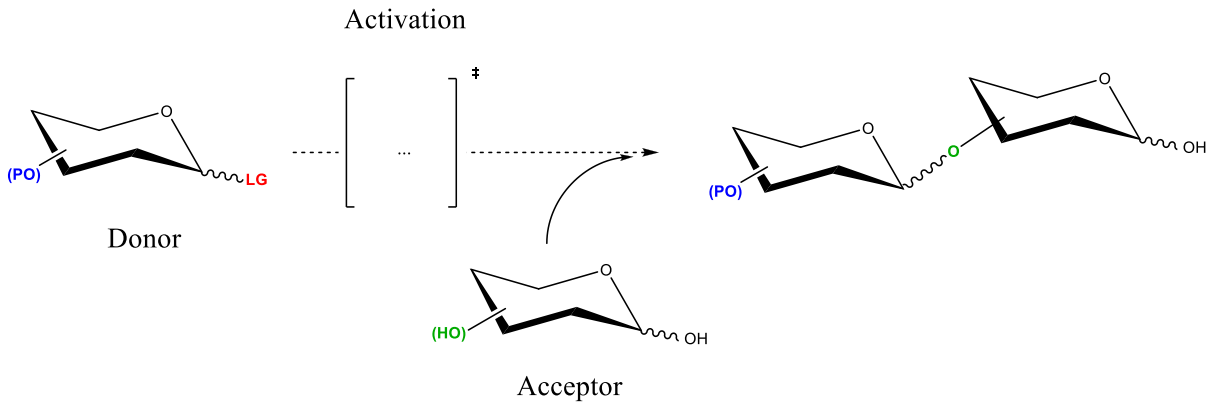


Fig. 1: Chemical glycosylation general principle. A promoter activates the leaving group (LG) of the glycosyl donor, the hydroxyl functions of which are protected (PO). An oxocarbenium intermediate (or several possible intermediate) is then formed and the anomeric carbon is the target of a nucleophilic attack by the acceptor leading to the glycoside product after deprotonation.

In addition to electronic considerations, steric issues must also be taken into account, since they can hinder access to one of the anomeric carbon faces (α or β) to be attacked. In addition to these factors, it is important to note that an essential feature of carbohydrates is the "anomeric", or Edward-Lemieux effect. If one considers the chair conformation of cyclohexane rings, the inherent steric hindrance favors an equatorial positioning of substituents, which leads to an energetically more stable conformation (example of methoxy subsituent on cyclohexane). However, in sugars, the

presence of the *endo*-cyclic oxygen can allow the anomeric subsituent to adopt less conventional axial configuration. This phenomenon probably arises from (a) an electrostatic repulsive effect and (b) a hyperconjugation due to an anti-periplanar alignment between the *O*-5 lone pair of electrons and the anti-bonding orbital of *C*-1, which is not possible in the case of equatorial positioning (Fig. 2). Depending on the nature of the substituents present on the sugar ring this anomeric affect will therefore also influence the stereoselectivity outcome of the glycosylation reaction.

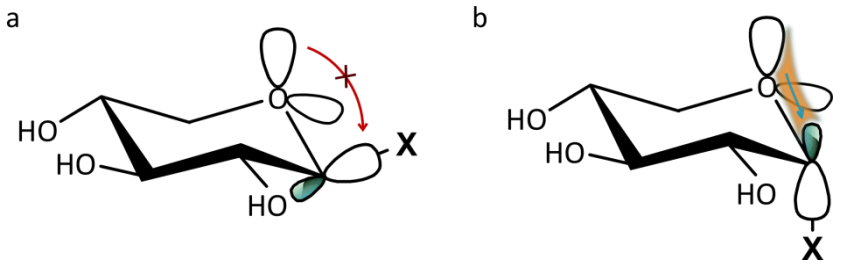


Fig. 2: Anomeric Effect. The electrostatic repulsive effect (a) between the heterom atom X and the *endo*-cyclic oxygen lone pair favors axial configuration, which is reinforced by a hyperconjugation effect (b).

Regarding the last step of deprotonation, little is known about its features since it is considered as a non-determining phase with respect to glycosylation end product stereochemistry.

So far, most work has focused on developing new LG, PO, promoters, and modifying reaction conditions. However, a better understanding of the underlying energetics and thermodynamics will be necessary in the future in order to advance glycochemistry. In this respect, the study of efficient systems that already exist in Nature will be a worthwhile pursuit and, in particular, structure-function studies of glycosynthetic enzymes that establish complex interactions with their sugar substrates to control stereo- and regio-selectivity for the formation of glycosidic bonds will be a rich source of information.

2. Opportunities offered by chemo-enzymatic approaches

As mentioned earlier, the need for well-defined molecules with perfectly controlled stereo- and regio-selectivity, cannot always be satisfied using classic organic chemistry. Indeed, such approaches often require selective protection/deprotection steps, and are complicated by the imperfect control of the configuration of the anomeric center, making subsequent anomeric resolution necessary. Apart from carbohydrates that are extracted from biological material, the preparation of oligo or polysaccharides often procures product mixtures displaying polydispersity (variable degree of polymerization, DP) and/or mixed linkages.

To circumvent the limits of synthetic glycochemistry, since the 1990's chemists have increasingly turned their attention towards the development of chemo-enzymatic strategies, employing enzyme-catalyzed transglycosylation (Shoda et al., 1993; Crout et al., 1998; Palcic, 1999). These powerful

methods have progressively provided the means to prepare ever more complex, tailor-made oligosaccharides (Perugino et al., 2005) and glyco-conjugates (Desmet et al., 2012). Moreover, enzyme-catalyzed reactions are available for the preparation of naturally-occurring (e.g. cellulose, xylans, chitin) and artificial polysaccharides (Kadokawa, 2011; Kobayashi & Makino, 2009). Finally, transglycosylation has proved to be a useful method for the glycosylation of amino acids, a process that is necessary for the preparation of glycoproteins (Nilsson et al., 1997), compounds that are components of cellular receptors and thus of utmost interest for therapeutic applications.

Regarding the nature of the enzymatic tools employed in chemo-enzymatic strategies, glycosyl transferases (GTs) (Lairson et al., 2008; Breton et al., 2012) or glycoside hydrolases (GHs) (Schmaltz et al., 2011) are the most widely exploited biocatalyts. GTs constitute Nature's principal glycosynthetic tools, but hydrolytic GHs that nevertheless display the ability to catalyze transglycosylation (i.e. principally retaining GHs) are often adopted by glycochemists. At first sight, this choice of tools may appear to be somewhat strange. However, GHs are widespread, robust (many are thermostable) enzymes that have proved to be quite amenable to production in heterologous expression systems. Moreover, although in recent years chemists have solved some of the challenges linked to the preparation of nucleotide sugars (i.e. the donors for GTs), this has long been a handicap (Field, 2011; Gantt et al., 2011). Therefore, the fact that GHs use quite simple inexpensive substrates (e.g. *p*NP-glycosides) is seen as a major advantage. Another clear advantage, is the extreme diversity of GHs, acting on different sugar substrates and offering the ability to synthesize a wide range of glycosidic bond types, and thus an almost unlimited array of glycosides (Wang et al., 2009; Schmaltz et al., 2011). Finally, it is most important to note that within the group of enzymes known as GHs, some enzymes exclusively or majoritarily catalyze the formation of glycosidic bonds (c.f. section I-3 of Chapter I). For example, using sucrose as a donor sugar, glucan sucrases can glucosylate a myriad of different acceptor, while showing no sign of hydrolytic activity (Monsan et al., 2010).

With regards to the preparation of tailor-made, branched oligosaccharides using chemo-enzymatic methods, some successes are noteworthy. These include the chemo-enzymatic synthesis of certain xyloglucan motifs that are found in plant cell wall (Fauré et al., 2006) or the preparation of antigenic oligosaccharides (Champion et al., 2009), with both examples demonstrating the power of this approach. Nevertheless, much progress is still to be made, notably to extend the use of chemo-enzymatic strategies to the preparation of rarer carbohydrate structures and ultimately to the synthesis of any natural or artificial carbohydrate or glycoconjugate.

Today, the use of chemo-enzymatic strategies for the preparation of hexose-containing compounds is already well-advanced, and many GHs are available to facilitate this. However, in the field of pentose sugars, less progress has been made, no doubt because of the lesser importance that is associated with this subclass of carbohydrates (Dumon et al., 2012). Moreover, when considering

carbohydrate from the point of view of sugar ring cyclisation, glycosynthesis involving furanose sugars is less advanced, despite the efforts of a few research groups in this area and despite new fundamental knowledge concerning the specific features of furanose rings (Taha et al., 2013). This PhD study is devoted to tackling chemo-enzymatic glycosynthesis in the field of pentose sugars, and specifically those in furanose form. Therefore, the current state of the art in this field will be described in Part II-3.

References

Breton, C., Fournel-Gigleux, S. & Palcic, M.M. (2012) Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.*, 22 (5), pp.540–549.

Champion, E., André, I., Moulis, C., Boutet, J., Descroix, K., Morel, S., Monsan, P., Mulard, L. a & Remaud-Siméon, M. (2009) Design of α-transglucosidases of controlled specificity for programmed chemoenzymatic synthesis of antigenic oligosaccharides. *J. Am. Chem. Soc.*, 131 (21), pp.7379–7389.

Crout, D.H.G. & Vic, G. (1998) Glycosidases and glycosyl transferases oligosaccharide synthesis. *Biocatal. Biotransfor.*, 2, pp.98–111.

Desmet, T., Soetaert, W., Bojarová, P., Křen, V., Dijkhuizen, L., Eastwick-Field, V. & Schiller, A. (2012) Enzymatic glycosylation of small molecules: challenging substrates require tailored catalysts. *Chem. Eur. J.*, 18 (35), pp.10786–10801.

Dumon, C., Song, L., Bozonnet, S., Fauré, R. & O'Donohue, M.J. (2012) Progress and future prospects for pentose-specific biocatalysts in biorefining. *Process Biochem.*, 47 (3), pp.346–357.

Fauré, R., Saura-Valls, M., Brumer, H., Planas, A., Cottaz, S. & Driguez, H. (2006) Synthesis of a library of xylogluco-oligosaccharides for active-site mapping of xyloglucan endo-transglycosylase. *J. Org. Chem.*, 71 (14), pp.5151–5161.

Field, R.A. (2011) Glycobiology: Challenging reaction equilibria. *Nat. Chem. Biol.*, 7 (10), pp.658–659.

Gantt, R.W., Peltier-Pain, P., Cournoyer, W.J. & Thorson, J.S. (2011) Using simple donors to drive the equilibria of glycosyltransferase-catalyzed reactions. *Nat. Chem. Biol.*, 7 (10), pp.685–691.

Jensen, H.H. & Bols, M. (2006) Stereoelectronic substituent effects. *Acc. Chem. Res.*, 39 (4), pp.259–265.

Kadokawa, J. (2011) Precision polysaccharide synthesis catalyzed by enzymes. *Chem. Rev.*, 111 (7), pp.4308–4345.

Kobayashi, S. & Makino, A. (2009) Enzymatic polymer synthesis: an opportunity for green polymer chemistry. *Chem. Rev.*, 109 (11), pp.5288–5353.

Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.*, 77, pp.521–555.

Monsan, P., Remaud-Siméon, M. & André, I. (2010) Transglucosidases as efficient tools for oligosaccharide and glucoconjugate synthesis. *Curr. Opin. Microbiol.*, 13 (3), pp.293–300.

Mydock, L.K. & Demchenko, A. V (2010) Mechanism of chemical *O*-glycosylation: from early studies to recent discoveries. *Org. Biomol. Chem.*, 8 (3), pp.497–510.

Nilsson, K.G.I., Ljunger, G. & Melin, P.M. (1997) Glycosidase-catalysed synthesis of glycosylated amino acids: synthesis of GalNAcα-Ser and GlcNAcβ-Ser derivatives. *Biotechnol. Lett.*, 19 (9), pp.889–892.

Palcic, M.M. (1999) Biocatalytic synthesis of oligosaccharides. *Curr. Opin. Biotechnol.*, 10 (6), pp.616–624.

Perugino, G., Cobucci-Ponzano, B., Rossi, M. & Moracci, M. (2005) Recent advances in the oligosaccharide synthesis promoted by catalytically engineered glycosidases. *Adv. Synth. Catal.*, 347 (7-8), pp.941–950.

Ranade, S.C. & Demchenko, A. V. (2013) Mechanism of Chemical Glycosylation: Focus on the Mode of Activation and Departure of Anomeric Leaving Groups. *J. Carbohydr. Chem.*, 32 (1), pp.1–43.

Schmaltz, R.M., Hanson, S.R. & Wong, C.-H. (2011) Enzymes in the synthesis of glycoconjugates. *Chem. Rev.*, 111 (7), pp.4259–4307.

Shoda, S., Kawasaki, T., Obata, K. & Kobayashi, S. (1993) A facile enzymatic synthesis of cellooligosaccharide derivatives using β-lactosyl fluoride. *Carbohydr. Res.*, 249 (1), pp.127–137.

Taha, H.A., Richards, M.R. & Lowary, T.L. (2013) Conformational analysis of furanoside-containing mono- and oligosaccharides. *Chem. Rev.*, 113 (3), pp.1851–1876.

Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. & Etzler, M.E. (2009) *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press.

Wang, L.-X. & Huang, W. (2009) Enzymatic transglycosylation for glycoconjugate synthesis. *Curr. Opin. Chem. Biol.*, 13 (5-6), pp.592–600.

I-2. Mechanism of Glycoside Hydrolases and mutants thereof

1. Definition

The IUBMB classification (EC w.x.y.z) primarily organizes enzymes according to the type of reaction that they catalyze: oxydoreductase (EC 1.), transferase (EC 2.), hydrolase (EC 3.), lyase (EC 4.), isomerase (EC 5.) and ligase (EC 6.). Further classification in sub-classes is defined by the chemical function on which the enzyme is active (EC w.**x**) and the acceptor involved (EC w.x.**y**). The final number (**z**) allows an accurate identification of the enzyme. Accordingly, GHs are enzymes that hydrolyze *O*-glycosylated compounds and are therefore classified in the EC 3.2.1.z group.

The extremely large number of quite different enzymes that fall under the generic term GH makes their classification particularly difficult, and is thus challenging for the IUMB system described above. To address this challenge, some years ago (Henrissat, 1991), B. Henrissat *et al.* developed a new sequence homology-based classification called CAZy, meaning Carbohydrate Active enZymes (http://www.cazy.org/). At the outset, this classification system was specifically dedicated to GHs, but it was quickly extended to other associated enzymes that also modify carbohydrates. Accordingly, nowadays the CAZy database is divided into several sections and as well as GHs and GTs, serves as a repertory for carbohydrate-esterases (CEs), polysaccharide-lyases (PLs), and auxiliary activities (AA) gathering essentially redox enzymes (Lombard et al., 2014). Over the years, CAZy has proven to be a useful tool for the prediction of enzyme structure, the localization of key catalytic residues, and to a lesser extent, for the prediction of enzyme function and substrate specificity (albeit muli-specificity/function can be observed within a single GH family). Moreover, CAZy is useful for phylogenetic studies notably for the investigation of enzyme evolution. In addition to the family structure, CAZy also employs a higher level organization, which bring families together into clans, which are defined on the basis of common 3-D folds ($(\beta/\alpha)_8$, β-jelly roll, $(\alpha/\alpha)_6$, α+β, 6-fold β-propeller, 5-fold β-propeller and β-helix) and catalytic machinery (Glu/Glu, Asp/Asp, Asp/Glu, for nucleophile/general acid-base catalytic residues or acid/base couples). At present[1], GHs are distributed into 133 different families and 14 different clans (from GH-A to N).

2. Catalytic mechanisms

2.1. Canonical mechanisms

By studying the effect of enzyme catalysis on the stereochemistry of substrate asymmetric carbons, in 1953 D.E. Koshland was able to describe two main enzyme mechanisms, based on whether the stereochemistry of the substrate asymmetric carbon is retained or inverted in the

[1] On the 7th of July, 2014

reaction product (Koshland Jr., 1953). However, it was only 37 years later that Michael Sinnot described the molecular basis of these two mechanisms, and highlighted the fact that irrespective of the outcome, GHs act through an acid/base general catalysis scheme (Figs. 1 and 2). Indeed in both inverting and retaining GHs, the reaction is catalyzed by a catalytic dyad, usually composed of two carboxylic acid groups, or sometimes a triad, when a catalytic "helper" involved in the nucleophile stabilization is needed for catalysis to occur.

In the case of inverting GHs, the acid catalyst donates a proton to the interglycosidic oxygen atom of the substrate's glycosidic bond and thus assists its breakdown. Simultaneously, the base catalyst activates a water molecule, which in turn launches a nucleophilic attack on the anomeric carbon (Davies et al., 1995), forming a product exhibiting inverted anomeric conformation (relative to the substrate). In this mechanism, both water and the sugar substrate need to be present at the catalytic center at the same time. Consequently to facilitate this, the distance between the two catalytic residues is usually approximately 10 Å, which is greater than that observed in GHs that operate via a retaining mechanism.

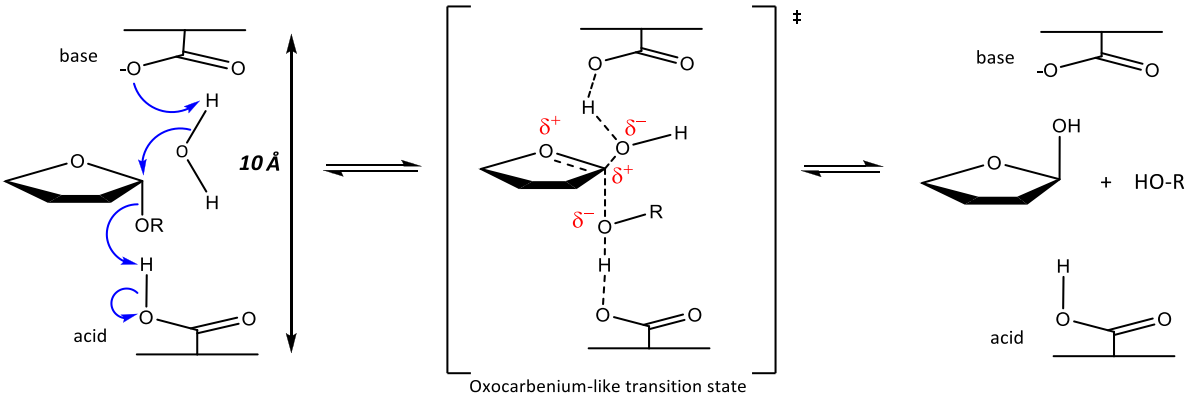


Fig. 1: Inverting GH catalytic mechanism

In the mechanism operated by so-called retaining GHs, the reaction is characterized by two distinct steps that occur successively (Fig. 2). In the the first step, called enzyme glycosylation, the **acid**/base catalyst donates a proton to the interglycosidic oxygen atom of the target glycosidic bond. Simultaneously, the nucleophile catalyst launches a nucleophilic attack on the anomeric carbon causing a first inversion of the anomeric configuration and formation of a covalent glycosyl-enzyme intermediate, accompanied by the departure of the substrate's aglycon moiety (R-OH). In the second step, called deglycosylation, the enzyme is deglycosylated via an attack of the anomeric carbon by an acceptor molecule (R'-OH) that is first activated by the acid/**base** catalyst. The outcome of this second step is a new inversion of the anomeric configuration and the generation of a glycoside product whose anomeric configuration is thus restored (double displacement) relative to the initial configuration of the substrate. In the case of hydrolysis, the nucleophilic attack is carried out by a

water molecule (R' = H), which triggers the breakdown of the glycosyl-enzyme bond. In the case of transglycosylation, the mechanism is identical, but the nucleophilic attack is launched by an activated hydroxyl belonging to an acceptor molecule (R' = acceptor). Unlike inverting GHs, in the retaining mechanism, the nucleophilic acceptor acts on a reaction intermediate and thus the distance between the two catalytic residues shorter, being on average 5.5 Å.

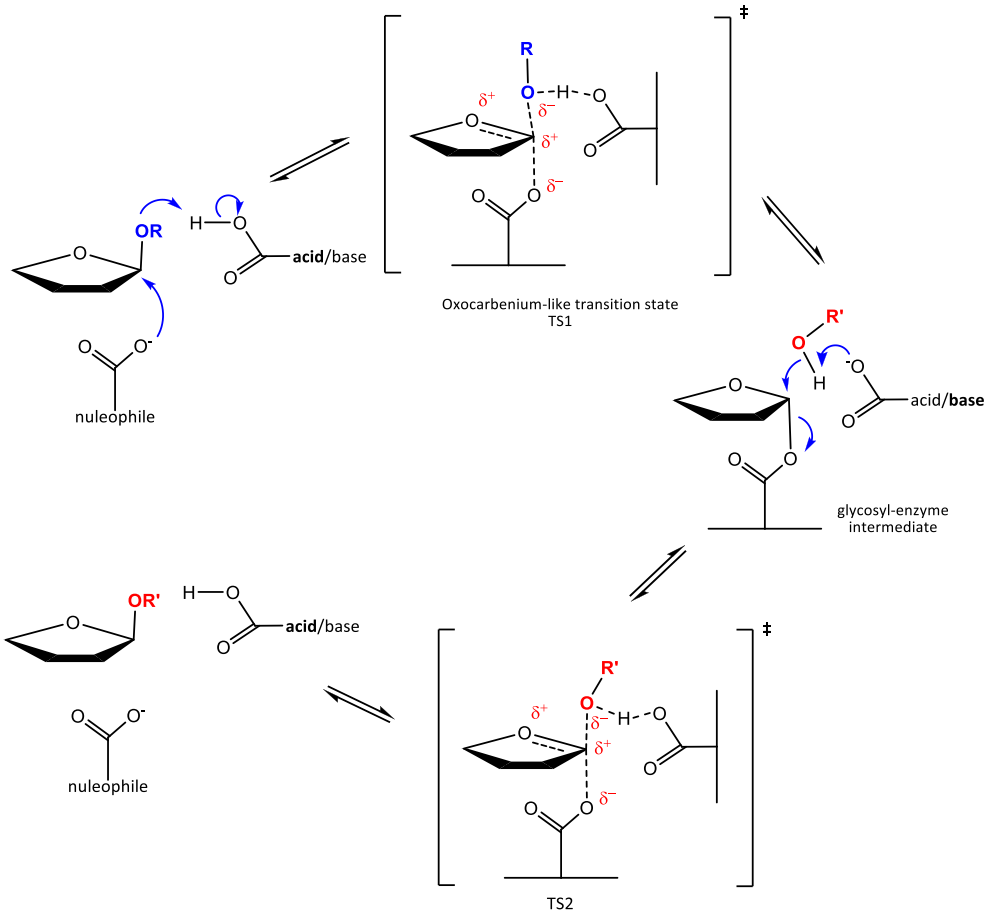


Fig. 2: Retaining GH catalytic mechanism

Beyond these two canonical mechanisms, or equivalent substrate-assisted mechanism, some other unusual catalytic pathways for glycoside cleavage have been described, notably in the case of enzymes belonging to GH4 and GH109 families. In these cases, catalysis involves hydration and elimination steps (Jongkees et al., 2014).

2.2. Modified Glycoside Hydrolases

To expand the glycosynthesis toolbox, a considerable amount of work has been devoted to the engineering of GHs, with the aim of producing enzymes that can efficiently synthesize glycosidic bonds, taking advantage of the outstanding substrate diversity of GHs. This knowledge-driven work, initiated more than 15 years ago by scientific pioneers, such as S. Withers or A. Planas has led to the

development of the well-known glycosynthase concept and derivatives thereof.

2.2.1. Glycosynthases

A glycosynthase is a GH in which the nucleophile catalytic residue has been mutated, generally eliminating the carboxylic acid function. Glycosynthases are crippled enzymes that are unable to catalyze reactions, except when an activated glycosyl donor is fed into the reaction. In this case, glycosynthases can mediate the transfer of the donor glycon moiety onto a sugar acceptor (Fig. 3). Usually, glycosynthases require the use of very activated donor sugars, such as glycosyl fluorides. Moreover, because the donor sugar needs to mimic the covalent glycosyl-enzyme intermediate, its anomeric configuration is inverted relative to the enzyme's normal substrate. When these conditions are met, the crippled glycosynthase is able to mediate transglycosylation, forming a product that cannot be hydrolyzed, since the anomeric position is no longer suitable for attack.

The first *exo*- and *endo*- glycosynthases to be reported were mutant derivatives of a retaining *exo*-β-glucosidase from *Agrobacterium sp*. (Mackenzie et al., 1998) and a retaining β-glucanase from *Bacillus licheniformis* (Malet et al., 1998) respectively. Having demonstrated the power of this technique on these enzymes, the glycosynthase concept has been progressively extended to many other GHs, including inverting GHs. In the latter case, this involved the exploitation of the Hehre resynthesis-hydrolysis mechanism, and was performed on a reducing-end xylose-releasing exo-glucanase (Rex), which nevertheless conserved some hydrolytic activity due to nucleophilic water (Honda et al., 2006).

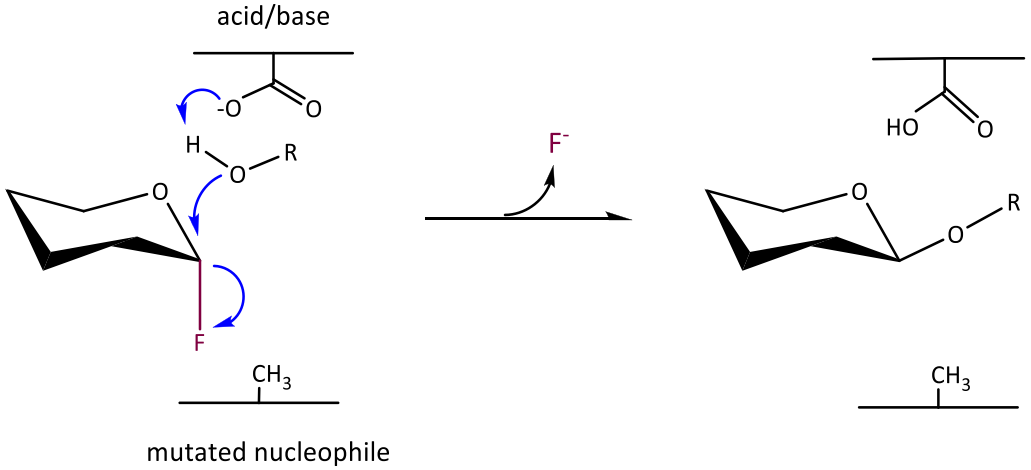


Fig. 3: Glycosynthase mechanism (nucleophile mutant of retaining GH)

Glycosynthases are generally prepared by mutating the nucleophile residue to Alanine, Glycine or Serine. Depending on the exact structure of the enzyme and to what extent the enzyme will form interactions with the departing fluoride, such mutations will lead to more or less active

glycosynthases. Very recently, a substitution of the catalytic nucleophile into aspartate was reported (Aragunde et al., 2014; Andrés et al., 2014) leading to so-called transitional glycosynthases.

Based on these explanations, since the nucleophile residue is mutated it is easy to understand why the transglycosylation product should not be hydrolyzed but one may wonder why the donor cannot be simply hydrolyzed through a nucleophilic attack of a water molecule that still can be activated by the acid/**base** residue. To answer this question, structural data on the aforementioned Rex-derived glycosynthase afforded explanation on the acquisition of glycosynthase activity (Hidaka et al., 2010). The main findings are that the electron lone pair of the water's oxygen, located in the same position as the nucleophilic water in the parental enzyme, displays a lack of orientation leading to a severe decrease in hydrolytic activity. Indeed, for this glycosynthase it is the so-called "water holder" Y198 that was mutated into F, eliminating therefore a hydrogen bond with the water molecule, the second one with the catalyst D263 being conserved. However, since this glycosynthase is derived from an inverting GH, acting in a single step displacement with no covalent glycosyl-enzyme intermediate formation, these structural findings could be extended to other inverting GHs harboring a water holder, but not to retaining GHs.

In the same way that glycosynthases can be made from retaining β-glycosidases, they can also be made from retaining α-glycosidases. Moreover, the concept has been extended by mutating the acid/base residue. In this case, when glycosyl fluorides are used as donors (Fig. 4), the enzyme acts as a so-called *O*-glycoligase (Kim et al., 2010). Unlike glycosynthases derived from β-D-glycosidases, when using an *O*-glycoligase a β-glycosyl-enzyme intermediate is formed. This β-intermediate appears to be reactive enough for coupling with an acceptor sugar to be possible, as opposed to α-glycosyl-enzyme intermediate obtained with acid/base mutant of β-D-glycosidases that are too stable.

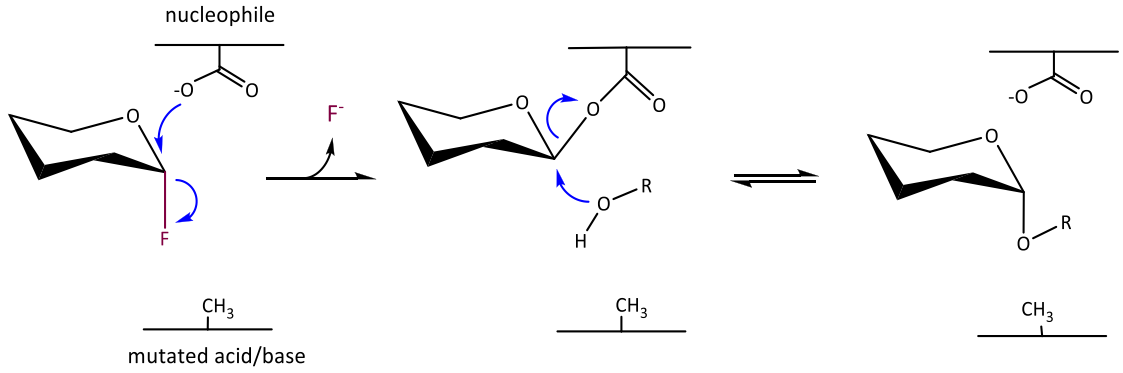


Fig. 4: α-*O*-glycoligase catalytic mechanism

Since the first report of a glycosynthase in 1998, the glycosynthase strategy has been applied to more than 24 GHs from families GH1, 2, 5, 7, 8, 10, 12, 16, 17, 26, 29, 31, 36, 52, 85 and 95 (Cobucci-Ponzano et al., 2012). Most of glycosynthases have been created from retaining GHs, except glycosynthases obtained from Rex (GH8) and those made using α-(1,2)-L-fucosidase (GH95) from *Bifidobacterium bifidum* (Wada et al., 2008), both of which are inverting GHs. Recently, two chitinases (GH18) were submitted to mutation in order to generate glycosynthases. Although these enzymes are hydrolytically-crippled and display the ability to perform transglycosylation (Martinez et al., 2012), they use oxazoline-activated donors rather than glycosyl fluorides. Moreover, these enzymes are mutated at the "stabilizer" amino acid position, since the catalytic mechanism operated by chitinases involves acetamido anchimeric assistance, activated by an amino acid "stabilizer" (Fig. 5).

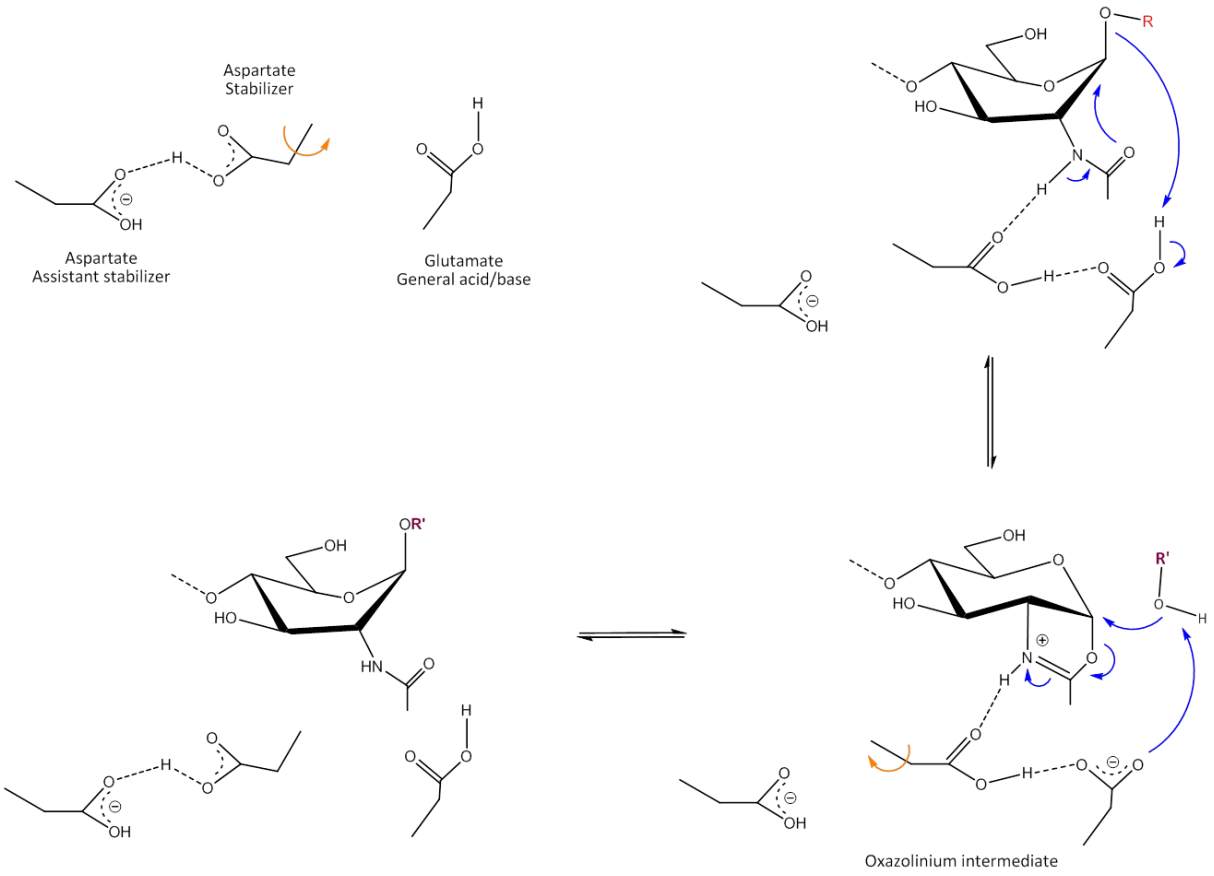


Fig. 5: Proposed GH18 chitinases catalytic mechanism involving a substrate-assisted mechanism of hydrolysis. The orange arrow depicts the rotation of the stabilizing aspartate toward the general acid/base upon substrate binding

Despite the fact that the glycosynthase concept can now be considered as a tried and tested one, it is noteworthy that no furanosynthases have ever been described. This might be due to a general lack of interest, but it appears more likely to be linked to the specific challenges that are raised when working with furanoses. Indeed, it is most probable that it has proven difficult to obtain stable, but

sufficiently-activated furanose donors, such as fluoride furanosides. Indeed, the successful preparation of such compounds has not yet been described in the literature.

2.2.2. Thioglycoligases and thioglycosynthases

So-called thioglycoligases are obtained by mutating the acid/base residue in retaining glycosidases. These enzymes are able to transglycosylate an activated donor sugar moiety onto a thiosugar acceptor. The absence of **acid**/base residue is compensated by a good leaving group (LG) on the donor sugar that does not need protonation of the interglycosidic oxygen (Fig. 6). The presence of a thiol function makes the acceptor nucleophilic enough to attack the anomeric position of the donor involved in the glycosyl-enzyme intermediate, without requiring acid/**base** catalyst-mediated activation. Therefore, the acceptor can react in the second step of the double displacement mechanism (deglycosylation step), whereas water cannot. However, intrinsic nucleophilicity of water may allow residual hydrolysis.

As a remark, up to now, no thioglycoside product was ever obtained with wild-type GH or a glycosynthase using appropriate donor and acceptor, with a good LG and a thiol(ate) function, respectively. This can theoretically be explained by a repulsion of the thiolate ion by the carboxylate function of the acid/base catalytic residue (Stick et al., 2005).

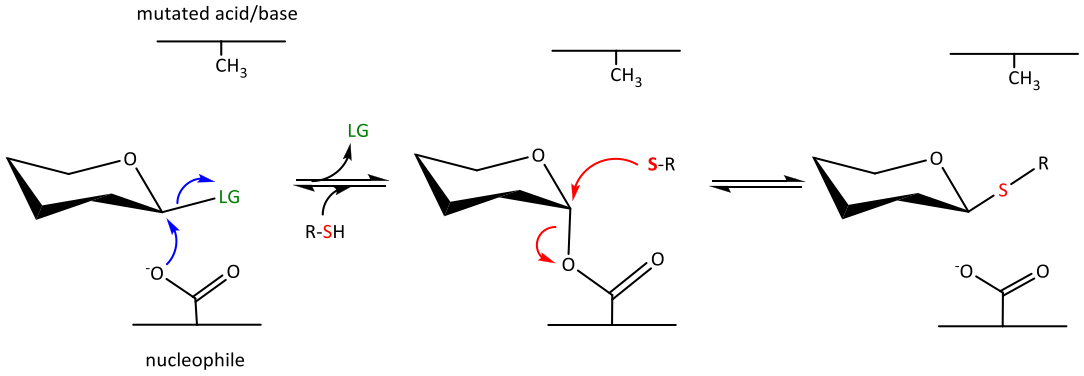


Fig. 6: Thioglycoligase mechanism principle

The efficiency of the first thioglycoligase obtained from *Agrobacterium sp.* was improved through site-saturation mutagenesis of the acid/base residue revealing that E170Q was the best mutation leading to a k_{cat} ~100-fold higher than that of any other mutant (Müllegger et al., 2005).

The thioglycoligase strategy was exclusively employed with sugars under pyranose form but never with furanose sugars till recently. The first study on a thioglycoligase using furanoses as donor sugar was reported in 2011 with an acid/base mutant of the α-L-arabinofuranosidase (GH51) from

Clostridium thermocellum (Almendros et al., 2011). They successfully catalyzed the reaction of ligation of a furanose donor harboring a good leaving group (1-thioimidoyl α-L-arabinofuranoside) onto a broad range of thio-aryl acceptors (thio-benzyl, thio-phenyl, etc.). Unfortunately, the transfer on thio-sugars failed. Thioglycoligation between two furanoses or pentoses still remains a challenge to take up.

Both strategies of thioglycoligases and glycosynthase were merged to generate a new class of enzymes, called thioglycosynthases. The latter are double mutants GH that lack both the nucleophile and acid/base residues (Fig. 7). As a consequence, this type of enzyme catalyzes the glycosidic bond formation from a glycosyl fluoride donor and a thiosugar acceptor (Jahn et al., 2004).

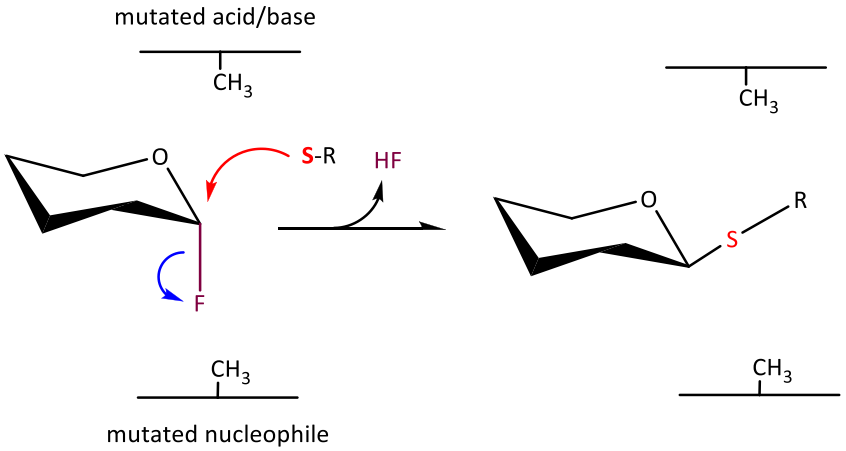


Fig. 7: Thioglycosynthase catalytic mechanism

From a more applicative point of view, beyond the challenge of creating thio-glyco linkages with high regioselectivity, thioglycosides are ideal carbohydrates mimetic for fundamental applications such as protein-ligand interactions studies since they are good inhibitors due to their highly reduced sensitivity to hydrolysis. As therapeutic applications (Witczak et al., 2005), these thio-sugars may serve as inhibitors of certain glycosidases involved in diseases (e.g., type II diabetis). Though several thio-sugars were already obtained via chemical routes (Driguez, 1997), thioglycoligases and thioglycosynthases will offer additional synthesis routes for this class of molecules.

The development of efficient enzymatic tools for glycosynthesis purposes implies to get an accurate understanding of which molecular determinants are driving transglycosylation. In this perspective, the next chapter provides an analytical review of the knowledge acquired up to now on GH-catalyzed transglycosylation characterization, structural determinants involved in hydrolysis/transglycosylation (H/T) partition and what can be done to make this enzymes evolve and develop synthetic activities.

References

Almendros, M., Danalev, D., François-Heude, M., Loyer, P., Legentil, L., Nugier-Chauvin, C., Daniellou, R. & Ferrières, V. (2011) Exploring the synthetic potency of the first furanothioglycoligase through original remote activation. *Org. Biomol. Chem.*, 9 (24), pp.8371–8378.

Andrés, E., Aragunde, H. & Planas, A. (2014) Screening glycosynthase libraries with a fluoride chemosensor assay independently of enzyme specificity: identification of a transitional hydrolase to synthase mutant. *Biochem. J.*, 458 (2), pp.355–363.

Aragunde, H., Castilla, E., Biarnés, X., Faijes, M. & Planas, A. (2014) A transitional hydrolase to glycosynthase mutant by Glu to Asp substitution at the catalytic nucleophile in a retaining glycosidase. *Carbohydr. Res.*, 389, pp.85–92.

Cobucci-Ponzano, B. & Moracci, M. (2012) Glycosynthases as tools for the production of glycan analogs of natural products. *Nat. Prod. Rep.*, 29 (6), pp.697–709.

Davies, G. & Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*, 3, pp.853–859.

Driguez, H. (1997) Thiooligosaccharides in glycobiology. *Glycoscience Synthesis of Substrate Analogs and Mimetics*, 187, pp.85–116.

Henrissat, B. (1991) A classification of glycosyl hydrolases based sequence similarities amino acid. *Biochem. J.*, 280, pp.309–316.

Hidaka, M., Fushinobu, S., Honda, Y., Wakagi, T., Shoun, H. & Kitaoka, M. (2010) Structural explanation for the acquisition of glycosynthase activity. *J. Biochem.*, 147 (2), pp.237 –244.

Honda, Y. & Kitaoka, M. (2006) The first glycosynthase derived from an inverting glycoside hydrolase. *J. Biol. Chem.*, 281 (3), pp.1426 –1431.

Jahn, M. & Withers, S.G. (2004) New approaches to enzymatic oligosaccharide synthesis: glycosynthases and thioglycoligases. *Biocatal. Biotransfor.*, 21, pp.156–166.

Jongkees, S.A.K. & Withers, S.G. (2014) Unusual enzymatic glycoside cleavage mechanisms. *Acc. Chem. Res.*, 47 (1), pp.226–235.

Kim, Y.-W., Zhang, R., Chen, H. & Withers, S.G. (2010) *O*-glycoligases, a new category of glycoside bond-forming mutant glycosidases, catalyse facile syntheses of isoprimeverosides. *Chem. Commun.*, 46 (46), pp.8725–8727.

Koshland Jr., D.E. (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Cambridge Philos. Soc.*, 28 (4), pp.416–436.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, 42 (Database issue), pp.D490–D495.

Mackenzie, L.F., Wang, Q., Warren, R.A.J. & Withers, S.G. (1998) Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.*, 120 (22), pp.5583–5584.

Malet, C. & Planas, A. (1998) From β-glucanase to β-glucansynthase: glycosyl transfer to α-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS lett.*, 440 (1-2), pp.208–212.

Martinez, E.A., Boer, H., Koivula, A., Samain, E., Driguez, H., Armand, S. & Cottaz, S. (2012) Engineering chitinases for the synthesis of chitin oligosaccharides: Catalytic amino acid mutations convert the GH-18 family glycoside hydrolases into transglycosylases. *J. Mol. Catal. B: Enzym.*, 74 (1-2), pp.89–96.

Müllegger, J., Jahn, M., Chen, H.-M., Warren, R.A.J. & Withers, S.G. (2005) Engineering of a thioglycoligase: randomized mutagenesis of the acid–base residue leads to the identification of improved catalysts. *Protein Eng. Des. Sel.*, 18 (1), pp.33 –40.

Stick, R. V. & Stubbs, K.A. (2005) From glycoside hydrolases to thioglycoligases: the synthesis of thioglycosides. *Tetrahedron: Asymmetry*, 16 (2), pp.321–335.

Wada, J., Honda, Y., Nagae, M., Kato, R., Wakatsuki, S., Katayama, T., Taniguchi, H., Kumagai, H., Kitaoka, M. & Yamamoto, K. (2008) 1,2-alpha-l-Fucosynthase: a glycosynthase derived from an inverting alpha-glycosidase with an unusual reaction mechanism. *FEBS lett.*, 582 (27), pp.3739–3743.

Witczak, Z.J. & Culhane, J.M. (2005) Thiosugars: new perspectives regarding availability and potential biochemical and medicinal applications. *Appl. Microbiol. Biotechnol.*, 69 (3), pp.237–244.

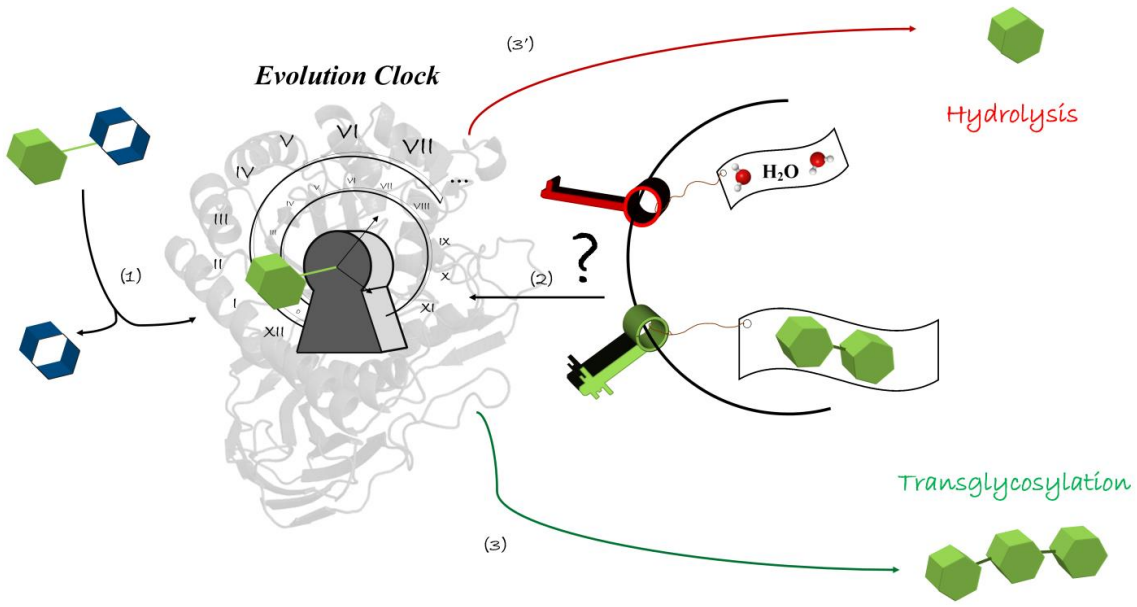
I-3. Glycosynthesis in a Waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases

(Critical Review, invited to be submitted)

Bastien BISSARO*†‡, Pierre MONSAN*†‡§, Régis FAURÉ*†‡ and Michael J. O'DONOHUE*†‡[1]

* Université de Toulouse; INSA,UPS,INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France
† INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France
‡ CNRS, UMR5504, F-31400 Toulouse, France
§ Toulouse White Biotechnology, UMS INRA/INSA 1337, UMS CNRS/INSA 3582, 3 Rue des Satellites 31400 Toulouse, France

[1] To whom correspondence should be addressed (email michael.odonohue@insa-toulouse.fr)



Keywords: Glycoside hydrolase, Transglycosylation, Evolution, Structure/function, Transition state theory

Abbreviations: AS, amylosucrases; CAZyme, carbohydrate-active enzyme; CGTase, cyclodextrin glucanotransferase; EG, *endo*-glucanase; ENGase, *endo*-β-*N*-acetylglucosaminidase; FS, fructansucrase; FT, fructosyltransferase; GH, glycoside hydrolase; GP, glycosyl phosphorylase; GS, glucansucrase; GT, glycosyltransferase; KIE, kinetic isotope effect; LG, leaving group; QM/MM, quantum mechanics/molecular mechanics; SA, sialidase; SUH, sucrose hydrolase; TG, transglycosylase; *tr*S, *trans*-sialidase; TS, transition state; TS1, glycosylation step-associated transition state; TS2, deglycosylation step-associated transition state; TST, transition state theory; VI, vacuolar invertase; XEH, xyloglucan *endo*-hydrolase; XET, xyloglucan *endo*-transglycosylase; T/H, transglycosylation/hydrolysis ratio.

Running title: Glycosynthesis in a Waterworld

Abstract

Carbohydrates are ubiquitous in Nature and play vital roles in many biological systems. Therefore, the synthesis of carbohydrate-based compounds is of considerable interest both for research and commercial purposes. However, for synthetic glycochemists carbohydrates are challenging, because of the large number of sugar subunits and the multiple ways in which these can be linked together. Therefore, to tackle the challenge of glycosynthesis, chemists are increasingly turning their attention towards enzymes, which are exquisitely adapted to the intricacy of these biomolecules.

In Nature, glycosidic linkages are mainly synthesized by Leloir glycosyltransferases, but can result from the action of non-Leloir transglycosylases or phosphorylases. Advantageously for the chemists, non-Leloir transglycosylases belong to the class of glycoside hydrolases, which are both readily available and exhibit a wide-range of substrate specificities. However, non-Leloir transglycosylases constitute notable exceptions, since these enzymes efficiently catalyze the formation of glycosidic bonds, while most glycoside hydrolases favor the mechanistically-related hydrolysis reaction. Unfortunately, because non-Leloir transglycosylases are almost indistinguishable from their hydrolytic counterparts, it is unclear how these enzymes overcome the omnipresence of water, thus avoiding the hydrolytic reaction. In the absence of this knowledge, it is impossible to rationally design non-Leloir transglycosylases using the vast diversity of glycoside hydrolases as the initial protein scaffolds.

This critical review proposes a new rational basis for the engineering of glycoside hydrolases. This proposal is based on the careful analysis of literature data describing non-Leloir transglycosylases and their relationship to glycoside hydrolase counterparts.

1. Introduction

Carbohydrates are ubiquitous in biological systems, being involved in a plethora of life-sustaining or threatening molecular events [1]. Therefore, the *in vitro* synthesis of well-defined complex carbohydrate-based compounds is of considerable importance, both for fundamental research in glycosciences and for the preparation of commercially-valuable products. In this regard, the synthesis of glycosidic bonds by carbohydrate-active enzymes (CAZymes) (i.e. transglycosylation) has been studied for over 60 years [2], being as old as the study of the mechanistically-related hydrolytic reaction. This is because the advantages of enzyme-catalyzed transglycosylation, particularly stereo- and regio-selectivity, have long been recognized by glycochemists, who have increasingly adopted them in order to simplify complex reactions that are usually conducted using more classical organic chemistry methods.

1.1. Enzymes available to the synthetic glycochemist

In Nature, the synthesis of glycosidic bonds is mainly performed by glycosyltransferases (GTs), thus it would be quite logical for these to be widely exploited by glycochemists [3,4]. However, this is not the case, because these enzymes require nucleotide sugars as donor substrates, which are still not readily available despite recent progress [5,6]. Moreover, experience shows that the heterologous production of GTs is often difficult to achieve, thus limiting the availability of these enzymes. Another class of CAZymes that is often used for glycosynthesis is the glycoside hydrolases (GHs), which are more abundant than GTs and cover an extremely wide range of substrate specificities. Nevertheless, although so-called retaining GHs possess an inherent ability to catalyze the formation of glycosidic bonds, this mechanistic outcome is usually subordinate to hydrolysis. Therefore, the use of GHs for glycosynthesis often depends on the ability of the glycochemist to suppress the latter activity, for example by acting on the thermodynamic equilibrium of the reaction (e.g. using co-solvents and reducing water activity) and thus forcing transglycosylation against hydrolysis [2,7]. However, such techniques are not always easy to implement and the results are often disappointing (e.g. poor selectivity and multiple glycosylations). For this reason, scientists have studied the fundamental basis of the hydrolysis/transglycosylation partition in GH-catalyzed reactions and likewise developed strategies to engineer glycosynthetic enzymes. Progress in this field is exemplified by the ‘glycosynthase concept’, first proposed in 1998 [8,9]. This ingenious technique, which has been extensively reviewed elsewhere [10–14], has so far been applied to GHs from a dozen or so different GH families and has benefited from much developmental work. Following the seminal work of Withers et al., a series of review articles dealing with enzyme-catalyzed transglycosylation have either focused on the enzymes [15], on the products [16,17] or on the catalytic mechanisms involved [18–20].

1.2. Transglycosylases – exceptions to the rule

Over the last 15 years, an increasing number of sequences encoding CAZymes has been added to the CAZy database [21–23] which at the time of writing contains more than 196 000 GH modules, assigned to 133 different GH families. Among the characterized GHs present in this database, only a few have been described as transglycosylases (TGs), meaning enzymes that essentially, or often exclusively, catalyze transglycosylation, even in dilute conditions and aqueous media. Intriguingly, these so-called TGs are highly related to hydrolytic GH counterparts, with any single TG being more related to other members of the GH family to which it is assigned, than to other TGs from other families. This fact underlines the tight evolutionary relationship between TGs and GHs and implies that transglycosylation in GHs is favored by subtle molecular adjustments rather than major modifications, such as major structural changes.

A large number of studies have focused on the identification of the molecular determinants that govern acceptor selectivity (i.e. water *vs* sugar moieties) and thus the hydrolysis/transglycosylation (H/T) partition in related GH/TG pairs. Nevertheless, despite some interesting findings, the conclusions of these studies have fell short of expectations [24,25], since they failed to reveal information of a more generic nature pertaining to the way in which the H/T partition is modulated in GHs. This is unfortunate, because the acquisition of such knowledge will allow protein engineers to exploit the vast biodiversity of GHs, conferring efficient glycosynthetic capability to any single GH. In turn, this knowledge gap is preventing a wider deployment of TGs in synthetic glycochemistry, a prospect that will revolutionize this field, providing access to hitherto inaccessible sugar structures.

In this review, we invite the reader to revisit the considerable knowledge that has been acquired in recent years notably that dealing with GH/TG pairs, but also pertaining to glycosynthases and pseudo-TGs obtained using protein engineering techniques. The ultimate aim of this review is to discuss this data in terms of the H/T partition and thus provide a much clearer theoretical framework for future work.

2. Transglycosylation in glycoside hydrolases

2.1. A mechanistic description of hydrolysis and transglycosylation in GHs

In 1953, Daniel E. Koshland provided the mechanistic framework to describe how GHs cleave glycosidic linkages via one of two general mechanisms, involving either retention or inversion of the anomeric configuration (from substrate and product) [26]. Regarding retaining GHs, which represent approximately 60% of characterized GHs, catalysis occurs in two main steps, the first one being ‘glycosylation’ and the second ‘deglycosylation’. Glycosylation begins with the formation of the Michaelis-Menten complex (E.S) and continues up to the formation of the covalent glycosyl-enzyme

intermediate (or equivalent oxazolinium ion intermediate), coupled to the release of a leaving group (Figure 1 and Box 1). Deglycosylation involves an acceptor molecule and gives rise to one of two outcomes, depending on the nature of the acceptor (Figure 1). If water is the acceptor, hydrolysis occurs, whereas the presence of a suitable sugar acceptor will allow transglycosylation to proceed. As mentioned earlier, some retaining GHs are strict TGs, but most are hydrolases, performing hydrolysis and transglycosylation in parallel and at a level defined by the ratio H/T.

2.2. From a simple lock to a locked door

As the mechanistic description above indicates, the actual functioning of GHs is far more complex than that described in 1894 by Emil Fisher's original 'lock and key' model [27]. Indeed, as Koshland pointed out, this early model is limited in several ways, but in particular because it omits the role of water [28]. In the case of GHs, since the mechanism involves both donor and acceptor molecules (which can be water), we would like to extend the lock and key analogy, adding a door handle, whose action is linked to the open/close state of the lock. Looking first at the model, one can describe a system in which the door opening process occurs in two steps: unlocking and then handle movement. The first step is achieved using a key and the second step is performed by simply exerting downward pressure on the handle. The slacker the door mechanism, the easier it is to open the door, even for the weakest of grips, making this type of door locking system accessible to all comers. In GHs, the lock is the donor subsite and the key is the donor molecule. The lock is open when a catalytic intermediate is formed and the door handle is actioned by an acceptor or a water molecule which is followed by product release. A slack door can be likened to a highly efficient GH. Since it is easy to open and is accessible to all-comers. The most frequent door-opener is water, which is omnipresent, being at 55 M. On the contrary, a stiff door, which requires a firm grip both to turn the key and exert pressure on the door handle, can only be opened by a stronger minority. In enzymatic terms, this minority corresponds to acceptor molecules that specifically interact with the enzyme. Overall, in our analogy, the stiffness of the door opening system is determined by how well transition state (TS) interactions are developed during catalysis, with hydrolysis being associated with efficient catalysis and thus highly developed TS interactions.

Box 1 On the meaning of catalytic constants for retaining GHs

Because k_{cat}, K_M and k_{cat}/K_M values are generally determined to compare wild-type and mutant enzymes, it is pertinent to recall some of the key features of these values [36]. Importantly, in most circumstances the Henri-Michaelis-Menten constant K_M (1913) cannot be equated to the affinity constant ($1/K_d$) [37], especially when considering mutated GHs that display highly modified catalytic capabilities. Indeed, when K_M is rewritten as $[k_3.(k_{-1} + k_2)]/[k_1.(k_2 + k_3)]$ it becomes clear that this constant includes terms that refer to both glycosylation and deglycosylation, whereas the catalytic performance constant $k_{cat}/K_M = k_1.k_2/(k_{-1} + k_2)$ only describes the glycosylation step (enzyme-substrate association and glycosidic bond cleavage) and is thus independent of rate-limiting step considerations (Figure 1). Therefore, while the constant k_{cat}/K_M can be considered as a reliable value to evaluate the impact of a mutation on the glycosylation step, the K_M value should be used with caution. Finally, rewriting the catalytic constant, $k_{cat} = k_2.k_3/(k_2 + k_3)$ reveals that when donors bearing a good leaving group (i.e. usually $pK_a^{LG} < 8.0$) are employed, k_{cat} is approximated by k_3, since the deglycosylation step becomes rate-limiting (i.e. $k_3 << k_2$), a situation that is assumed to be true for most GHs. Therefore, assuming that a suitable donor is used, the measurement of the k_{cat} value provides information about the extent to which mutations affect the deglycosylation step, for example by improving acceptor binding, lowering the TS2 energy barrier or improving product diffusion out of the active site.

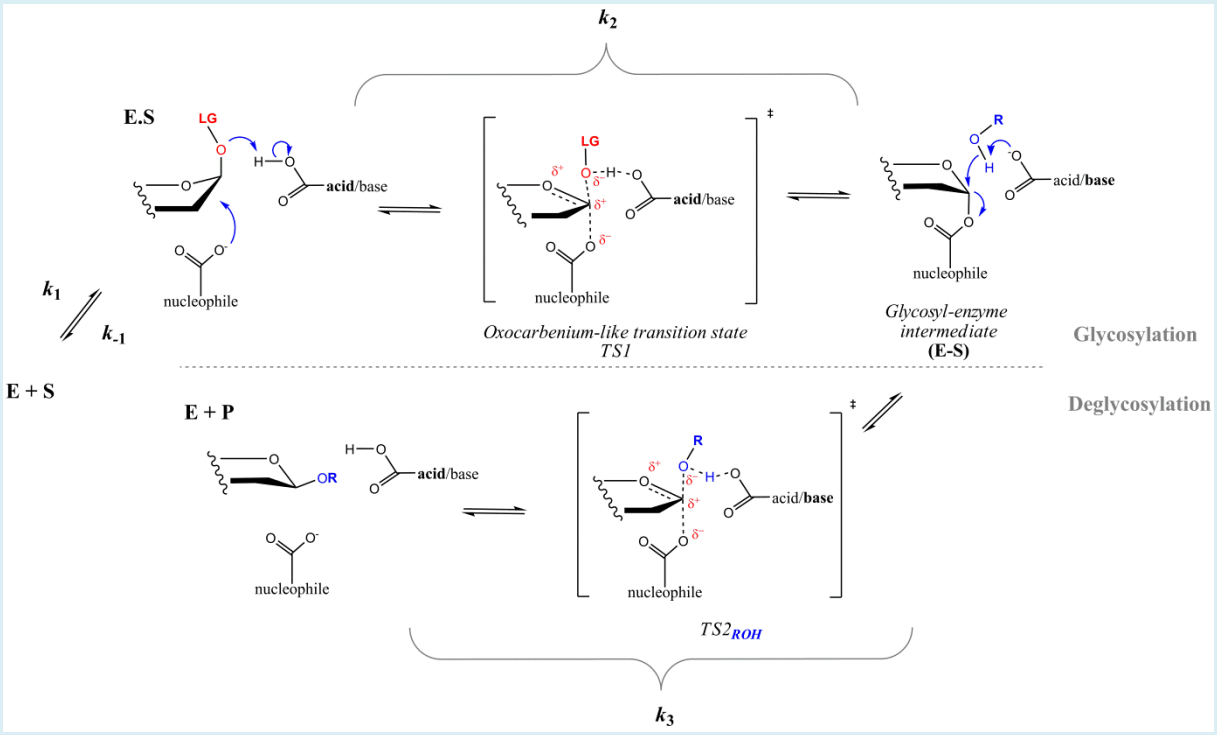


Figure 1 Two-step displacement mechanism of retaining GHs. The donor leaving group (LG) can be either an activated group (e.g. *p*NP) or a sugar (e.g. fructose for glucansucrases). Regarding deglycosylation, the covalent glycosyl-enzyme intermediate can be either attacked by a water molecule (hydrolysis, R = H) or an external acceptor (transglycosylation, R = sugar, alkyl chain, etc.). In the case of secondary hydrolysis the LG becomes R (i.e. donor = transglycosylation product) and R = H (hydrolysis).

2.3. Transition states in glycoside hydrolases and H/T balance

TS: the power of GHs

Glycosidic bonds are extremely stable and display half-lives of several million years. This fact can be illustrated by simply remarking that papyruses from ancient Egypt can still be seen in our museums today. However, in the presence of GHs, the half-life of glucosidic bonds in cellulose for example are reduced to the millisecond range [29].

This incredible catalytic potency of GHs, and enzymes in general, was first rationalized by Linus Pauling in 1946 [30], who proposed that the formation of TS is directly responsible for reaction rate enhancements, which in the case of GHs can be 10^{17}-fold higher than those of uncatalyzed reactions [31].

Retaining GH-catalyzed reactions are characterized by two TS, the first one preceding the formation of the glycosyl-enzyme intermediate (TS1) and the second one (TS2) characterizing disruption of this covalent intermediate and preceding formation of the reaction products (Figures 1 and 2). When compared to the enzyme-free reaction, the enthalpy of activation (ΔH) is significantly lowered, and the degree to which it is decreased correlates with the catalytic efficiency of the enzyme [31] (Box 2).

Box 2 Basics of enzyme thermodynamics

The Arrhenius equation (1889) provides a link between kinetics and thermodynamics, since the rate constant (k) can be expressed as a function of the energy of activation (E_a) and temperature (T), where R is the universal gas constant (Equation 1). Similarly, transition state theory, and in particular the Eyring-Polanyi [62] equation (Equation 2), relates the rate to temperature and thermodynamic parameters, such as the Boltzmann (k_B) and Planck (h) constants, and the Gibbs free energy variable (ΔG_{TS}). ΔG_{TS} includes the activation enthalpy (ΔH) and entropy (ΔS) (ΔG = ΔH - T.ΔS) and denotes the differences of free energy between the ground state (E.S) and TS. When performing site-directed mutagenesis on an enzyme, if the apparent free energy associated with glycosylation is altered for the mutant enzyme relative to the wild-type enzyme, TS can be deduced using these relationships (Equation 3).

$$\boldsymbol{k} = -d[\text{Substrate}]/dt = A.\exp(-E_a/RT) \qquad \text{(Eq. 1)}$$

$$\boldsymbol{k} = (k_B.T/h).\exp(-\Delta G_{TS}/RT) \qquad \text{(Eq. 2)}$$

$$\boldsymbol{\Delta E_a} = \Delta(\Delta G_{TS}) = -RT.\ln([k_{cat}/K_M]_{mut}/[k_{cat}/K_M]_{wt}) \quad \text{(Eq. 3)}$$

Therefore, these equations provide an evaluation of the impact of mutations, or substrate modifications, on the global catalytic efficiency with respect to TS destabilization.

The driving force behind enzyme TS takes the form of local energy expenditure, which is the price of TS stabilization. This energy is in turn derived from the ability of enzymes to form quite intricate interactions with the donor moieties of substrates [32]. Accordingly, tight donor recognition goes hand in hand with efficient electron sharing and the formation of strong, low-barrier hydrogen bonds (< 2.5 Å), two factors that are synonymous with efficient enzyme catalysis.

When discussing enzyme catalysis, it is also relevant to mention enzyme dynamics, because these constitute a key feature of the process [33,34]. Indeed, attempts to investigate catalytic phenomena such as the modulation of H/T using methods such as X-ray crystallography have often failed to provide any useful information, probably because of the absence of dynamics. Nevertheless, the role of dynamics in TS formation is less clear [35], although it is plausible that they contribute to TS properties.

On TS properties

The structures adopted by TS along the reaction pathway adopt a coplanar geometry between C5 (or C4), O5 (or O4), C1 and C2 for pyranoses (or furanoses), which implies the formation of an oxocarbenium ion-like state (sp^2-hybridization) [38]. In this case, the anomeric carbon is subject to electrophilic migration (Figure 1) towards the nucleophile catalyst [39,40]. To favor orbital overlap between the electron lone pair of the endocyclic oxygen and C1 (necessary for cationic character establishment), the sugar undergoes ring distortion, moving away from the lowest energy chair conformation [41], as illustrated by structural [42,43] and computational [44] analyses.

Recently, *in silico* approaches have been employed to demonstrate that maximum charge development and TS coordinate points do not necessarily occur at the same time point [45]. In other work, it has been shown that the presence of a hydrophobic platform within the subsite-1, present in almost all GHs (α or β, retaining or inverting) might play a critical role [46].

Regarding the energetic properties of TS, the contribution of the 2-hydroxyl group is a well-known feature of retaining β-glycosidases (5-10 $kcal.mol^{-1}$, compared to < 2 $kcal.mol^{-1}$ for other hydroxyl groups) [18,47,48]. This is because in β-glycosidases, the 2-hydroxyl group hydrogen bonds to the catalytic nucleophile, thus favoring a greater share of positive charge and directly affecting oxocarbenium cation formation [49], though to different extents depending on the GH family [19]. In the case of retaining α-glycosidases and α-glycosyltransferases [50], this contribution plays a lesser role (5.2 and 1.9 $kcal.mol^{-1}$), probably because of different electronic patterns within the trio constituted by the nucleophile's carboxylic acid function, the endocyclic oxygen and the anomeric carbon of the sugar moiety [18]. In retaining β-glycosidases, the nucleophile carboxyl oxygens establish a *syn* interaction with the 2-hydroxyl group and the anomeric center, whereas in retaining α-glycosidases the equivalent *syn* interaction involves the endocyclic oxygen and the anomeric

carbon center. A direct consequence of this is that a greater share of positive charge is either localized on the anomeric center, or on the endocyclic oxygen, for retaining β- or α-glycosidases respectively. Taking this difference into account, when considering TS electronic patterns it is plausible that this feature could be a key determinant of the principal activity displayed by any given glycosidase. Indeed, it is noteworthy that many 'true' non-Leloir TGs are α-retaining enzymes (e.g. glucansucrases, CGTases), which form a β-linked covalent intermediate, this latter displaying an inherently greater reactivity compared to its α- counterpart [51]. Furthermore, α-retaining GHs are all equipped with *anti*-protonators, which means that unlike *syn*-protonators the interaction of the acid/base catalyst with the lone pair of the endocyclic oxygen is impossible [52,53]. In principle, the absence of this interaction is detrimental for TS stabilization, although some GHs display compensatory interactions (e.g. provided by conserved tyrosines in some β-retaining glycosidases) [46,53]. Therefore, for any given GH, the study of the impact of charge distribution at TS and the anomery of the glycosyl-enzyme intermediate on the selectivity between water and sugar acceptors should be a useful source of information on the enzyme's H/T partition.

From a temporal point of view, TS are highly transient, displaying lifetimes estimated to be within a single bond vibration timescale (i.e. approximately 10 fs, or 10^{-15} s) [54], far lower than the global k_{cat}, which occurs on a millisecond timescale in most GHs. Regarding water molecule, its diffusion takes approximately 1 ps (10^{-12} s) and does not constitute a rate-limiting step, unlike bond breaking and formation, which are much more critical (see below).

Differences between TS1 and TS2

So far, from a practical point of view, the kinetic isotope effect (KIE), also called 'isotope fractionation' [55], has proved to be the only experimental approach that can provide details about TS formation and properties (i.e. geometry and electronic environment) [54,56]. Therefore, it is this technique that was used to reveal that in the reaction catalyzed by *Agrobacterium sp.* β-glucosidase, the oxocarbenium ion character was stronger for the deglycosylation TS than for that of the glycosylation step [57], despite the fact that the TS1 and TS2 in retaining GHs are usually considered to share very similar features. Nevertheless, it is clear that the study of TS, in particular the activation barrier of TS2, is hampered by the lack of experimental approaches that can provide sound data.

Recently, Quantum Mechanics/Molecular Mechanics (QM/MM) approaches have been developed to provide insight into the properties of TS and the extent of bond breaking at each individual step [58,59]. Accordingly, based on findings from QM/MM it has been postulated that TS2 is more dissociative, meaning that the (C1-nucleophile) bond is almost broken before the nascent C1-OR bond (with OR from acceptor HOR, with R= H for water) is formed [60]. This provides interesting insight into the enzyme-catalyzed chemistry of the second reaction step and is consistent with the

fact that water-mediated deglycosylation is rate-limiting. Moreover, QM/MM has revealed that conserved, non-catalytic active site residues, which are involved in hydrogen bonding with the sugar moiety, contribute to TS stabilization to different extents, this being dependent on the exact position of the hydroxyl moiety and the reaction step under consideration [61]. This is consistent with previous experimental findings that reported on the different contributions of the sugar hydroxyl groups [48]. Despite these encouraging results, QM and MM are still in their infancy and thus findings need to be more extensively corroborated by experimental data.

Unfortunately, in the case of retaining GHs, the study of TS is always limited to those developed during hydrolytic reactions, despite the fact that other reagents, such as hydroxylated molecules, can act as acceptors for the deglycosylation step (i.e. transglycosylation). Therefore, in the quest to elucidate the determinants of H/T modulation, it is rather evident that water and carbohydrate-mediated deglycosylation involve different behaviours. Although diffusion issues should be considered as important, thermodynamics are at heart of the enzyme-catalyzed chemical reaction and are probably much more critical, as underlined by *in silico* approaches. Therefore, the key questions regarding the H/T partition appear to concern the properties of the deglycosylation transition state ($TS2_{ROH}$) and the impact of the nature of the reacting acceptor substrate (ROH) on it.

3. Naturally-occurring transglycosylases: elucidating Nature's design strategy

In the following section, naturally-occurring TGs are defined as retaining GHs that display a dominant or exclusive ability to transfer glycosyl units onto acceptor sugars (e.g. xyloglucan *endo*-transglycosylases or XET, sucrase-type enzymes, cyclodextrin glucanotransferases or CGTases and *trans*-sialidases or *tr*S). For practical reasons, in the specific case of TGs, when referring to the partition between hydrolysis and transglycosylation, the ratio T/H is used in preference to the more usual H/T ratio. Moreover, herein, we only discuss enzymes for which there is a sufficient amount of knowledge concerning structure-function relationships.

In guise of a general introduction to this section, the reader is referred to Table 1, which underlines the fact that sugar-transferring enzymes are usually catalytically-less efficient (e.g. k_{cat}/K_M values) than hydrolytic counterpart enzymes (85- to 1165-fold lower for GH1 β-glycosidase and GH13 sucrose-acting enzymes, respectively). This catalytic sluggishness is likely to be correlated with more energy-demanding TS (for both glycosylation and deglycosylation steps), which lower overall catalytic turnover. In this respect, it is also interesting to recall that in a previous study that set out to correlate enzyme and substrate flexibility with catalytic performance, it was proposed that the enzymes we observe today are the result of evolutionary processes that have transformed intrinsically slow, broad specificity prototypes into more efficient catalysts [63]. Of course, this is a

rather simplistic view of enzyme evolution and other data suggest that enzymes might have evolved in both directions [64,65], and indeed, some GHs mainly from plants display both hydrolysis and transglycosylation activities and thus present intermediate cases (i.e. mixed activity) [66–71].

3.1. Xyloglucan *endo*-transglycosylases

From the perspective of understanding the determinants of the T/H partition, XETs are extremely interesting enzymes. These are mainly grouped in family GH16 (members of GH-B clan), along with their hydrolytic counterparts, xyloglucan *endo*-hydrolases (XEH), as well as other hydrolytic enzymes that display a wide variety of substrate specificities. The molecular phylogeny of xyloglucan *endo*-transglycosylase/hydrolase (XTH) genes, enzyme characteristics and *in vivo* functional differences provide criteria that can be used to define three major groups in this class of enzymes, which are involved in plant cell wall metabolism.

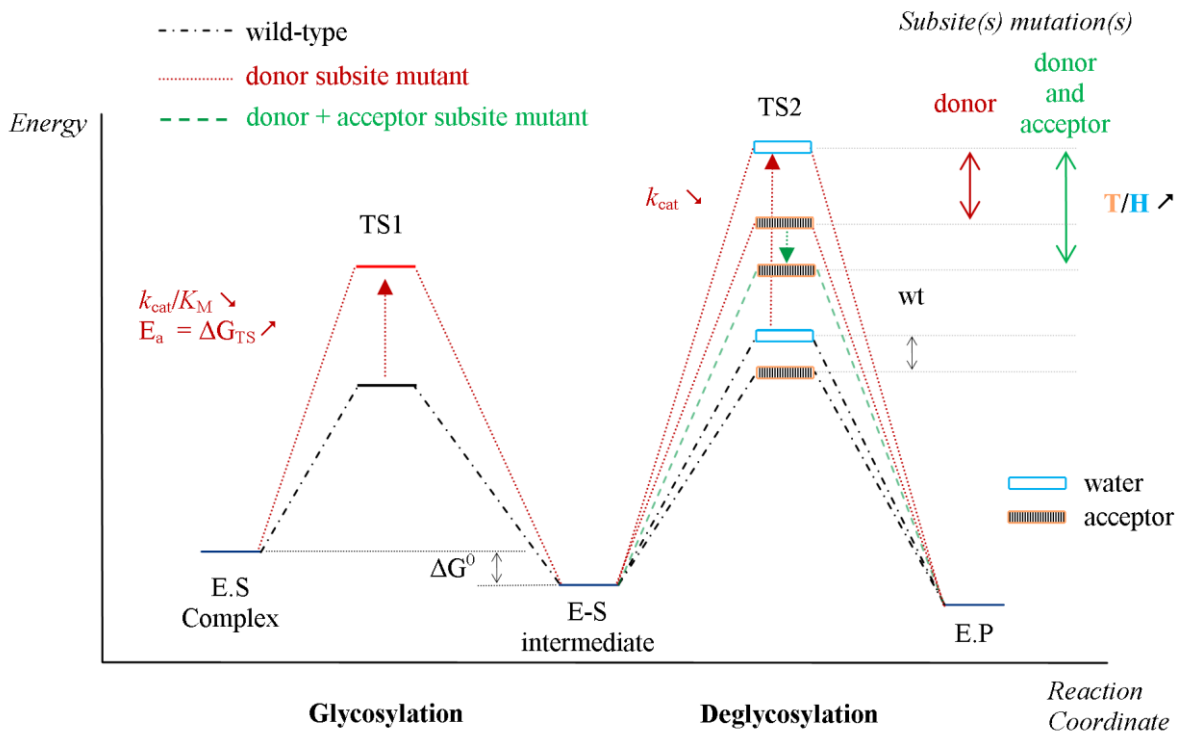


Figure 2 Energetic diagram of the two-step displacement mechanism of retaining GHs (black dash-dot) and alternative energetic pathways for evolved transglycosylases (red dot or green dash for donor and donor+acceptor subsites mutants, respectively). A water-mediated (blue open rectangle) TS2 destabilization coupled to an acceptor-mediated (orange dashed rectangle) stabilization would lead to a T/H partition increase. Alternatively, the E-S intermediate energy level could also be re-up favoring as well an acceptor-mediated deglycosylation. A combination of both phenomena is not excluded to explain transglycosylases behavior.

Group I and II members exclusively exhibit XET activity, which is also the predominant feature of group III-B, unlike members of group III-A (XEH) that are mainly hydrolytic [72–75]XETs catalyze the non-hydrolytic cleavage and religation of xyloglucan molecules through a ping-pong bi-bi mechanism with competitive inhibition, since its competing substrates can act as both the donor and the acceptor [76,77].

As explained earlier, the canonical double-displacement mechanism of glycosyl transfer involves the formation of a covalent glycosyl-enzyme intermediate, which rapidly accumulates (< 2 min) and displays a relatively long lifetime (hydrolytic half-time of this intermediate for *Ptt*XET16-34, from *Populus tremula x tremuloides*, is approximately 3 h with $k_{hydr.} = 1.10^{-4}$ s^{-1}), associated with an estimated free-energy change (ΔG^0, Figure 2) of formation of approximately 1.5-2.0 $kcal.mol^{-1}$ (Piens et al., 2008; Steele & Fry, 1999; Sulova, Takacova, Steele‹, Fry, & Farkas, 1998). For deglycosylation to occur, the presence of suitable sugar acceptors (e.g. xylogluco-oligosaccharides) is required. When this criterion is fulfilled, the catalytically-competent covalent xylogluco-oligosaccharyl-XET adduct can be fully deglycosylated in 30 min. In this respect, it is also noteworthy that when *Ptt*XET16-34, a XET from hybrid poplar, was supplied with activated β-D-xyloglucan-oligosaccharidic donors (e.g. LG = 2-chloro-4-nitrophenyl or fluoride), no activity (neither transglycosylation nor hydrolysis) was observed [81,82]. This underlines the fact that the donor aglycon moiety must be a sugar, presumably because a specific molecular structure is required to drive TS1. This requirement can be avoided through the creation of a PttXET16-34-based glycosynthase since the reaction only proceeds through the 'pseudo' second step ($TS2_{ROH}$) of their 'original' retaining-mechanism.

Consistently, it is remarkable that donor substrate binding is dominated by a higher affinity of the xyloglucan moiety within the positive subsites, an interaction that is driven by the presence of aromatic residues. This increased affinity for the positive subsites is thought to be necessary (though not sufficient *per se*) for transglycosylation [77,81].

Despite a lack of sequence identity within family GH16, all of its members display a typical β-jellyroll fold that is composed of two large curved β-sheets, stacked in a sandwich-like manner. However, in the case of XTHs specific structural features reflect the specialization of these enzymes toward their highly branched substrates [81]. Notably, according to Brumer et al., starting from an ancestral (hydrolytic) licheninase active on linear 1,3-1,4-β-glucans [83], the deletion of a loop procured the ability to bind highly branched substrates, such as xyloglucan, a characteristic that is shared by both *endo*-glucanases (EG) and XTHs that display hydrolytic and/or transglycosylation activities. Illustrative of this are *Pt*EG16 from *Populus trichocarpa*, which is able to hydrolyze the xylogluco-oligosaccharide XXXGXXXG, and its counterpart *Ptt*XET16-34, which performs transglycosylation using the same substrate [76,83]. Moreover, the extension of the C-terminal

domain differentiates the XTHs from EGs. This XTH feature provides xyloglucan specificity (i.e. targets branched substrates) to this group of GH16 enzymes [77,83]. Finally, regarding XETs and XEHs, X-ray structure data has revealed that in some cases these differ in two loops located between β-strands β6 and β7, and between β8 and β9, in the vicinity of the active site [74]. The importance of this last observation has been demonstrated through the creation of a β8/β9 loop deletion in the *Tropaeloum majus* XEH (*Tm*NXG1-ΔYNIIG mutant), a loop that forms part of subsite +1 and interacts with the D-glucosyl residue. This mutation procured an increased T/H ratio in the initial phase of the reaction, with a 2-fold increase in transglycosylation rate being coupled to a 5.7-fold decrease in hydrolysis rate.

Moreover, structural and molecular dynamics work performed on *Ptt*XET16-34 and *Tm*NXG1 has revealed a correlation between the nature of the principal activity and subsite binding interactions, which are combined with subtle differences in dynamic behavior [84]. As a matter of fact, in XETs, the number of H-bonds formed between the enzyme and the acceptor moiety is greater than in XEHs, whereas in XEHs the number of H-bonds formed with the donor moiety is higher. Moreover, a key to the transglycosylation ability of XETs appears to be more flexibility in subsite -1, which is detrimental for activity, except when a sugar is present in subsite +1.

3.2. Sucrase-type enzymes

Sucrases are *exo*-enzymes that include glucansucrases (GS) and fructansucrases (FS). Using sucrose as a substrate, these enzymes are able to synthesize homopolysaccharides composed of D-glucosyl or D-fructosyl subunits respectively, with different linkage specificities [93,94]. GS are classified in both GH13 and GH70 family, with GH13 GS being designated amylosucrases (AS). Transglucosylating AS and GS have been extensively studied both in our group [93,95] and in L. Dijkhuizen's group [94,96]. Although AS, GS and FS act on the same substrate, they actually exhibit different protein folds, with AS and GS being characterized by a $(\beta/\alpha)_8$-barrel architecture and belonging to clan GH-H (α-amylase superfamily), which is divided into 40 subfamilies [97], and FS belonging to clan GH-J (5-bladed-β-propeller). Nevertheless, all three enzyme groups operate through a retaining mechanism.

Table 1 Comparison of catalytic constants between glycosynthetic and hydrolytic natural GHs

GH family	Enzyme	Substrate	k_{cat}[a] (s^{-1})	K_M[a] (mM)	k_{cat}/K_M[a] ($s^{-1}.mM^{-1}$)	Reference
1	Rice *Os*BGlu31 (*exo*)	Ferulic acid (acceptor)[b]	1.21	0.05	25.42	[66]
		*p*NP-β-D-Glc*p* (donor)[b,c]	1.21	9.33	0.13	
	Agrobacterium β-glucosidase (*exo*)	*p*NP-β-D-Glc*p*	169	0.078	2170	[57]
13	*Bc* strain 251 CGTase (*endo*)	β-cyclization	329	-	-	[85]
		hydrolysis	3.9	-	-	
	Barley α-amylase (*endo*)	Blue starch	248	0.52 ($mg.mL^{-1}$)	477 ($s^{-1}.mL.mg^{-1}$)	[86]
		CNP-β-D-maltoheptaoside[c]	122	1.1	111	
	*Np a*mylosucrase (*exo*)	Sucrose (< 20 mM)[d]	0.55	1.9	3.45	[87]
		Sucrose (> 20 mM)	1.28	50.2	0.0255	
	Xag Sucrose hydrolase (*exo*)	Sucrose	66.5	2.24	29.7	[88]
16	*Ptt*XET16-34 (*endo*)	XGO_{Glc8} (transglycosylation)[e]	0.08	0.4	0.2	[74]
	*Tm*NXG1 (XEH)	XGO_{Glc8} (hydrolysis)[e]	0.071	0.08	0.85	
		XGO_{Glc8} (transglycosylation)[e]	0.015	0.5	0.028	
32	Wheat FT (1-SST) (*exo*)	Sucrose (1-kestose production)	0.78	551	-	[89]
	Wheat VI (*exo*)	Sucrose (hydrolysis)	608	15	-	
33	*Tc*trS (*exo*)	Sialyllactose (tranglycosylation)	12.6	1.2	10.5	[90]
	*Tc*trS (*exo*)	Sialyllactose (hydrolyse)	0.18	0.29	0.62	[91]
	*Tr*SA (*exo*)		151.4	0.27	554.7	

[a] Determined in the optimal operating conditions for each enzyme. Specific activity is provided when the k_{cat} value is unavailable.

[b] Kinetic parameters were determined either for the acceptor (with 30 mM donor) or for the donor (with 0.25 mM acceptor)

[c] *p*NP, 4-nitrophenyl; and CNP, 2-chloro-4-nitrophenyl.

[d] For low sucrose concentration (< 20 mM), hydrolysis is dominant.

[e] Xyloglucan-oligosaccharides mixture composed of XXXG, XLXG, XXLG, and XLLG moieties (using the nomenclature developed by Fry et al. [92]) and based on (D-Glc*p*)$_8$ backbone.

Until recently, the structure of GS remained elusive [98–100], thus hampering progress in the understanding of structure-function relationships in these enzymes [101]. Nevertheless, using sequence-based approaches, it was possible to identify transition state stabilizers (histidines) that are present in both GH13 and GH70, being conserved in α-amylases, CGTases and GS. These residues are essential for overall catalysis (hydrolysis and transglycosylation) and their mutation is often highly detrimental for activity (< 0.5% residual activity) [101].

Concerning GH70 GS, the analysis of the impact of mutations of key catalytic residues and others located in the acceptor subsites has led to the conclusion that such mutations can be grouped into one of three categories: those affecting *(i)* D-glucosidic linkage specificity, *(ii)* glucan solubility and *(iii)* overall enzyme activity [94]. Structural data analysis revealed that subsite +1 residues form H-bonds with the D-fructosyl moiety, as do residues in subsite +2 with the D-glucosyl moiety, these latter playing an important role in determining the linkage ratio [101].

Results from the study of a reuteransucrase (GH70) from *Lactobacillus reuteri* suggest that steric hindrances play a major role in chain elongation, since the deletion of a variable N-terminal domain procured an increase in transglycosylation (3 to 4-fold) at the expense of hydrolysis [102]. Similarly, the creation of a single point mutation (N1179E) within the same subgroup of enzymes led to a T/H ratio increase [103]. In another study, it was reported that a GH70 4,6-α-glucanotransferase is able to perform a disproportionation reaction on α-(1,4)-linked malto-oligosaccharides, but is unable to use sucrose as a substrate, despite the high energy (6.6 kcal.mol^{-1}) associated with its glycosidic linkage [104]. In the light of this observation it was proposed that this enzyme represents an evolutionary intermediate between GH13 and GH70 [105].

Compared to GH70 GS, the data available for GH13 AS is more abundant. These enzymes all display a similar 5-domain structure with a deep pocket at the bottom of which sucrose binds to subsites -1 and +1 [106,107]. Three arginines (R226, R415 and R446), located in acceptor subsites +2/+3, +4 and +1 respectively, are particularly important in the transglucosylation reaction, since these play a crucial role in the docking and positioning of acceptors [108,109]. In a study of the AS from *Neisseria polysaccharea* (*Np*AS, subfamily 4 of GH13) the acceptor subsites were submitted to mutagenesis with the aim of improving transglucosylation using unnatural acceptors. Although, quite impressive increases in transglucosylation were achieved (395-fold increase), which were accompanied by decreased apparent K_M values, no evidence of significant structural changes that would alter sucrose binding was detected [110]. Therefore, it was concluded that modified loop flexibility and enzyme dynamics are likely to be the determinants of altered substrate recognition and responsible thus for the establishment of a catalytically-productive state. Overall, this study revealed a certain plasticity of subsite +1, because it was possible to isolate mutants that could

glucosylate a series of different acceptors, and suggested that the improvement of transglucosylation using unnatural acceptors was facilitated by improved interactions in the acceptor subsites. In another study, recognition of D-glucosyl moieties in subsite -1 was investigated. This revealed that despite the fact that AS exhibits slow rates, the D-glucosyl is specifically recognized by a complex network of interactions [111]. To provide further insight into the transglycosylating character of *Np*AS, it is worthwhile to compare this enzyme with a hydrolytic counterpart, such as the sucrose hydrolase from *Xanthomonas axonopodis* pv. *Glycines* (*Xag*SUH). Although this GH13 member shares 57% sequence identity with *Np*AS and is structurally similar (identical 5-domain structure with rmsd value of 1.78 Å) [88], *Xag*SUH catalyzes sucrose hydrolysis, and is incapable of catalyzing transglucosylation [112]. One main difference between *Xag*SUH and *Np*AS has been revealed by acquiring structural snapshots along the catalytic coordinate. This revealed that upon sucrose binding in *Xag*SUH a pocket-shaped active site is formed through rigid-body movements of the B and B' domains towards the active site. Moreover, it is noteworthy that the majority of active site residues are conserved between the two enzymes, except for three arginines (R226, R415 and R446) that are substituted by other residues (glycine or leucine) in *Xag*SUH. Significantly, as mentioned earlier, these arginines are essential in *Np*AS for transglucosylation, although the introduction of homologous arginines in *Xag*SUH by mutagenesis failed to confer transglucosylation properties to the enzyme, an observation that is consistent with the fact that improvements in transglycosylation first require diminution of hydrolysis [88], especially given the fact that the value of k_{cat} on sucrose is 120-fold higher than that of *Np*AS (Table 1). Surprisingly, this fact was not evoked by the authors, who suggested that hydrolysis in *Xag*SUH might be caused by a collateral effect of D-fructose release, which would disorder the B-domain and thus expose the enzyme-bound D-glucosyl moiety to bulk solvent and thus hydrolysis.

In summary, the identification of the key factors that determine the T/H partition in GS has proved to be quite difficult, although it appears evident that interactions in acceptor subsites play an important role. In this respect, and taking into account the fact that the active site in these enzymes is often buried, it has been suggested that the presence of an acceptor group in sucrases during catalysis protects the covalent intermediate from water-mediated attack [113]. This might be so if one assumes that sucrose can bind in the subsites -1 and +1 in the presence of the acceptor, which would then presumably occupy other acceptor subsites (i.e. from subsite +2). However, this hypothesis assumes that upon formation of the covalent intermediate the acceptor is somehow displaced towards subsite +1 and that there is considerable flexibility within the active site, allowing for example the D-fructose leaving group to depart unhindered. Unfortunately, at least in the case of AS, this is not generally the case, even if the active site topology of *Deinococcus radiodurans Dr*AS is a

particularly open one [113]. Moreover, even if the active site of all AS are highly flexible and accessible, the hypothesis does not really explain how water-mediated deglycosylation is avoided, especially in an enzyme such as *Dr*AS, which despite its open active site topology, still mainly performs transglucosylation [113,114]. Therefore, alternative hypotheses are required, not to explain how water is prevented from entering active sites, but rather to explain how the presence of water is rendered irrelevant with respect to deglycosylation.

Regarding fructose-specific sucrases (FS), which can act on sucrose and/or fructans, these are gathered within families GH68 and GH32 (clan GH-J). The FS in GH68 (i.e. levansucrases, inulosucrases) usually display a dominant hydrolytic activity, accounting for 70-80% of substrate (levan, inulin) conversion. A previous study performed on the single domain levansucrase, SacB from *Bacillus subtilis*, revealed that the addition of transitional and complete C-terminal domains from other FS led to reductions in hydrolysis (down to 10% of substrate conversion), accompanied by a 5-fold increase in transfructosylation. Upon analysis of the chimeric enzymes, the authors remarked that the k_{cat} value associated with hydrolysis was unaltered and thus attributed the increase in the T/H ratio to more favorable acceptor subsite interactions provided by a structural adjustment in the catalytic site mediated by the addition of additional domains [115].

The GH32 family comprises fructan-acting enzymes (β-D-fructofuranosidases and inulinases), but also sucrose-acting enzymes, and contains an additional β-sandwich domain relative to GH68 FS. The GH32 sucrases or invertases (as they are often known) are able to transfer the D-fructosyl moiety of sucrose either onto water (hydrolysis), leading to the production of fructose (i.e. inverted sugar), or onto a sucrose acceptor (transglycosylation), thus catalyzing the synthesis of fructan. In the latter case, the enzymes are designated as fructosyltransferases (FTs). Within the plant kingdom, sucrose can be degraded by vacuolar (VIs) or cell wall invertases, and from a phylogenetic standpoint FTs and VIs belong to the same GH32 subgroup, sharing high sequence identity (ca. 65%) and structural homology. Using phylogenetic tree analysis, it has been proposed that FTs have evolved from ancestral VIs [65]. Among the different VIs, it is noteworthy that three amino acid sequence motifs are highly conserved **(I): the sucrose-binding box motif "**WMNDPNG**"**, which contains the catalytic nucleophile **D**, **(II):** the **"EC"** motif, which includes the catalytic acid/base **E** and **(III):** the **"RDP"** motif, with **D** being identified as a TS-stabilizing residue [116,117]. The first N in the sucrose-binding box is involved in a hydrogen bond network, forming links with the nucleophile **D** and W. Importantly, within the sucrose-binding box, W is always replaced by a Y in FTs from the same subgroup, and similarly, the first N is very often substituted in FTs by S. Engineering of these alternative residues into VIs demonstrated that the disruption of the hydrogen bond network involving the nucleophile **D** (i.e. W23Y and N25S) enhanced transglycosylation up to 17-fold when compared to wild type VIs

[89]. Similarly, other studies performed on VIs from yeast [118] or onion [65] led to similar conclusions, although the increase in transglycosylation is more modest. Furthermore, a shift of optimum pH from 3.8-4.8 to 4.8-5.7 was observed for a yeast VI mutant (W19Y-N21S), consistent with an alteration of the ionization state of the catalytic residues [118]. Interestingly, the reverse experiment involving the substitution of Y by W in two different FTs failed to procure a more hydrolytic VI-like enzyme [119,120], which suggests that it is much easier to disrupt rather than create a hydrogen bond network !

From a kinetic point of view, compared to FTs (Table 1) VIs are more efficient catalysts. FTs do not display a saturation profile (i.e. K_M of hundreds of mM relative to 2-20 mM range for VIs), but are nonetheless very good at transfructosylation (70-80% substrate conversion) compared to VIs (2-5% of substrate conversion). When considering mutated VIs, these can be seen as intermediate cases, since for most of the available examples K_M values were increased from 4 to 34-fold [65,89,118], resulting in severely reduced k_{cat}/K_M values, an alteration that is indicative of higher TS1 energy levels.

Acceptor substrate selectivity among GH32 has also been investigated using a mutagenesis approach to modify residues located in +1 or +2 subsites. However, this type of mutation has so far not conferred significant transglycosylation ability to invertases, although in at least one case both regioselectivity (β-(2,6)/β-(2,1)) and catalytic efficiency were significantly altered [118,121]. On the other hand, the mutagenesis of putative acceptor subsite residues in a FT proved to be quite detrimental for transglycosylation [120]. Therefore, based on available data on FT/VIs it is possible to conclude that the modification of acceptor subsite determinants can be used to improve acceptor recognition and positioning for transglycosylation, but this is insufficient to destabilize water-mediated deglycosylation (i.e. $TS2_{water}$) in invertases. To achieve this, it is much better to target the proton network in the donor subsite.

3.3. Cyclodextrin glucanotransferases and α-amylases

Involved in starch depolymerization, CGTases and their hydrolytic counterparts, α-amylases, belong to GH13 and thus to clan GH-H. These enzymes share a common structural architecture, which is defined by three domains, A, B and C, although CGTases possess two extra domains D and E. It has been proposed that CGTases have evolved from α-amylases, since the latter display greater sequence diversity and are more widespread through the different taxonomic groups [64]. Regarding the natural function of CGTases, it is likely that by providing cyclodextrins of defined size (i.e. α, β or γ), CGTases procure 'tailored' substrates for α-amylases and thus accelerate starch saccharification.

Crystallographic analysis of CGTases has revealed an extensive active site structure, extending from

at least a subsite -7 to a subsite +3. Due to their architecture, these *endo*-enzymes are able to catalyze intra-molecular transglycosylation (β-cyclization) through the transfer of the covalently bound sugar unit onto the 4-hydroxyl group of the non-reducing end of the same donor molecule [122]. Compared to GS (*exo*-enzymes with only one donor subsite), donor subsite interactions in CGTases are much more developed and, taking into account the high transfer rates that characterize these enzymes (10^2-10^3 IU.mg^{-1}, Table 1), it is probable that the transition state energy barrier is lower than that of GS.

In addition to the synthesis of cyclodextrins, CGTases have also been shown to be capable of hydrolysis or to perform the transfer of the bound glycosyl intermediate onto another α-glucan chain (i.e. disproportionation) [123]. In order to prevent hydrolysis, it appears that CGTases have acquired acceptor subsites that favor sugar recognition. This is illustrated by mutagenesis work that was performed on the acceptor subsites (+2 and +3) of the *Bacillus circulans* 251 CGTase (*Bc*CGTase). The substitution of F183 and F259 in *Bc*CGTase by N or S resulted in a 10 to 300-fold decrease in transglycosylation activity (β-cyclization) and a 3 to 20-fold increase in hydrolysis [124]. Similarly, the simultaneous mutation of equivalent residues (F184Q and F260W) in the CGTase from *Thermoanerobacterium thermosulfurigenes* strain EM1 (*Tabium* CGTase), combined with a third mutation (A231V), converted this enzyme into an α-amylase-like hydrolytic enzyme [125]. Impressively, this mutant no longer displayed detectable CGTase activity, with the T/H ratio being 0.0012 (compared to 5 for the parental CGTase). Consistent with these results, another study focusing on aglycon subsites in liquefying (hydrolytic) and maltogenic (transglycosylating) α-amylases revealed that increased hydrophobicity in subsites +2/+3 of the α-amylase from *Bacillus licheniformis* (BLA) increased the T/H ratio, reducing the hydrolysis rate (associated with a higher K_M value) on starch by one third [126]. Likewise, the sequence comparison of hydrolytic and maltogenic α-amylases revealed the presence in subsite +1 of a conserved histidine or glutamate residue, respectively [127]. The introduction of the substitution H235E in BLA created a transglycosylation activity, which is undetectable in the wild-type enzyme, but did not drastically affect the efficiency of hydrolysis (72% residual) [128]. Overall, mutations in the acceptor subsites of CGTases generally provoke a diminution of transglycosylation activity [123], whereas donor subsite mutations mostly alter cyclodextrin specificity (α, β and γ ratio, for cyclodextrins composed of 6, 7 and 8 glucose units respectively). In this respect, it is noteworthy that a five-residue loop localized in subsites -3/-4 of α-amylases has been described as a key determinant (steric hindrance) of the T/H partition, since it is absent in CGTases. To test this hypothesis, this loop was deleted in the α-amylase Novamyl (residues 191 to 195), combining this modification with acceptor subsite mutations (F189L/T190Y). This protein engineering work provided CGTase-like behavior [129], but the reverse experiment (i.e. introduction

of a loop in US132 CGTase) failed, since it procured a mutant that was unable to catalyze hydrolysis or even initial β-cyclization [130]. This failure once again underlines the complexity of the phenomenon and supports the notion that hydrolysis is driven by optimized interactions in the donor subsites which in turn contribute to the formation of TS. In this respect, it is interesting to mention that the successful conversion of the aforementioned *Tabium* CGTase into a hydrolase was almost certainly facilitated by the fact that the parental enzyme already displays unusually high hydrolytic ability. This implies that in *Tabium* CGTase the donor interactions required for hydrolysis are already in place and thus it just remains necessary to delete the determinants of transfer activity.

More generally, these studies highlight the role of aromatic/hydrophobic residues in acceptor subsites. Notably, it appears obvious the presence of aromatic residues provides both a stacking platform for better acceptor docking [131] and a hydrophobic barrier, which limits the presence of water in the active site, with both of these factors favoring transglycosylation. Similarly, such features were also suggested to be part of an evolutionary relationship between α-amylases and 4-α-glucanotransferase within family GH57 [132].

Assuming that CGTases are indeed the consequence of the evolution of α-amylases, presumably the former have somehow dealt with the well-developed donor subsite interactions that favor hydrolysis [133]. Theoretically, the existence of intermediate CGTases that display high 'residual' hydrolytic activity, such as the GH13 *Tabium* CGTase [125] or the one from *Bacillus sp. SK 13.002* strain [134], should provide clues as to how this has been achieved, although in reality unravelling subtle molecular differences might actually be a considerable challenge [135].

3.4. Transferring *vs* hydrolyzing sialidases

Sialidases (SA) and *trans*-sialidases (*tr*S) are members of family GH33 and belong to the GH-E clan. These enzymes catalyze either the hydrolysis or the synthesis of sialyl-glycoconjugates respectively, operating via a classical ping-pong bi-bi mechanism with acid/base catalysis [136–138]. *tr*S exhibit both activities although when a suitable acceptor is available transglycosylation is approximately 10-fold higher than hydrolysis [139]. Moreover, in the case of *Tctr*S, the *tr*S from *Trypanosoma cruzi*, the K_M value for the acceptor is lower than that of the donor (10 μM and in the millimolar range for the lactose and sialic acid moieties respectively) [90,137]. Both SA and *tr*S possess similar catalytic domains, displaying a six-bladed β-propeller topology, which are connected via a long α-helix and a large hydrophobic interface to a domain displaying a β-sandwich fold and lectin-like topology. This latter does not appear to be directly involved in transglycosylation activity [140]. The molecular architecture of the active sites of these enzymes displays several common features including eight strictly invariant residues and a hydrophobic pocket that binds the *N*-acetyl group of the sialic acid moiety, suggesting a mutual evolutionary origin and a similar mode of action

for the entire family [91,137,139].

Nevertheless, it has been suggested that subtle structural differences are likely to be responsible for the two observed selectivities. This is illustrated by the comparison of the hydrolytic *Trypanosoma rangeli*'s SA (*Tr*SA) and the transglycosylating *Tctr*S. Although these enzymes share 70% amino acid identity, their active sites display distinctive features [137,140–142]. *Tctr*S exhibits a narrower, more hydrophobic substrate-binding pocket, which implies that the reactive center is less solvent-exposed, and results in an alternative hydrogen bonding pattern with the sialyl donor moiety. Additionally, residue Q284 in *Tr*SA is replaced by P283 in *Tctr*S, a substitution that alters the conformation of the neighboring residue, W312 (W313 in *Tr*SA). In *Tctr*S, W312 and Y119 (S120 in *Tr*SA) form the two lateral walls of the acceptor binding site, providing the basis for stacking interactions with the sugar acceptor. Moreover, it is noteworthy *Tctr*S appears to display greater active site flexibility than *Tr*SA [91,140,142,143], a point that is exemplified by the study of the inherent motions of Y119 and the strictly conserved Y342 (catalytic nucleophile) residues. According to Demir and Roitberg, structural rearrangements that are triggered by 'allosteric' binding of the sialyl-conjugate donor forming a covalent sialyl-enzyme intermediate lead to the creation of a productive acceptor sugar binding site [143].

Overall, the fine-tuning of enzyme-donor substrate interactions, conformational flexibility (notably loops), solvent exposure and the presence of an acceptor sugar-binding site are crucial to obtain *trans*-sialidase activity. Therefore to switch between hydrolysis and transglycosylation, *Tr*SA has been submitted to mutagenesis, introducing five mutations designed to modify the structure and dynamics of the donor subsite and to create a suitable acceptor subsite. These provided *Tr*SA$_{5mut}$, which displayed detectable *trans*-sialidase activity, although this activity was only 1% of that exhibited by the true *trans*-sialidase, *Tctr*S [91]. Further mutation of *Tr*SA$_{5mut}$, introducing either I37L or G342A, which affect the donor subsite, procured a higher transglycosylation rate, which was 11% of that exhibited by *Tctr*S. Therefore, it appears that the acquisition of improved *trans*-sialidase activity requires alterations in the donor subsite, notably to alter the flexibility of the tyrosine nucleophile residue and thus diminish hydrolytic activity. In this respect, it is also significant that while *Tr*SA is inhibited by DANA (K_i = 1.5 µM for *Tr*SA), a structural analog of the transition state sialic acid oxocarbenium ion, the mutated *Tr*SA described above is less sensitive to inhibition (K_i = 1.54 mM) [91], as is the case for *Tctr*S (K_i = 12.3 mM). This implies that the acquisition of *trans*-sialidase activity may involve a modification of the TS that is developed during the glycosylation step.

More recently, using QM/MM approaches Roitberg *et al*. evaluated the free-energy profiles for the conversion of the Michaelis complex to the covalent glycosyl-enzyme intermediate in *Tr*SA, *Tr*SA$_{5mut}$ and *Tctr*S [144,145]. In many SA enzymes, the free-energy barrier (ΔG^{TS1}) to reach the

glycosylated-enzyme intermediate (15.2 and 15.0 kcal.mol^{-1} for *Tr*SA and *Tr*SA$_{5mut}$, respectively) is approximately 5 kcal.mol^{-1} lower than that of *Tctr*S (20.8 kcal.mol^{-1}). Besides,the free-energy change (ΔG^0, Figure 2) for *Tctr*S is nearly zero (-0.89 kcal.mol^{-1}), unlike those of reactions catalyzed by *Tr*SA and *Tr*SA$_{5mut}$ (-10.9 and -9.8 kcal.mol^{-1}, respectively), in which a higher stability glycosyl-enzyme intermediate is formed. However, the deglycosylation step appears to be favorable for *tr*S-like enzymes, with the difference being approximately 5 kcal.mol^{-1} (i.e. 21.6, 24.8 and 26.1 kcal.mol^{-1} for *Tctr*S, *Tr*SA$_{5mut}$ and *Tr*SA, respectively). In the light of these findings, it was possible to perform further *in silico* design of a *trS* and propose *Tr*SA$_{10mut}$, a mutant based on *Tr*SA$_{5mut}$ that contains five additional substitutions. Among these, I40L and G345A (*Tr*SA numbering), which are close to the catalytic nucleophile tyrosine, are thought to be responsible for the proposed impact on the T/H balance. This enzyme is theoretically characterized by a weak stabilization of the covalent intermediate and a free-energy barrier for deglycosylation (-3.2 and 19.1 kcal.mol^{-1}, respectively) that is lower than that of *Tctr*S. Nevertheless, the free-energy barrier of the glycosylation step (ΔG^{TS1} = 16 kcal.mol^{-1}) is closer to that of a typical hydrolytic SA.

Finally, on the edge of the acceptor substrate binding cleft, it is noteworthy that *Tctr*S exhibits a seven-amino acid loop (VTNKKKQ) whose composition, physico-chemical properties and dynamics differ from the equivalent loop (IADMGGR) in *Tr*SA. Using an enzyme engineering approach it was shown that the loop in *Tctr*S promotes transglycosylation, increasing product yield, and reduces hydrolysis, effects that were attributed to a perturbation of the water binding network [146].

4. Engineered transglycosylases

Although TGs have only been identified in a few GHs families, hydrolytic GHs from other families have been submitted to protein engineering in order to modify their H/T balance. In the following section, the different strategies that have been adopted are described along with the results that have been obtained.

4.1. Modification of donor subsite interactions

4.1.1. Enzyme engineering

One of the very first protein engineering studies aimed specifically at increasing the T/H ratio was performed on a GH1 β-glycosidase from *Thermus thermophilus* using a random mutagenesis/screening methodology. The mutation of two conserved residues F401 and N282 in this enzyme increased K_M values (> 6-fold) and significantly improved transglycosylation (up to 78% synthesis yield compared to 8% for the wild-type enzyme) [147]. In a follow-up study, using a site-directed approach, the same authors probed the importance of conserved residues in the donor (-1)

subsite [148] and revealed that these play an important role in TS stabilization (29 to 3577-fold decrease k_{cat}/K_M values) but do not induce major structural changes.

Likewise, working on AMY1, a GH13 α-amylase (clan GH-H), it was shown that the mutation of a subsite -2 residue (M53W) leads to increased lifetime of the glycosyl-enzyme and thus to the acquired ability to perform transglycosylation using *p*NP-α-D-maltoheptaose as the donor [86]. It is also noteworthy that the introduction of a range of mutations at position 53 procured k_{cat}/K_M values that were 59 to 5000-fold lower than that of the parental enzyme (mainly due to up to a 20-fold increase in K_M values). Likewise, it is significant that the presence of tryptophan at position 53 is a common occurrence in GH13 CGTases, which is consistent with the impact of the mutation M53W in AMY1.

The mutation of conserved donor subsites residues produces a similar effect to the one described above in other GH families. This is exemplified by protein engineering work performed on a GH18 chitinase from *Serratia marcescens* (*Sm*ChiA). The latter possesses a long active site cleft positioned at the top of a $(\beta/\alpha)_8$ barrel, the donor subsite of which was targeted with the aim of prolonging the retention time of the donor glycosyl moiety [149]. The introduction of the mutation W167A (subsite -3) procured a higher transglycosylation yield (45% of the substrate converted into transglycosylation products, compared to 8% for wild type *Sm*ChiA) and subsequent determination of the 3D structure of the mutated enzyme revealed that repositioning of D313 (subsite -1) had occurred. This is significant because D313 is involved in the stabilization of the oxazolinium intermediate and interacts with E315, a residue that is putatively responsible for water molecule activation during hydrolysis. Therefore, the mutation W167A might both prolong residency of the donor glycosyl moiety and/or diminish hydrolysis. Similar examples of such a coupled effect (i.e. improved transglycosylation and diminished hydrolysis) are provided by work performed on chitinases from *Serratia proteamaculans* (*Sp*ChiD) [150] and *Aspergillus fumigatus* (*Af*ChiB) [151], with mutations being introduced at the catalytic center and in subsite -1 respectively. Furthermore, QM/MM calculations performed on a hyper-transglycosylating variant (D142N) of ChiB from *S. marcescens* (*Sm*ChiB) predicted that the mutation, which is within a highly conserved DxDxE motif, would affect both TS stabilization and the catalytic water molecule [152].

A further example concerns two homologous α-galactosidases (AgaA and AgaB) from family GH36 (clan GH-D) [153]. Despite being highly related (97% identity), AgaA displays a 'rather low' K_M value for raffinose (K_M = 3.8 mM) and exhibits high hydrolytic activity and no detectable ability to catalyze transglycosylation. On the other hand, AgaB displays a higher K_M value (200 mM) for raffinose and exhibits the ability to catalyze autocondensation reactions (i.e. transglycosylation). In this context, the mutation of residue 355 (Ala in AgaA and Glu in AgaB) provides the means to switch between the

two phenotypes, with for example the substitution A355E in AgaA procuring AgaB-like behavior and *vice versa*. Although residue 355 is located far from the active site (20 Å), structural analysis revealed that the presence of a Glu at position 355 provokes the displacement of the conserved W336, which is present in subsite -1 where it provides the basis for sugar stacking. This modification widens the active site and thus probably disturbs the binding of raffinose.

Regarding another example of a galactose-acting enzyme family, random mutagenesis and screening performed on the GH42 β-galactosidase from *Geobacillus stearothermophilus* (BgaB) pinpointed a residue (R109) for subsequent site-saturation mutagenesis. This procured a final mutant (R109W) that displayed improved ability (23% yield compared to 2% for the parental enzyme) to transfer D-galactosyl moieties onto lactose [154]. R109 is a highly conserved amino acid among GH42 β-galactosidases that, according to 3D structure analyses, is involved in hydrogen binding with the D-galactosyl moiety. Therefore, mutation of this residue probably leads to the destabilization of donor binding in subsite -1 (K_M values on lactose increase from 1.7 to 114 mM), coupled to decreased hydrolysis rate (15% residual) and thus alterations in the T/H ratio that favor transglycosylation. Overall, in terms of TS, it is likely that the mutation R109W increases the TS energy barriers for glycosylation and deglycosylation, thus rendering water-mediated deglycosylation less competitive.

Regarding rational engineering work focused on the nucleophile catalyst, several studies have revealed that modifications of the latter can also have drastic effects on the ability of water to deglycosylate the glycosyl-enzyme intermediate. Recently, it was reported that the introduction of a sulfinate function (i.e. SOO^-), to replace the catalytic nucleophile of the GH13 dextran glucosidase, provoked a drastic drop in k_{cat} (0.27%), an acidic pK_a shift (from 3.9 to 1.5) and an increase in transglucosylation yields [155]. According to the authors of this work, the observed effects can be attributed to differences in the TS energy barriers between water and acceptor-mediated enzyme deglycosylation. In this respect, shortening or lengthening (E78D or carboxymethylation of the mutant E78C) of the nucleophile residue in the GH11 xylanase from *Bacillus circulans* was also shown to be detrimental for global catalytic efficiency, shortening having a greater impact (1600-5000-fold decrease) than nucleophile lengthening (16-100-fold) [156]. However, in this study no information concerning the impact on the T/H ratio was reported. Nevertheless, in a very recent study, nucleophile shortening (E134D) in a GH16 EG was shown to introduce glycosynthase-like activity [157]. The resulting enzyme, which retained 2% residual hydrolytic activity and displayed a modified pK_a value (5.8 instead of 7.0 for the parental enzyme) for its acid/base catalytic residue, was described as a hydrolase-glycosynthase intermediate [158]. Unfortunately, no information regarding the reactivity of the glycosyl-enzyme covalent intermediate towards water or sugar acceptors was reported.

4.1.2. Substrate modifications

The previous section described how modifications in enzyme donor subsites can favor transglycosylation. In a similar manner, several authors have revealed that substrate modifications can procure the same overall effect (i.e. altering the TS energy barrier for water-mediated deglycosylation). An excellent example of this was reported for the GH51 α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*Tx*Abf). This enzyme was shown to display much better transglycosylation yields in the presence of the non-natural donor sugar *p*NP-β-D-Gal*f* (75% when using Bn-α-D-Xyl*p* as the acceptor) than with *p*NP-α-L-Ara*f* (7%) [159,160]. Compared to *p*NP-α-L-Ara*f*, *p*NP-β-D-Gal*f* possesses an extra hydroxymethyl moiety at position C5, a difference that is sufficient to decrease by 100-fold the hydrolytic rate and radically increase the K_M value (> 50 mM, compared to 0.72 mM on *p*NP-α-L-Ara*f*), changes that are clearly indicative of modified glycosylation and deglycosylation steps. It is noteworthy, that similar results have been subsequently observed for the GH51 Abf from *Clostridium thermocellum* [161].

Another example, described over 20 years ago, concerns the GH1 β-glucosidase from *Agrobacterium faecalis* (Abg). When acting on *p*NP-β-D-Xyl*p*, the value of k_{cat}/K_M was divided by 140-fold compared to that obtained with *p*NP-β-D-Fuc*p* (a substrate that contains an extra methyl group at C5), while the T/H ratio for the autocondensation reaction was 4.3 [57]. Moreover, it was shown that when Abg acts on dNP-3-deoxy-3-fluoro-glucoside, the reaction mechanism is altered and is described by a biphasic profile of V_i = f(S) plot, with transglycosylation (autocondensation) occurring above a certain threshold concentration [48]. Through the measurement of k_{cat} values (Box. 2) it was possible to demonstrate that this effect was due to an increase of TS energy barriers associated with glycosylation (6.4 kcal.mol^{-1}) and deglycosylation (12 kcal.mol^{-1}) steps, respectively.

Glycosynthases, which are crippled enzymes in which the catalytic nucleophile is usually mutated into A, S or G, are also good examples of how the donor substrate can diminish the potency of water. In the glycosynthase-mediated reaction, glycosyl fluorides displaying inverted anomery are used as donors. These mimic the covalent glycosyl-enzyme intermediate and, in the presence of an acceptor, allow deglycosylation to occur. Importantly, once the transglycosylation product is formed, it cannot be hydrolyzed by the enzyme, since this is catalytically-impotent, being unable to perform glycosylation. However, this rather neat account of how glycosynthases work fails to explain why water does not compete with the acceptor. In fact, the rather elusive answer to this question concerns the donor-borne fluorine atom [162]. This atom is likely to extensively perturb the donor-enzyme interaction that is necessary to overcome the TS energy barrier and thus only counter-interactions provided by the presence of a sugar in the acceptor subsite are sufficient to decrease the energy of TS2 [25]. In this respect, it is also interesting to note that a correlation has been observed

between the ability to create efficient glycosynthases from GHs and the behavior of the corresponding parental GH with 2-fluorosugars [25]. When high k_{trans} and k_{trans}/k_{H2O} were measured in the presence of such inhibitors, it was predicted that the enzyme would form the basis of a good glycosynthase. Actually, it is highly probable that, in both cases, 'equivalent' TS destabilizations are responsible for this improved selectivity for sugar acceptors over water molecules.

In summary, although very few studies have actually measured catalytic efficiencies of GHs in the presence of different donors and compared transglycosylation rates, it appears likely that the destabilization of the hydrogen bonding network in donor subsites is responsible for increased transglycosylation, irrespective of whether this is achieved through enzyme mutagenesis or donor substrate modification.

4.2. Acceptor subsite interactions and impact on the deglycosylation step

The first links between acceptor specificity and transglycosylation were revealed in the 70's through work performed on lysozyme [163]. The role of residues in acceptor subsites +1/+2 (originally denoted E and F) were shown to be important for transglycosylation activity by contributing to higher binding free energy of the incoming acceptor [164,165]. Similarly, a hypothetical link between transglycosylation and acceptor binding interactions was proposed on the basis of work performed on the GH10 xylanase from *Streptomyces lividans* (XlnA) [166]. In this work, the mutation of a subsite +3 residue (N173D) decreased transglycosylation activity in the presence of xylo-oligosaccharide (DP > 3) acceptors, thus revealing the sensitivity of this reaction to changes in acceptor binding. In another study involving a GH10 xylanase (Xyn10A from *Pseudomonas cellulosa*) the introduction of alanines in subsites +2 (N182), +3 (Y255) and +4 (Y220) provoked a strong decrease in transglycosylation activity [167]. This loss of activity was correlated with increased k_{cat} and K_M on xylan (21- and 22-fold, respectively), suggesting that decreased 'affinity' in the aglycon region accelerates the leaving group departure and facilitates the access of water to the active site.

To better understand the determinants of transglycosylation in members of family GH5, GH5 *endo*-β-(1,4)-mannanases have been subjected to acceptor subsite engineering. The results of such work revealed that stacking interactions (+1 subsite of Man C) and hydrogen bonding with the acceptor moiety (revealed by the deleterious effect of R171K in + 2 subsite of *Tr*Man5A) are critical features for efficient transglycosylation [168,169].

The question of acceptor affinity has also been addressed in the case of the aforementioned chitinase, *Sm*ChiA [152]. By further mutating the hyper-transglycosylating variant, *Sm*ChiA-D313N, introducing the aglycon mutation (subsite +2) F396W, it was possible to further enhance transglycosylation by several fold (not numerically quantified), Unfortunately, in this study the effect

of the mutation F396W alone was not studied, making it impossible to state whether it needs to be combined with D313N in order to observe an effect on transglycosylation. Nevertheless, this example is particularly interesting, because it underlines the importance of the aromatic surface area. The mutation F396W increases this parameter, which might explain the higher rate of the transglycosylation reaction (compared to D313N alone) and the lower catalytic efficiency (the relative k_{cat}/K_M for D313N/F396W and D313N are 1.4 and 4.9% respectively) of this enzyme [152]. Consistent with this result are separate observations that the removal of aromatic residues located in the aglycon subsites of other chitinases, such as that of *Cycas revolute* (*Cr*Chi-A), *Sp*ChiD or *Af*ChiB1 decreases or annihilates transglycosylation activity [150,151,170]. In the case of *Af*ChiB1, computational simulation suggested that the complete loss of transglycosylation coupled to the maintenance of hydrolysis in the mutant W137E would be due to the loss of a stacking interaction between W137 and the acceptor sugar. According to the authors, this interaction would be necessary for the efficient attack of the covalent glycosyl-enzyme intermediate.

Another example of how modifications in the acceptor binding site of GHs can lead to T/H modulation in favor of transglycosylation has been provided for family GH85 *endo*-β-*N*-acetylglucosaminidases (ENGase). The enhancement of transglycosylation procured by either increased hydrophobicity in the acceptor subsite (achieved by the mutation Y217F in the ENGase from *Mucor hiemalis*), or the mutation of so-called 'gate keeper' amino acids (W216 and W244 in the ENGase from *Arthrobacter protophormiae*) was partially attributed to alterations in active site dynamics [171,172].

Although the introduction of aromatic side-chains substitutions into the acceptor binding regions of GHs is frequently used to modulate the T/H ratio, this is neither an exclusive nor a general strategy. Indeed, the introduction of other mutations (P402D and F328A) into the +1 region of the active site of a family GH36 α-D-galactosidase (+1 subsite) actually led to an enlargement of the entrance to the active site and consequently influenced the orientation of the bound acceptor. Such mutations led to 4 to 16-fold increases in the yield of transglycosylation products, although in absolute terms the amounts produced were modest [173]. Another example reveals that the presence of aromatic residues in the acceptor subsite does not automatically favor transglycosylation. Indeed, the introduction of the mutation F116A into the acceptor subsite of the GH39 β-xylosidase from *Bacillus halodurans* (BhXyl39) revealed that the effect of this mutation on transglycosylation was dependent on the nature of the acceptor molecule, with transglycosylation being increased when octanol was used, but decreased in the case of pentanol [174].

In summary, generally-speaking one can affirm that although certain mutations in the acceptor subsites of GHs enhance the T/H ratio, such mutations are much less likely to drastically affect hydrolysis. Consequently, such mutations mostly exhibit limited potential to increase

transglycosylation, since any beneficial effects that maybe associated with acceptor subsite mutations are masked to some extent by the persistence of hydrolysis.

4.3. Water 'activation' and channels

Evidently, water is a key external element in hydrolytic reactions catalyzed by GHs. Therefore, in order to better diminish hydrolysis, it is quite logical that scientists have attempted to elucidate the molecular determinants of water access to the active site and the way in which 'catalytic water' is bound in a productive manner for catalysis.

In the case of T4 lysozyme, which is actually an inverting GH, the O^{γ} of T26 was identified as a solvent binding determinant. Consistent with this hypothesis, mutation of T26 to histidine (T26H) procured transglycosylase activity, with a T/H ratio of 10:1 being observed after a 60 min reaction period [175]. This quite surprising feat (considering that the parental enzyme is an inverting one) is explicable if one considers that the N^{ε} of H26, which lies close to the putative catalytic water binding site, permits the formation of a covalent glycosyl-enzyme intermediate.

In other studies, tyrosine residues have been identified as water binding determinants. For example, the introduction of a tyrosine near to the active site (V286Y) of the *Bacillus licheniformis* α-amylase led to a 5-fold increase in hydrolytic activity on starch compared to the wild type enzyme, consistent with the hypothesis that the tyrosine OH group favors water access to the active site [126]. Similarly, the key role of tyrosine residues has been demonstrated in other GH13 α-amylases [176,177] and in *endo*-xylanases [178,179], with the elimination (mutation to F, A, S or N) of tyrosine generally leading to the loss of anchoring points for water molecules, coupled to increased hydrophobicity and thus overall lower hydrolysis. It is also noteworthy, that the mutation of the water-binding tyrosine (Y198F) in a GH8 *exo*-acting, inverting xylanase (Rex) led to the creation of an unusual glycosynthase that displays a high (4.7 s^{-1}) fluoride release rate when fed with α-xylobiosyl fluoride despite the fact that the enzyme's general base was present [180]. However, in this case the mutation did not remove the nucleophilic water *per se*, but rather led the non-productive orientation of the water's lone electron pair, due to the absence of the H-bond ordinarily furnished by Y198 [162].

Tyrosines that provide catalytic assistance have also been revealed in GH85 enzymes, in particular the *Arthrobacter protophormiae* ENGase A. The creation of mutations Y205F (donor subsite) and Y299F (acceptor region) in this enzyme led to 2.5 and 3-fold increases respectively in transglycosylation yield, though hydrolysis was not affected in the same way. Indeed mutation of Y205 significantly reduced hydrolysis (44% residual activity), consistent with the postulate that the hydroxyl moiety of Y205 activates the catalytic water molecule. However, mutation of Y299 did not

affect hydrolysis [172], an observation that implies that the substitution Y299F mainly improves acceptor binding, and thus transglycosylation.

Overall, the mutation of residues that interact with catalytic water in GHs appears to be a good strategy to reduce hydrolysis and thus modulate the H/T ratio. Attractively, such a strategy reduces the risk of major active site perturbations, while targeting a critical element of hydrolysis.

Another way to alter the hydrolytic potency of GHs is to actually modify water access to the active site. Although protein are generally tightly packed and surrounded by bulk water, internal water molecules are observed in cavities and channels, which in some cases may provide access to the active site, following paths that are determined by a combination of hydrogen bonding, electrostatic interactions and hydrophobic effects [181]. An elegant example of such a channel was found in a substrate-complexed form of the GH13 α-amylase from *Pseudoalteromonas haloplanctis*. X-ray crystallography revealed a series of seven well-ordered water molecules that followed a path from the surface to the enzyme's active site [182]. Similarly, evidence that the alteration of water channels can provide the means to modulate the H/T ratio was shown in the case of the neopullulanase from *Bacillus stearothermophilus*. Here, the introduction of a hydrophobic residue at the entry point of water into the catalytic active site was sufficient to increase transglycosylation. Inversely, the introduction of hydrophilicity produced the opposite effect, increasing hydrolysis [183]. Likewise, in another α-amylase (human salivary type), a water channel composed of aromatic residues was also demonstrated to have a catalytic relevance [184]. Two recent studies probed the role of water channels with respect to the H/T balance in GH1 β-glucosidases from *Thermotoga maritima* and *Thermus thermophilus* respectively [185,186]. In the latter, internal water dynamics were extensively studied using hydrogen-deuterium exchange mass spectrometry (DXMS) and molecular dynamics techniques [186]. Finally, structural analysis of a 1,3-α-3,6-anhydro-L-galactosidase from family GH117 has also revealed the presence of a putative water channel that runs from the protein's surface to the active site [187].

From an evolutionary perspective, the presence of specific water binding determinants (e.g. water-binding tyrosine residues) or water delivery systems within the active site can be considered as extra features that contribute to the enzyme's hydrolytic potency. Nevertheless, the presence of such features is manifestly facultative for activity, because TS stabilization is the key element for catalysis to occur. Indeed, the lack of specific water interactions during the deglycosylation step may explain why this step is usually rate-limiting in the majority of GHs.

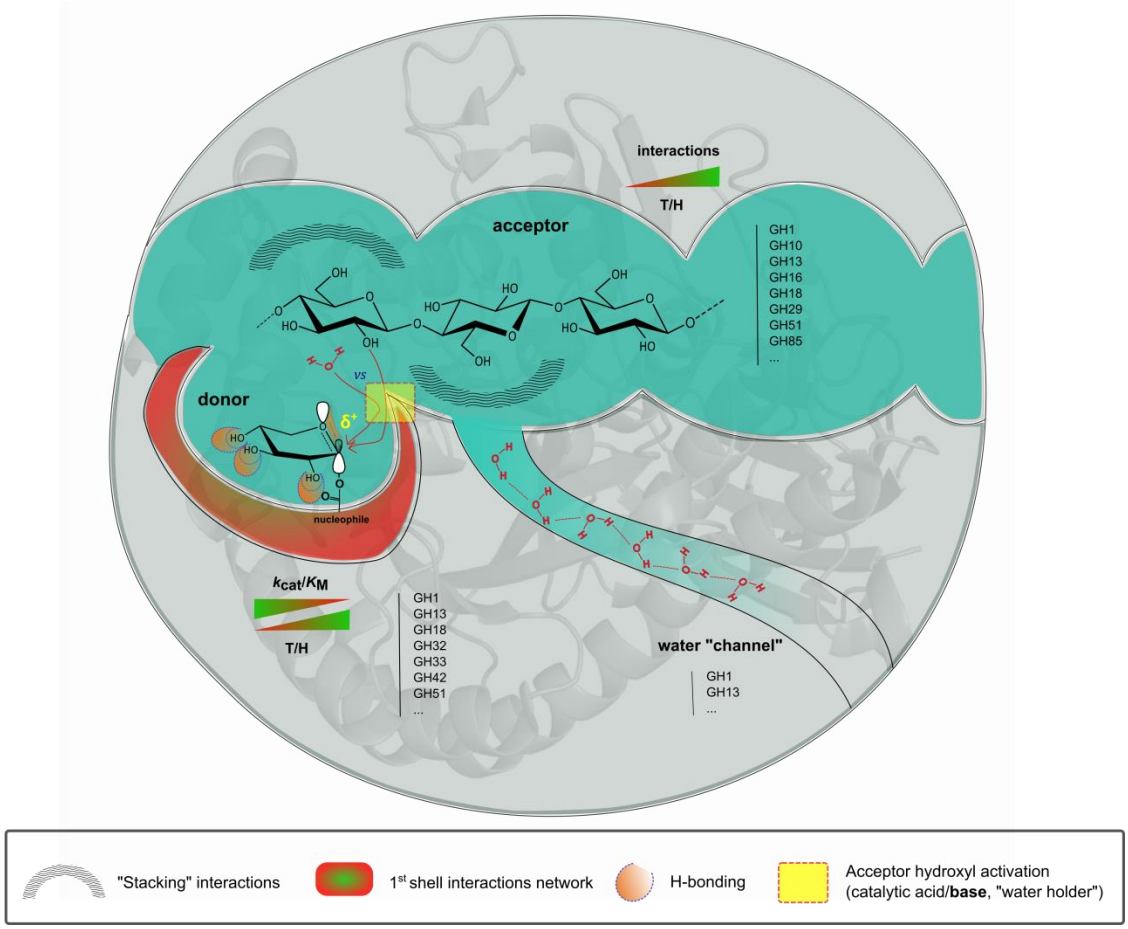


Figure 3 View of an enzyme active site and potential molecular factors influencing the T/H balance. Along the reaction pathway this snapshot represents the transition state of the deglycosylation step in the course of which the anomeric center of the covalent glycosyl-enzyme intermediate undergoes a nucleophilic attack by either an activated water molecule (hydrolysis) or a glycosidic acceptor (transglycosylation).

5. Conclusions

Hopefully, this quite extensive review of available literature has succeeded in revealing a certain number of generic features that allow GHs to perform transglycosylation in adverse reaction conditions, dominated by the presence of water.

Unsurprisingly, the first lesson that can be drawn from the examples that are described herein is that TGs are probably characterized by modified transition states when compared to their GH counterparts, as illustrated by our 'locked door' analogy. This leads to the conclusion that the first condition that must be met if a GH is to be converted into an efficient TG is that new transition state structures or electronic displacement systems must be developed that lower the efficiency of water-mediated deglycosylation, thus radically reducing the inherent advantage of ubiquitous water over other suitably-nucleophilic acceptors. The analysis of available data indicates that the destabilization of $TS2_{water}$ is generally brought about by altering interactions in the donor subsite, which inevitably

results in higher transition state energy barriers (c.f. enhanced stiffness of the lock mechanism) for both deglycosylation and glycosylation (TS2 and TS1 respectively), since they display similar properties. In turn, this translates into less efficient catalysis (i.e. low values of k_{cat}/K_M).

The second lesson that can be drawn from the cumulative knowledge presented here is that transglycosylation is also promoted by the presence of substrate-specific interactions in the acceptor subsites (better grip on the door handle). These are useful, because they partially compensate for the destabilization of the donor subsite interactions, possibly providing a situation where $TS2_{acceptor}$ energy is lower than that of $TS2_{water}$. Moreover, these can be very specific for certain acceptor molecules, thus their presence does not necessarily imply more general access for acceptors, including water. In this respect, it is noteworthy that the conversion of *endo*-GHs towards TGs might not be necessarily reflected by significantly lowered k_{cat}/K_M values since both donor and acceptor regions are involved in the donor substrate recognition, with relative contributions that are probably GH family-dependent. As a consequence, the analysis of *endo*-enzymes evolution may constitute more intricate case of studies than *exo*-acting ones.

Finally, despite the fact that the natural solvent for most enzymes is water, most of this remains as bulk water. For catalysis many GHs display specific features that either fix specific water molecules within the vicinity of the catalytic center and/or deliver water to the active site via channels. In both cases, these water-specific molecular determinants can be modified in order to reduce the presence of catalytic water, thus further promoting transglycosylation.

In the case of natural TGs, evolutionary processes have no doubt resulted in the combination of the aforementioned features. However, it is certain that the foremost of these must be the destabilization of $TS2_{water}$, a modification that we consider to be the main driving force of the GH→TG transition. Unfortunately, from an enzyme engineering perspective, the rational modification of TS structures (geometry and/or electronic features) remains impossible due to insufficient knowledge of these. Nevertheless, in this respect, the use of combined approaches, associating QM/MM and experimental validation, should provide new information in the near future. Moreover, with the increasing integration of knowledge in the fields of chemistry and biology, it is likely that new comprehension of carbohydrate reactivity in chemically-catalyzed glycosylation will rapidly contribute to better understanding of the enzyme-substrate interactions that are necessary for transglycosylation to occur.

From an evolutionary standpoint, it appears that TGs have mostly evolved from their hydrolytic counterparts, although *endo*-acting XTH might be an exception to this general rule [83]. Therefore, it is intriguing to speculate upon the circumstances that might have necessitated the conversion of a catalytically-optimized GHs into today's relatively inefficient TGs. Surely, it would have been more

logical to leave glycosynthesis to the dedicated glycosynthetic enzymes (i.e. GTs) that we know today, unless evolutionary divergence between GHs and TGs occurred at a point when no other alternatives were available?

Despite the long history of carbohydrate chemistry, the fact that some fundamental features of glycosylation (e.g. the factors that influence anomeric center reactivity and stereo-electronic effects) are still subjects of research [188,189] underlines the complexity of this important field and reveals the fact that there are still considerable opportunities for progress. In the case of synthetic glycochemistry, the extension of the available palette of TGs is undoubtedly an attractive grail for enzyme engineers. From this point of view, the knowledge that is presented herein should provide a rational starting point for would-be enzyme engineers wishing to convert GHs into TGs.

Acknowledgements

The authors would like to thank Drs. Leila Lo Leggio, Marco Moracci, Antoni Planas and Morten Sørlie for their critical appraisal of this manuscript.

References

1 Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W. and Etzler, M. E. (2009) In Essentials of Glycobiology, Cold Spring Harbor Lab Press

2 Edelman, J. (1956) The formation of oligosaccharides by enzymic transglycosylation. Adv. Enzym. Relat. Areas Mol. Biol. **17**, 189–232

3 Breton, C., Fournel-Gigleux, S. and Palcic, M. M. (2012) Recent structures, evolution and mechanisms of glycosyltransferases. Curr. Opin. Struct. Biol. **22**, 540–549.

4 Lairson, L. L., Henrissat, B., Davies, G. J. and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. **77**, 521–555

5 Field, R. A. (2011) Glycobiology: challenging reaction equilibria. Nat. Chem. Biol. **7**, 658–659

6 Gantt, R. W., Peltier-Pain, P., Cournoyer, W. J. and Thorson, J. S. (2011) Using simple donors to drive the equilibria of glycosyltransferase-catalyzed reactions. Nat. Chem. Biol. **7**, 685–691

7 Cote, G. L. and Tao, B. Y. (1990) Oligosaccharide synthesis by enzymatic transglycosylation. Glycoconjugate J. **7**, 145–162

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

9 Malet, C. and Planas, A. (1998) 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

10 Moracci, M., Trincone, A. and Rossi, M. (2001) Glycosynthases: new enzymes for oligosaccharide synthesis. J. Mol. Catal. B: Enzym. **11**, 155–163

11 Perugino, G., Trincone, A., Rossi, M. and Moracci, M. (2004) Oligosaccharide synthesis by glycosynthases. Trends Biotechnol. **22**, 31–37

12 Faijes, M. and Planas, A. (2007) In vitro synthesis of artificial polysaccharides by glycosidases and glycosynthases. Carbohydr. Res. **342**, 1581–1594

13 Rakić, B. and Withers, S. G. (2009) Recent developments in glycoside synthesis with glycosynthases and thioglycoligases. Aust. J. Chem. **62**, 510–520

14 Cobucci-Ponzano, B. and Moracci, M. (2012) Glycosynthases as tools for the production of glycan analogs of natural products. Nat. Prod. Rep. **29**, 697–709

15 Schmaltz, R. M., Hanson, S. R. and Wong, C.-H. (2011) Enzymes in the synthesis of glycoconjugates. Chem. Rev. **111**, 4259–4307

16 Kobayashi, S. and Makino, A. (2009) Enzymatic polymer synthesis: an opportunity for green polymer chemistry. Chem. Rev. **109**, 5288–5353

17 Kadokawa, J. (2011) Precision polysaccharide synthesis catalyzed by enzymes. Chem. Rev. **111**, 4308–4345

18 Zechel, D. L. and Withers, S. G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. Acc. Chem. Res. **33**, 11–18

19 Vasella, A., Davies, G. J. and Böhm, M. (2002) Glycosidase mechanisms. Curr. Opin. Chem. Biol. **6**, 619–629

20 Vuong, T. V and Wilson, D. B. (2010) Glycoside hydrolases: catalytic base/nucleophile diversity. Biotechnol. Bioeng. **107**, 195–205

21 Henrissat, B. (1991) A classification of glycosyl hydrolases based sequence similarities amino acid. Biochem. J. **280**, 309–316

22 Henrissat, B. and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. **7**, 637–644

23 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. **42**, D490–D495

24 Street, I. P., Kempton, J. B. and Withers, S. G. (1992) Inactivation of a *β*-glucosidase through the accumulation of a stable 2-deoxy-2-fluoro-*α*-D-glucopyranosyl-enzyme intermediate: a detailed investigation. Biochemistry **31**, 9970–9978

25 Ducros, M.-A., Tarling, C. A., Zechel, D. L., Brzozowski, A. M., Frandsen, T. P., Ossowski, I. Von, Withers, S. G. and Davies, G. J. (2003) Anatomy of glycosynthesis : structure and kinetics of the *Humicola insolens* Cel7B E197A and E197S glycosynthase mutants. Chem. Biol. **10**, 619–628

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

27 Fischer, E. (1894) Einfluss der configuration auf die wirkung der enzyme. Ber. Dtsch. Chem. Ges. **27**, 2985–2993

28 Koshland, D. E. (1995) The key–lock theory and the induced fit theory. Angew. Chem. Int. Ed. **33**, 2375–2378

29 Wolfenden, R. and Snider, M. J. (2001) The depth of chemical time and the power of enzymes as catalysts. Acc. Chem. Res. **34**, 938–945

30 Pauling, L. (1946) Molecular architecture and biological reactions. Chem. Eng. News **24**, 1375–1377

31 Wolfenden, R., Lu, X. and Young, G. (1998) Spontaneous hydrolysis of glycosides. J. Am. Chem. Soc. **120**, 6814–6815

32 Warshel, A. (1981) Electrostatic basis of structure-function correlation in proteins. Acc. Chem. Res. **14**, 284–290

33 Tokuriki, N. and Tawfik, D. S. (2009) Protein dynamism and evolvability. Science **324**, 203–207

34 Chouard, T. (2011) Breaking the protein rules. Nature **471**, 151–153

35 Glowacki, D. R., Harvey, J. N. and Mulholland, A. J. (2012) Taking Ockham's razor to enzyme dynamics and catalysis. Nat. Chem. **4**, 169–176.

36 Eisenthal, R., Danson, M. J. and Hough, D. W. (2007) Catalytic efficiency and k_{cat}/K_M: a useful comparator? Trends Biotechnol. **25**, 247–249

37 Johnson, K. A. (2013) A century of enzyme kinetic analysis, 1913 to 2013. FEBS Lett. **587**, 2753–2766

38 Sinnott, M. L. (1990) Catalytic mechanism of enzymic glycosyl transfer. Chem. Rev. **90**, 1171–1202

39 Vocadlo, D. J. and Davies, G. J. (2008) Mechanistic insights into glycosidase chemistry. Curr. Opin. Chem. Biol. **12**, 539–555

40 Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. and Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 α-L-arabinofuranosidase. EMBO J. **22**, 4922–4932

41 Davies, G. J., Planas, A. and Rovira, C. (2012) Conformational analyses of the reaction coordinate of glycosidases. Acc. Chem. Res. **45**, 308–316

42 Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M. and Davies, G. J. (1996) Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. Biochemistry **35**, 15280–15287

43 Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M. and Brzozowski, A. M. (1998) Snapshots along an enzymatic reaction coordinate : analysis of a retaining *β*-glycosidase. Biochemistry **37**, 11707–11713

44 Biarnés, X., Nieto, J., Planas, A. and Rovira, C. (2006) Substrate distortion in the Michaelis complex of *Bacillus* 1,3-1,4-*β*-glucanase. Insight from first principles molecular dynamics simulations. J. Biol. Chem. **281**, 1432–1441

45 Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A. and Rovira, C. (2011) Catalytic itinerary in 1,3-1,4-β-glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state. J. Am. Chem. Soc. **133**, 20301–20309

46 Nerinckx, W., Desmet, T. and Claeyssens, M. (2003) A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. FEBS Lett. **538**, 1–7

47 Zechel, D. L., Reid, S. P., Stoll, D., Nashiru, O., Warren, R. A. J. and Withers, S. G. (2003) Mechanism, mutagenesis, and chemical rescue of a *β*-mannosidase from *cellulomonas fimi*. Biochemistry **42**, 7195–7204

48 Namchuk, M. N. and Withers, S. G. (1995) Mechanism of *Agrobacterium β*-glucosidase: kinetic analysis of the role of noncovalent enzyme/substrate interactions. Biochemistry **34**, 16194–16202

49 Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. a and Withers, S. G. (1998) Insights into transition state stabilization of the β-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants. Nat. Struct. Biol. **5**, 812–818

50 Sierks, M. R. and Svensson, B. (1992) Kinetic identification of a hydrogen bonding pair in the glucoamylase-maltose transition state complex. Protein Eng. **5**, 185–188

51 Kim, Y.-W., Zhang, R., Chen, H. and Withers, S. G. (2010) *O*-glycoligases, a new category of glycoside bond-forming

mutant glycosidases, catalyse facile syntheses of isoprimeverosides. Chem. Commun. **46**, 8725–8727.
52 Heightman, T. D. and Vasella, A. T. (1999) Recent insights into inhibition , structure , and mechanism of configuration-retaining glycosidases. Angew. Chem. Int. Ed. **38**, 750–770
53 Nerinckx, W., Desmet, T., Piens, K. and Claeyssens, M. (2005) An elaboration on the syn-anti proton donor concept of glycoside hydrolases: electrostatic stabilisation of the transition state as a general strategy. FEBS Lett. **579**, 302–312
54 Schramm, V. L. (2011) Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. Annu. Rev. Biochem. **80**, 703–732
55 Bigeleisen, J. and Mayer, M. G. (1947) Calculation of equilibrium constants for isotopic exchange reactions. J. Chem. Phys. **15**, 261–267
56 Cleland, W. W. (2003) The use of isotope effects to determine enzyme mechanisms. J. Biol. Chem. **278**, 51975–84
57 Kempton, J. B. and Withers, S. G. (1992) Mechanism of *Agrobacterium β*-glucosidase: kinetic studies. Biochemistry **31**, 9961–9969
58 Wang, J., Hou, Q., Sheng, X., Gao, J., Liu, Y. and Liu, C. (2013) Theoretical study on the deglycosylation mechanism of rice BGlu1 β-glucosidase. Int. J. Quantum Chem. **113**, 1071–1075
59 Wang, J., Hou, Q., Dong, L., Liu, Y. and Liu, C. (2011) QM/MM studies on the glycosylation mechanism of rice BGlu1 β-glucosidase. J. Mol. Graph. Model. **30**, 148–152
60 Brás, N. F., Ramos, M. J. and Fernandes, P. A. (2010) DFT studies on the β-glycosidase catalytic mechanism: The deglycosylation step. J. Mol. Struct. THEOCHEM **946**, 125–133
61 Badieyan, S., Bevan, D. R. and Zhang, C. (2012) Probing the active site chemistry of *β*-glucosidases along the hydrolysis reaction pathway. Biochemistry **51**, 8907–8918
62 Eyring, H. (1935) The Activated Complex in Chemical Reactions. J. Chem. Phys. **3**, 107–115
63 Demetrius, L. (1998) Role of enzyme-substrate flexibility in catalytic activity: an evolutionary perspective. J. Theor. Biol. **194**, 175–194
64 Del-Rio, G., Morett, E. and Soberon, X. (1997) Did cyclodextrin glycosyltransferases evolve from α-amylases? FEBS Lett. **416**, 221–224
65 Ritsema, T., Hernández, L., Verhaar, A., Altenbach, D., Boller, T., Wiemken, A. and Smeekens, S. (2006) Developing fructan-synthesizing capability in a plant invertase via mutations in the sucrose-binding box. Plant J. **48**, 228–237
66 Luang, S., Cho, J.-I., Mahong, B., Opassiri, R., Akiyama, T., Phasai, K., Komvongsa, J., Sasaki, N., Hua, Y., Matsuba, Y., Ozeki, Y., Jeo, J.-S. and Ketudat Cairns, J. R. (2013) Rice Os9BGlu31 is a transglucosidase with the capacity to equilibrate phenylpropanoid, flavonoid, and phytohormone glycoconjugates. J. Biol. Chem. **288**, 10111–10123
67 Schröder, R., Atkinson, R. G. and Redgwell, R. J. (2009) Re-interpreting the role of endo-β-mannanases as mannan endotransglycosylase/hydrolases in the plant cell wall. Ann. Bot-London **104**, 197–204
68 Johnston, S. L., Prakash, R., Chen, N. J., Kumagai, M. H., Turano, H. M., Cooney, J. M., Atkinson, R. G., Paull, R. E., Cheetamun, R., Bacic, A., Brummell, D. A. and Schröder R. (2013) An enzyme activity capable of endotransglycosylation of heteroxylan polysaccharides is present in plant primary cell walls. Planta **237**, 173–187
69 Franková, L. and Fry, S. C. (2011) Phylogenetic variation in glycosidases and glycanases acting on plant cell wall polysaccharides, and the detection of transglycosidase and trans-β-xylanase activities. Plant J. **67**, 662–81
70 Franková, L. and Fry, S. C. (2012) Trans-α-xylosidase and trans-β-galactosidase activities, widespread in plants, modify and stabilize xyloglucan structures. Plant J. **71**, 45–60
71 Fry, S. C. (2004) Primary cell wall metabolism : tracking the careers of wall polymers in living plant cells. New Phytol. **161**, 641–675
72 Campbell, P. and Braam, J. (1999) Xyloglucan endotransglycosylases: diversity of genes, enzymes and potential wall-modifying functions. Trends Plant. Sci. **4**, 361–366
73 Rose, J. K. C., Braam, J., Fry, S. C. and Nishitani, K. (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. Plant Cell Physiol. **43**, 1421–1435
74 Baumann, M. J., Eklöf, J. M., Michel, G., Kallas, A. M., Teeri, T. T., Czjzek, M. and Brumer, H. (2007) Structural evidence for the evolution of xyloglucanase activity from xyloglucan *endo*-transglycosylases: biological implications for cell wall metabolism. Plant Cell **19**, 1947–1963
75 Eklöf, J. M. and Brumer, H. (2010) The XTH gene family: an update on enzyme structure, function, and phylogeny in xyloglucan remodeling. Plant Physiol. **153**, 456–466
76 Saura-Valls, M., Fauré, R., Ragàs, S., Piens, K., Brumer, H., Teeri, T. T., Cottaz, S., Driguez, H. and Planas, A. (2006) Kinetic analysis using low-molecular mass xyloglucan oligosaccharides defines the catalytic mechanism of a *Populus* xyloglucan endotransglycosylase. Biochem. J. **395**, 99–106
77 Saura-Valls, M., Fauré, R., Brumer, H., Teeri, T. T., Cottaz, S., Driguez, H. and Planas, A. (2008) Active-site mapping of a *Populus* xyloglucan *endo*-transglycosylase with a library of xylogluco-oligosaccharides. J. Biol. Chem **283**, 21853–21863
78 Sulová, Z., Takáčová, M., Steele, N. M., Fry, S. C. and Farkaš, V. (1998) Xyloglucan endotransglycosylase : evidence for the existence of a relatively stable glycosyl–enzyme intermediate. Biochem. J. **330**, 1475–1480
79 Steele, N. M. and Fry, S. C. (1999) Purification of xyloglucan endotransglycosylases (XETs): a generally applicable and simple method based on reversible formation of an enzyme–substrate complex. Biochem. J. **340**, 207–211
80 Piens, K., Fauré, R., Sundqvist, G., Baumann, M. J., Saura-Valls, M., Teeri, T. T., Cottaz, S., Planas, A., Driguez, H. and

Brumer, H. (2008) Mechanism-based labeling defines the free energy change for formation of the covalent glycosyl-enzyme intermediate in a xyloglucan *endo*-transglycosylase. J. Biol. Chem. **283**, 21864–21872

81 Johansson, P., Brumer, H., Baumann, M. J., Kallas, A. M., Henriksson, H., Denman, S. E., Teeri, T. T. and Jones, T. A. (2004) Crystal structures of a poplar xyloglucan endotransglycosylase reveal details of transglycosylation acceptor binding. Plant Cell **16**, 874–886

82 Fauré, R., Saura-Valls, M., Brumer, H., Planas, A., Cottaz, S. and Driguez, H. (2006) Synthesis of a library of xylogluco-oligosaccharides for active-site mapping of xyloglucan *endo*-transglycosylase. J. Org. Chem. **71**, 5151–5161

83 Eklöf, J. M., Shojania, S., Okon, M., McIntosh, L. P. and Brumer, H. (2013) Structure-function analysis of a broad specificity *Populus trichocarpa* endo-*β*-glucanase reveals an evolutionary link between bacterial licheninases and plant *XTH* gene products. J. Biol. Chem. **288**, 15786–15799

84 Mark, P., Baumann, M. J., Eklöf, J. M., Gullfot, F., Michel, G., Kallas, A. M., Teeri, T. T., Brumer, H. and Czjzek, M. (2009) Analysis of nasturtium *Tm*NXG1 complexes by crystallography and molecular dynamics provides detailed insight into substrate recognition by family GH16 xyloglucan *endo*-transglycosylases and *endo*-hydrolases. Proteins **75**, 820–836

85 Van der Veen, B. A., Uitdehaag, J. C., Penninga, D., van Alebeek, G. J., Smith, L. M., Dijkstra, B. W. and Dijkhuizen, L. (2000) Rational design of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 to increase α-cyclodextrin production. J. Mol. Biol. **296**, 1027–1038

86 Mori, H., Bak-Jensen, K. S. and Svensson, B. (2002) Barley α-amylase Met53 situated at the high-affinity subsite −2 belongs to a substrate binding motif in the β→α loop 2 of the catalytic $(\beta/\alpha)_8$-barrel and is critical for activity and substrate specificity. Eur. J. Biochem. **269**, 5377–5390

87 Potocki de Montalk, G., Remaud-Simeon, M., Willemot, R. M., Sarçabal, P., Planchot, V. and Monsan, P. (2000) Amylosucrase from *Neisseria polysaccharea*: novel catalytic properties. FEBS Lett. **471**, 219–223

88 Kim, M.-I., Kim, H.-S., Jung, J. and Rhee, S. (2008) Crystal structures and mutagenesis of sucrose hydrolase from *Xanthomonas axonopodis* pv. *glycines*: insight into the exclusively hydrolytic amylosucrase fold. J. Mol. Biol. **380**, 636–647

89 Schroeven, L., Lammens, W., Van Laere, A. and Van den Ende, W. (2008) Transforming wheat vacuolar invertase into a high affinity sucrose:sucrose 1-fructosyltransferase. New Phytol. **180**, 822–831

90 Ribeirão, M., Pereira-Chioccola, V. L., Eichinger, D., Rodrigues, M. M. and Schenkman, S. (1997) Temperature differences for *trans*-glycosylation and hydrolysis reaction reveal an acceptor binding site in the catalytic mechanism of *Trypanosoma cruzi trans*-sialidase. Glycobiology **7**, 1237–1246

91 Paris, G., Ratier, L., Amaya, M. F., Nguyen, T., Alzari, P. M. and Frasch, A. C. C. (2005) A sialidase mutant displaying *trans*-sialidase activity. J. Mol. Biol. **345**, 923–934

92 Fry, S. C., York, W. S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J.-P., Kato, Y., Lorences, E. P., Maclachlan, G. A., McNeil, M., Mort, A. J., Grant Reid, J. S., Seitz, H. U., Selvendran, R. R., Voragen, A. G. J. and White A. R. (1993) An unambigous nomenclature for xyloglucan-derived oligosaccharides. Physiol. Plant. **89**, 1–3

93 Monsan, P., Remaud-Siméon, M. and André, I. (2010) Transglucosidases as efficient tools for oligosaccharide and glucoconjugate synthesis. Curr. Opin. Microbiol. **13**, 293–300

94 Van Hijum, S. A. F. T., Kralj, S., Ozimek, L. K., Dijkhuizen, L. and van Geel-schutten, I. G. H. (2006) Structure-Function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. Microbiol. Mol. Biol. Rev. **70**, 157–176

95 Monchois, V., Willemot, R. M. and Monsan, P. (1999) Glucansucrases: mechanism of action and structure-function relationships. FEMS Microbiol. Rev. **23**, 131–151

96 Leemhuis, H., Pijning, T., Dobruchowska, J. M., van Leeuwen, S. S., Kralj, S., Dijkstra, B. W. and Dijkhuizen, L. (2013) Glucansucrases: three-dimensional structures, reactions, mechanism, α-glucan analysis and their implications in biotechnology and food applications. J. Biotechnol. **163**, 250–272

97 Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M. and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. Protein Eng. Des. Sel. **19**, 555–562

98 Vujičić-Zagar, A., Pijning, T., Kralj, S., López, C. A., Eeuwema, W., Dijkhuizen, L. and Dijkstra, B. (2010) Crystal structure of a 117 kDa glucansucrase fragment provides insight into evolution and product specificity of GH70 enzymes. Proc. Natl. Acad. Sci. USA **107**, 21406–21411

99 Ito, K., Ito, S., Shimamura, T., Weyand, S., Kawarasaki, Y., Misaka, T., Abe, K., Kobayashi, T., Cameron, A. D. and Iwata, S. (2011) Crystal structure of glucansucrase from the dental caries pathogen *Streptococcus mutans*. J. Mol. Biol. **408**, 177–186

100 Brison, Y., Pijning, T., Malbert, Y., Fabre, É., Mourey, L., Morel, S., Potocki-Véronèse, G., Monsan, P., Tranier, S., Remaud-Siméon, M., et al. (2012) Functional and structural characterization of *α*-(1→2) branching sucrase derived from DSR-E glucansucrase. J. Biol. Chem. **287**, 7915–7924

101 Leemhuis, H., Pijning, T., Dobruchowska, J. M., Dijkstra, B. W. and Dijkhuizen, L. (2012) Glycosidic bond specificity of glucansucrases: on the role of acceptor substrate binding residues. Biocatal. Biotransfor. **30**, 366–376

102 Kralj, S., van Geel-Schutten, G. H., van der Maarel, M. J. E. C. and Dijkhuizen, L. (2004) Biochemical and molecular characterization of *Lactobacillus reuteri* 121 reuteransucrase. Microbiol. **150**, 2099–2112

103 Kralj, S., van Leeuwen, S. S., Valk, V., Eeuwema, W., Kamerling, J. P. and Dijkhuizen, L. (2008) Hybrid

reuteransucrase enzymes reveal regions important for glucosidic linkage specificity and the transglucosylation/hydrolysis ratio. FEBS J. **275**, 6002–6010

104 Monsan, P. F. and Ouarne, F. (2009) Oligosaccharides derived from sucrose. In Prebiotics and Probiotics Science and Technology (Charalampopoulos, D; and Rastall, R. A., eds.), pp 293–336, Springer

105 Kralj, S., Grijpstra, P., van Leeuwen, S. S., Leemhuis, H., Dobruchowska, J. M., van der Kaaij, R. M., Malik, A., Oetari, A., Kamerling, J. P. and Dijkhuizen, L. (2011) 4,6-α-Glucanotransferase, a novel enzyme that structurally and functionally provides an evolutionary link between glycoside hydrolase enzyme families 13 and 70. Appl. Environ. Microbiol. **77**, 8154–8163

106 Skov, L. K., Mirza, O., Henriksen, A., De Montalk, G. P., Remaud-Siméon, M., Sarçabal, P., Willemot, R. M., Monsan, P. and Gajhede, M. (2001) Amylosucrase, a glucan-synthesizing enzyme from the *α*-amylase family. J. Biol. Chem. **276**, 25273–25278

107 Guérin, F., Barbe, S., Pizzut-Serin, S., Potocki-Véronèse, G., Guieysse, D., Guillet, V., Monsan, P., Mourey, L., Remaud-Siméon, M., André, I. and Tranier, S. (2012) Structural investigation of the thermostability and product specificity of amylosucrase from the bacterium *Deinococcus geothermalis*. J. Biol. Chem. **287**, 6642–6654

108 Albenne, C., Skov, L. K., Mirza, O., Gajhede, M., Feller, G., D'Amico, S., André, G., Potocki-Véronèse, G., van der Veen, B., Monsan, P. and Remaud-Simeon, M. (2004) Molecular basis of the amylose-like polymer formation catalyzed by *Neisseria polysaccharea* amylosucrase. J. Biol. Chem. **279**, 726–734

109 Cambon, E., Barbe, S., Pizzut-Serin, S., Remaud-Simeon, M. and André, I. (2014) Essential role of amino acid position 226 in oligosaccharide elongation by amylosucrase from *Neisseria polysaccharea*. Biotechnol. Bioeng. **111**, 1719–1728

110 Champion, E., Guérin, F., Moulis, C., Barbe, S., Tran, T. H., Morel, S., Descroix, K., Monsan, P., Mourey, L., Mulard, L., Tranier, S., Remaud-Siméon, M. and André, I. (2012) Applying pairwise combinations of amino acid mutations for sorting out highly efficient glucosylation tools for chemo-enzymatic synthesis of bacterial oligosaccharides. J. Am. Chem. Soc. **134**, 18677–18688

111 Daudé, D., Champion, E., Morel, S., Guieysse, D., Remaud-Siméon, M. and André, I. (2013) Probing substrate promiscuity of amylosucrase from *Neisseria polysaccharea*. ChemCatChem **5**, 2288–2295

112 Kim, H., Park, H., Heu, S. and Jung, J. (2004) Molecular and functional characterization of a unique sucrose hydrolase from *Xanthomonas axonopodis* pv. glycines. J. Bacteriol. **186**, 411–418

113 Skov, L. K., Pizzut-Serin, S., Remaud-Siméon, M., Ernst, H. A., Gajhede, M. and Mirza, O. (2013) The structure of amylosucrase from *Deinococcus radiodurans* has an unusual open active-site topology. Acta Crystallogr. Sect. F **69**, 973–978

114 Pizzut-Serin, S., Potocki-Véronèse, G., van der Veen, B. A., Albenne, C., Monsan, P. and Remaud-Simeon, M. (2005) Characterisation of a novel amylosucrase from *Deinococcus radiodurans*. FEBS Lett. **579**, 1405–1410

115 Olvera, C., Centeno-Leija, S., Ruiz-Leyva, P. and López-Munguía, A. (2012) Design of chimeric levansucrases with improved transglycosylation activity. Appl. Environ. Microbiol. **78**, 1820–1825

116 Lammens, W., Le Roy, K., Schroeven, L., Van Laere, A., Rabijns, A. and Van den Ende, W. (2009) Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications. J. Exp. Bot. **60**, 727–740

117 Van den Ende, W., Lammens, W., Van Laere, A., Schroeven, L. and Le Roy, K. (2009) Donor and acceptor substrate selectivity among plant glycoside hydrolase family 32 enzymes. FEBS J. **276**, 5788–5798

118 Lafraya, Á., Sanz-Aparicio, J., Polaina, J. and Marín-Navarro, J. (2011) Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase. Appl. Environ. Microbiol. **77**, 6148–6157

119 Ritsema, T., Verhaar, A., Vijn, I. and Smeekens, S. (2005) Using natural variation to investigate the function of individual amino acids in the sucrose-binding box of fructan:fructan 6G-fructosyltransferase (6G-FFT) in product formation. Plant Mol. Biol. **58**, 597–607

120 Altenbach, D., Rudiño-Pinera, E., Olvera, C., Boller, T., Wiemken, A. and Ritsema, T. (2009) An acceptor-substrate binding site determining glycosyl transfer emerges from mutant analysis of a plant vacuolar invertase and a fructosyltransferase. Plant Mol. Biol. **69**, 47–56

121 Álvaro-Benito, M., Sainz-Polo, M. A., González-Pérez, D., González, B., Plou, F. J., Fernández-Lobato, M. and Sanz-Aparicio, J. (2012) Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis β*-fructofuranosidase, an enzyme that produces prebiotics. J. Biol. Chem. **287**, 19674–19686

122 Uitdehaag, J. C. M. (1999) The cyclization mechanism of cyclodextrin glycosyltransferase (CGTase) as revealed by a *γ*-cyclodextrin-CGTase complex at 1.8-Å resolution. J. Biol. Chem. **274**, 34868–34876

123 Leemhuis, H., Kelly, R. M. and Dijkhuizen, L. (2010) Engineering of cyclodextrin glucanotransferases and the impact for biotechnological applications. Appl. Microbiol. Biotechnol. **85**, 823–835

124 Van der Veen, B. A., Leemhuis, H., Kralj, S., Uitdehaag, J. C., Dijkstra, B. W. and Dijkhuizen, L. (2001) Hydrophobic amino acid residues in the acceptor binding site are main determinants for reaction mechanism and specificity of cyclodextrin-glycosyltransferase. J. Biol. Chem. **276**, 44557–44562

125 Kelly, R. M., Leemhuis, H. and Dijkhuizen, L. (2007) Conversion of a cyclodextrin glucanotransferase into an α-amylase: assessment of directed evolution strategies. Biochemistry **46**, 11216–11222

126 Rivera, M. H., López-Munguía, A., Soberón, X. and Saab-Rincón, G. (2003) α-Amylase from Bacillus licheniformis mutants near to the catalytic site: effects on hydrolytic and transglycosylation activity. Protein Eng. **16**, 505 –514

127 Kim, T., Park, C., Cho, H., Cha, S., Kim, J., Lee, S., Moon, T., Kim, J., Oh, B. and Park, K. (2000) Role of the glutamate

332 residue in the transglycosylation activity of *Thermus* maltogenic amylase. Biochemistry 6773–6780
128 Tran, P. L., Cha, H.-J., Lee, J.-S., Park, S.-H., Woo, E.-J. and Park, K.-H. (2014) Introducing transglycosylation activity in *Bacillus licheniformis* α-amylase by replacement of His235 with Glu. Biochem. Biophys. Res. Commun., doi: 10.1016/j.bbrc.2014.08.019
129 Beier, L., Svendsen, A., Andersen, C., Frandsen, T. P., Borchert, T. V and Cherry, J. R. (2000) Conversion of the maltogenic α-amylase Novamyl into a CGTase. Protein Eng. **13**, 509–513
130 Jemli, S., Ben-Ali, M., Ben-Hlima, H., Khemakhem, B. and Bejar, S. (2012) Mutations affecting the activity of the cyclodextrin glucanotransferase of *Paenibacillus pabuli* US132: insights into the low hydrolytic activity of cyclodextrin glucanotransferases. Biologia **67**, 636–643
131 Asensio, J. L., Ardá, A., Cañada, F. J. and Jiménez-Barbero, J. (2013) Carbohydrate-aromatic interactions. Acc. Chem. Res. **46**, 946–954
132 Janeček, S. and Blesák, K. (2011) Sequence-structural features and evolutionary relationships of family GH57 α-amylases and their putative α-amylase-like homologues. Prot. J. **30**, 429–435
133 Janeček, S., Svensson, B. and MacGregor, E. A. (2003) Relation between domain evolution, specificity, and taxonomy of the α-amylase family members containing a C-terminal starch-binding domain. Eur. J. Biochem. **270**, 635–645
134 Sun, T., Letsididi, R., Pan, B. and Jiang, B. (2013) Production of a novel cyclodextrin glycosyltransferase from *Bacillus* sp. SK13.002. Afr. J. Microbiol. Res. **7**, 2311–2315
135 Bornscheuer, U. T. and Kazlauskas, R. J. (2004) Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. Angew. Chem., Int. Ed. **43**, 6032–6040
136 Watts, A. G., Damager, I., Amaya, M. L., Buschiazzo, A., Alzari, P., Frasch, A. C. and Withers, S. G. (2003) *Trypanosoma cruzi* trans-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. J. Am. Chem. Soc. **125**, 7532–7533
137 Amaya, M. F., Watts, A. G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A., Paris, G., Frasch, A. C., Withers, S. G. and Alzari, P. M. (2004) Structural insights into the catalytic mechanism of *Trypanosoma cruzi trans*-sialidase. Structure **12**, 775–784
138 Damager, I., Buchini, S., Amaya, M. F., Buschiazzo, A., Alzari, P., Frasch, A. C., Watts, A. and Withers, S. G. (2008) Kinetic and mechanistic analysis of *Trypanosoma cruzi* trans-sialidase reveals a classical ping-pong mechanism with acid/base catalysis. Biochemistry **47**, 3507–3512
139 Montagna, G., Cremona, M. L., Paris, G., Amaya, M. F., Buschiazzo, A., Alzari, P. M. and Frasch, A. C. C. (2002) The *trans*-sialidase from the african trypanosome *Trypanosoma brucei*. Eur. J. Biochem. **269**, 2941–2950.
140 Buschiazzo, A., Tavares, G. A., Campetella, O., Spinelli, S., Cremona, M. L., París, G., Amaya, M. F., Frasch, A. C. and Alzari, P. M. (2000) Structural basis of sialyltransferase activity in trypanosomal sialidases. EMBO J. **19**, 16–24
141 Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C. and Alzari, P. M. (2002) The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. Mol. Cell **10**, 757–768
142 Amaya, M. F., Buschiazzo, A., Nguyen, T. and Alzari, P. M. (2003) The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T. cruzi trans*-sialidase. J. Mol. Biol. **325**, 773–784
143 Demir, O. and Roitberg, A. E. (2009) Modulation of catalytic function by differential plasticity of the active site: case study of *Trypanosoma cruzi trans*-sialidase and *Trypanosoma rangeli* sialidase. Biochemistry **48**, 3398–3406
144 Pierdominici-Sottile, G., Horenstein, N. A. and Roitberg, A. E. (2011) A free energy study of the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. From the Michaelis complex to the covalent intermediate. Biochemistry **50**, 10150–10158
145 Pierdominici-Sottile, G., Palma, J. and Roitberg, A. E. (2013) Free energy computations identify the mutations required to confer trans-sialidase activity into *Trypanosoma rangeli* sialidase. Proteins **82**, 424–435
146 Jers, C., Michalak, M., Larsen, D. M., Kepp, K. P., Li, H., Guo, Y., Kirpekar, F., Meyer, A. S. and Mikkelsen, J. D. (2014) Rational design of a new *Trypanosoma rangeli* trans-sialidase for efficient sialylation of glycans. PLoS One **9**, e83902
147 Feng, H.-Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C. and Dion, M. (2005) Converting a *β*-glycosidase into a *β*-transglycosidase by directed evolution. J. Biol. Chem. **280**, 37088–37097
148 Teze, D., Hendrickx, J., Czjzek, M., Ropartz, D., Sanejouand, Y.-H., Tran, V., Tellier, C. and Dion, M. (2014) Semi-rational approach for converting a GH1 β-glycosidase into a β-transglycosidase. Protein Eng. Des. Sel. **27**, 13–19
149 Aronson, N. N., Halloran, B. A., Alexeyev, M. F., Zhou, X. E., Wang, Y., Meehan, E. J. and Chen, L. (2006) Mutation of a conserved tryptophan in the chitin-binding cleft of *Serratia marcescens* chitinase A enhances transglycosylation. Biosci., Biotechnol., Biochem. **70**, 243–251
150 Madhuprakash, J., Tanneeru, K., Purushotham, P., Guruprasad, L. and Podile, A. R. (2012) Transglycosylation by chitinase D from *Serratia proteamaculans* improved through altered substrate interactions. J. Biol. Chem. **287**, 44619–44627
151 Lü, Y., Yang, H., Hu, H., Wang, Y., Rao, Z. and Jin, C. (2009) Mutation of Trp137 to glutamate completely removes transglycosyl activity associated with the *Aspergillus fumigatus* AfChiB1. Glycoconjugate J. **26**, 525–534
152 Zakariassen, H., Hansen, M. C., Jøranli, M., Eijsink, V. G. H. and Sørlie, M. (2011) Mutational effects on transglycosylating activity of family 18 chitinases and construction of a hypertransglycosylating mutant. Biochemistry **50**, 5693–5703

153 Merceron, R., Foucault, M., Haser, R., Mattes, R., Watzlawick, H. and Gouet, P. (2012) The molecular mechanism of thermostable *α*-galactosidases AgaA and AgaB explained by X-ray crystallography and mutational studies. J. Biol. Chem. **287**, 39642–39652

154 Placier, G., Watzlawick, H., Rabiller, C. and Mattes, R. (2009) Evolved β-galactosidases from *Geobacillus stearothermophilus* with improved transgalactosylation yield for galacto-oligosaccharide production. Appl. Environ. Microbiol. **75**, 6312–6321

155 Saburi, W., Kobayashi, M., Mori, H., Okuyama, M. and Kimura, A. (2013) Replacement of the catalytic nucleophile aspartyl residue of dextran glucosidase by cysteine sulfinate enhances transglycosylation activity. J. Biol. Chem. **288**, 31670–31677

156 Lawson, S. L., Wakarchuk, W. W. and Withers, S. G. (1996) Effects of both shortening and lengthening the active site nucleophile of *Bacillus circulans* xylanase on catalytic activity. Biochemistry **35**, 10110–10108

157 Andrés, E., Aragunde, H. and Planas, A. (2014) Screening glycosynthase libraries with a fluoride chemosensor assay independently of enzyme specificity: identification of a transitional hydrolase to synthase mutant. Biochem. J. **458**, 355–363

158 Aragunde, H., Castilla, E., Biarnés, X., Faijes, M. and Planas, A. (2014) A transitional hydrolase to glycosynthase mutant by Glu to Asp substitution at the catalytic nucleophile in a retaining glycosidase. Carbohydr. Res. **389**, 85–92

159 Rémond, C., Plantier-Royon, R., Aubry, N., Maes, E., Bliard, C. and O'Donohue, M. J. (2004) Synthesis of pentose-containing disaccharides using a thermostable α-L-arabinofuranosidase. Carbohydr. Res. **339**, 2019–2025

160 Rémond, C., Plantier-Royon, R., Aubry, N. and O'Donohue, M. J. (2005) An original chemoenzymatic route for the synthesis of β-D-galactofuranosides using an α-L-arabinofuranosidase. Carbohydr. Res. **340**, 637–644

161 Chlubnova, I., Filipp, D., Spiwok, V., Dvorakova, H., Daniellou, R., Nugier-Chauvin, C., Kralova, B. and Ferrières, V. (2010) Enzymatic synthesis of oligo-D-galactofuranosides and L-arabinofuranosides: from molecular dynamics to immunological assays. Org. Biomol. Chem. **8**, 2092–2102

162 Hidaka, M., Fushinobu, S., Honda, Y., Wakagi, T., Shoun, H. and Kitaoka, M. (2010) Structural explanation for the acquisition of glycosynthase activity. J. Biochem. **147**, 237–244

163 Pollock, J. J. and Sharon, N. (1970) Acceptor specificity of the lysozyme-catalyzed transglycosylation reaction. Biochemistry **9**, 3913–3925

164 Toshima, G., Kawamura, S., Araki, T. and Torikata, T. (2003) Histidine-114 at subsites E and F can explain the characteristic enzymatic activity of guinea hen egg-white lysozyme. Biosci., Biotechnol., Biochem. **67**, 540–546

165 Kawamura, S., Eto, M., Imoto, T., Ikemizu, S., Araki, T. and Torikata, T. (2004) Functional and structural effects of mutagenic replacement of Asn37 at subsite F on the lysozyme-catalyzed reaction. Biosci., Biotechnol., Biochem. **68**, 593–601

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

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

168 Rosengren, A., Hägglund, P., Anderson, L., Pavon-Orozco, P., Peterson-Wulff, R., Nerinckx, W. and Stålbrand, H. (2012) The role of subsite +2 of the *Trichoderma reesei* β-mannanase TrMan5A in hydrolysis and transglycosylation. Biocatal. Biotransform. **30**, 338–352

169 Dilokpimol, A., Nakai, H., Gotfredsen, C. H., Baumann, M. J., Nakai, N., Abou Hachem, M. and Svensson, B. (2011) Recombinant production and characterisation of two related GH5 endo-β-1,4-mannanases from *Aspergillus nidulans* FGSC A4 showing distinctly different transglycosylation capacity. BBA-Proteins Proteom. **1814**, 1720–1729

170 Taira, T., Fujiwara, M., Dennhart, N., Hayashi, H., Onaga, S., Ohnuma, T., Letzel, T., Sakuda, S. and Fukamizo, T. (2010) Transglycosylation reaction catalyzed by a class V chitinase from cycad, *Cycas revoluta*: A study involving site-directed mutagenesis, HPLC, and real-time ESI-MS. BBA-Proteins Proteom. **1804**, 668–675

171 Umekawa, M., Huang, W., Li, B., Fujita, K., Ashida, H., Wang, L.-X. and Yamamoto, K. (2008) Mutants of *Mucor hiemalis* endo-*β*-*N*-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities. J. Biol. Chem. **283**, 4469–4479

172 Yin, J., Li, L., Shaw, N., Li, Y., Song, J. K., Zhang, W., Xia, C., Zhang, R., Joachimiak, A., Zhang, H.-C., et al. (2009) Structural basis and catalytic mechanism for the dual functional endo-β-N-acetylglucosaminidase A. PLoS One **4**, e4658

173 Bobrov, K. S., Borisova, A. S., Eneyskaya, E. V, Ivanen, D. R., Shabalin, K. A., Kulminskaya, A. A. and Rychkov, G. N. (2013) Improvement of the efficiency of transglycosylation catalyzed by α-galactosidase from *Thermotoga maritima* by protein engineering. Biochemistry (Moscow) **78**, 1112–1123

174 Ochs, M., Belloy, N., Dauchez, M., Muzard, M., Plantier-Royon, R. and Rémond, C. (2013) Role of hydrophobic residues in the aglycone binding subsite of a GH39 β-xylosidase in alkyl xylosides synthesis. J. Mol. Catal. B: Enzym. **96**, 21–26

175 Kuroki, R., Weaver, L. H. and Matthews, B. W. (1999) Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. Proc. Natl. Acad. Sci. USA **96**, 8949 –8954

176 Matsui, I., Yoneda, S., Ishikawa, K., Miyairi, S., Fukui, S., Umeyama, H. and Honda, K. (1994) Roles of the aromatic residues conserved in the active center of *Saccharomycopsis α*-amylase for transglycosylation and hydrolysis

activity. Biochemistry **33**, 451–458
177 Mizuno, M., Tonozuka, T., Uechi, A., Ohtaki, A., Ichikawa, K., Kamitori, S., Nishikawa, A. and Sakano, Y. (2004) The crystal structure of *Thermoactinomyces vulgaris* R ‐ 47 α-amylase II (TVA II) complexed with transglycosylated product. Eur. J. Biochem. **271**, 2530–2538
178 Collins, T., De Vos, D., Hoyoux, A., Savvides, S. N., Gerday, C., Van Beeumen, J. and Feller, G. (2005) Study of the active site residues of a glycoside hydrolase family 8 xylanase. J. Mol. Biol. **354**, 425–435
179 Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J. and Yaguchi, M. (1994) Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. Protein Sci. **3**, 467–475
180 Honda, Y., Fushinobu, S., Hidaka, M., Wakagi, T., Shoun, H., Taniguchi, H. and Kitaoka, M. (2008) Alternative strategy for converting an inverting glycoside hydrolase into a glycosynthase. Glycobiology **18**, 325 –330
181 Purkiss, A., Skoulakis, S. and M., G. J. (2001) The protein–solvent interface : a big splash. Phil. Trans. R. Soc. Lond. A **359**, 1515–1527
182 Aghajari, N., Roth, M. and Haser, R. (2002) Crystallographic evidence of a transglycosylation reaction : ternary complexes of a psychrophilic α-amylase. Biochemistry **41**, 4273–4280
183 Kuriki, T., Kaneko, H., Yanase, M., Takata, H., Shimada, J., Handa, S., Takada, T., Umeyama, H. and Okada, S. (1996) Controlling substrate preference and transglycosylation activity of neopullulanase by manipulating steric constraint and hydrophobicity in active center. J. Biol. Chem. **271**, 17321–17329
184 Ramasubbu, N., Ragunath, C., Sundar, K., Mishra, P. J. and Kandra, L. (2005) Structure-function relationships in human salivary α-amylase : role of aromatic residues a b c. Biologia **60**, 47–56
185 Frutuoso, M. A. and Marana, S. R. (2013) A single amino acid residue determines the ratio of hydrolysis to transglycosylation catalyzed by β-glucosidases. Protein Pept. Lett. **20**, 102–106
186 Teze, D., Hendrickx, J., Dion, M., Tellier, C., Woods, V. L., Tran, V. and Sanejouand, Y.-H. (2013) Conserved water molecules in family 1 glycosidases: a DXMS and molecular dynamics study. Biochemistry **52**, 5900–5910
187 Rebuffet, E., Groisillier, A., Thompson, A., Jeudy, A., Barbeyron, T., Czjzek, M. and Michel, G. (2011) Discovery and structural characterization of a novel glycosidase family of marine origin. Environ. Microbiol. **13**, 1253–1270
188 Jensen, H. H. and Bols, M. (2006) Stereoelectronic substituent effects. Acc. Chem. Res. **39**, 259–265
189 Mydock, L. K. and Demchenko, A. V. (2010) Mechanism of chemical *O*-glycosylation: from early studies to recent discoveries. Org. Biomol. Chem. **8**, 497–510

Part II

PENTOSES IN GLYCOSYNTHESIS: FROM PLANT CELL WALLS TO TAILOR-MADE MOLECULES

Within Nature, pentose carbohydrates are abundant in plant cell walls and thus constitute a colossal reserve of renewable material that should not be underexploited or neglected in biorefinery schemes. In this section, after providing an overall description of plant cell wall structures, underlining their complexity, focus is given to the main pentose-containing hemicelluloses, arabinoxylans. The very recent knowledge pertaining to the *in planta* biosynthesis of arabinoxylans (AX) as well as the diverse properties of its constituent elements, arabinoxylo-oligosaccharides (AXOS), are described. In the final part of this section, the *in vitro* methods used for the preparation of plant cell wall-like structures will be reviewed, notably those involving chemo-enzymatic approaches, and giving special focus to the preparation of pentose and furanose-containing compounds.

II-1. Plant Cell Walls and Biocatalysis: Two Keys for Sustainability

"*Sustainable development is development that meets the needs of the present without compromising future generations to meet their own needs*" (Bruntland commission, 1987).

According to demographic predictions, in 2050 nine billion human beings will be living on Earth and all of them will need to have access to energy and food. The consequence of this is an inevitable rise in demand for raw materials. Nevertheless, regarding energy, current policy aims to replace fossil-based energy, a transition that is driven by the desire to lower environmental impacts of human activities and to anticipate the end of cheap oil, which is predicted to occur around 2040 (Lichtenthaler & Peters, 2004). More recently, the fear of dwindling oil resources has been somewhat appeased by the exploitation of shale gas and by the revival of the coal industry, the latter being an exceptionally cheap fuel that is being massively exploited to fuel China's industrial growth. Nevertheless, the use of any fossil resource is by definition doomed because of its finite nature. Therefore, The transition to a partial and maybe later total bio-based system production taking into account sustainability criteria is inevitable and has to be prepared now (Langeveld, Dixon, & Jaworski, 2010).

It is noteworthy that the concept of the bio-economy (i.e. one that uses bio-based resources as raw materials for energy and manufacturing) is underpinned by several facts. First, it is certainly not the first time that industrialists and scientists have sought ways to produce chemicals and energy from renewable resources. Chemurgy, which emerged in the USA in the 1910's, advocated the development of non-food uses of existing crops and profitable application of agricultural wastes and residues (e.g. acetone or glycerol production using fermentation processes during the 1st World War). However, faced with fierce competition from coal and oil, this branch of applied chemistry was given up due to the withdrawal of governmental support. Second, the drive towards the bio-economy is also motivated by a growing demand from society for more sustainable and safer processes. This pressing requirement is at the heart of the Green Chemistry movement[2], initiated by Paul Anastas's Pollution Prevention Act (1990), which calls for hazard control rather than the management of consequences (i.e. controlling intrinsic *vs* circumstantial factors). Thirdly, it is significant that modern society generates huge amounts of vegetal waste. The better conversion of crop residues (e.g., straw and husk from rice cultivation or sugar cane bagasse) is necessary. To illustrate this, in India, one of the biggest rice producers, rice straw residues represent 97.2 million

[2] Historically, Green Chemistry is known under several names, such as Environmentally Benign Chemistry, Clean Chemistry, Atom Economy or Benign by Design Chemistry. The most widely accepted definition for Green Chemistry would be "the design, development and implementation of chemical processes and products to reduce or eliminate substances hazardous to human health and the environment" (Poliakoff, Fitzpatrick, Farren, & Anastas, 2002).

tons, 23% of which are not collected, burnt or left in the field (Gadde, Menke, & Wassmann, 2009). Indeed, one main advantage of biomass-based energy and chemicals is that farmers and foresters already produce a great deal of residues. Though an important part is necessary to protect habitat, soil and nutrient cycles, millions of tons could be safely collected. For instance, 39 million tons of crop residues go unused in USA every year (Perlack, DOE (USA) 2005). Therefore, to supply the "bio" raw material to the bio-based economy, biomass (from plants, insects) is available, the point is to find sustainable methods to exploit it.

While chemistry and physics were the principle driving forces behind industrial development throughout the 20th century, biology will be the principle source of innovation and progress in the 21st century, with the rise of this discipline being largely attributable to previous achievements made in physics and chemistry. The Organization of Economic Co-operation and Development (OECD) defines biotechnology as **"the application of scientific and engineering principles to the processing of materials by biological agents"**. Industrial or "white" biotechnology is the use of biotechnology for the production of energy, chemicals and materials with greatly improved environmental performances relative to petrochemical-based processes (Hermann, Blok, & Patel, 2007). It is a technology based on enzymes and microorganisms able to convert raw materials to provide compounds, which can be directly used as such, or involved as building blocks in further modifications or construction (e.g., bio-based polymers). Importantly, these biotechnological tools will facilitate the production of novel chemicals and materials displaying new properties, which would be difficult or even impossible to access via conventional processes, or at least the eco-friendly production of drop-in bio-based replacements of existing chemicals.

As part of potential biological tools, biocatalysts are used since the beginning of the 20th century in a wide range of fields, from pharmaceutical to polymer production, including health, cosmetics, food, feed, biofuels and agro domains. Despite many advantages listed hereinafter, biocatalysts need to be as efficient as chemo-catalytic processes in terms of cost and reaction time. These economic key features are still lagging behind facing robust chemical processes although some bioprocesses are already well established, such as those for the production of PLA (Bevan & Franssen, 2006), 1-3-propanediol (Dupont), biodiesel or oleochemicals (Gallezot, 2012).

The advantages offered by biocatalysts are hereinafter ordered and described, using the different principles of Green Chemistry as the classification matrix.

On raw materials

The need to adopt a more sustainable economic growth trajectory, particularly regarding the production and use of energy and chemicals, is motivating the use of plant biomass resources as an exploitable source of renewable carbon (Ragauskas et al., 2006) for biorefinery purposes (O'Donohue

& Monsan, 2007) . In this respect, carbohydrates, which constitute 75% of the annual renewable carbon production (or approximately 200 billion tons) (Lichtenthaler & Peters, 2004) are at the heart of the emerging bio-economy. However, carbohydrates are much more complex in terms of structure and diversity than fossil hydrocarbons, thus their use as a raw material is quite tricky.

The main repositories of carbohydrates are plant cell walls (Fig. 1), which are composite structures of cellulose (35-50%), hemicelluloses (20-35%), lignins (10-25%) and structural proteins, in which the exact proportions of the different components is variable depending on botanic origin, tissues and developmental stage (Badal C, 2000; Carpita, 1996; Cosgrove, 2005). Cellulose microfibrils constitute the main structural component of plant cell walls, whereas hemicelluloses constitute secondary features that play nonetheless essential roles, coating and cross-linking cellulose microfibrils, while ensuring cell wall plasticity (Salmen & Olsson, 1998).

Hemicelluloses are heterogeneous polymers that can be composed of either pentoses (D-xylose, L-arabinose), hexoses (D-mannose, D-glucose, D-galactose) and/or uronic acid subunits. The exact nature of the hemicelluloses present within a cell wall will depend on the aforementioned features, with primary cell wall hemicelluloses being different to those observed in secondary cell walls (Scheller & Ulvskov, 2010).

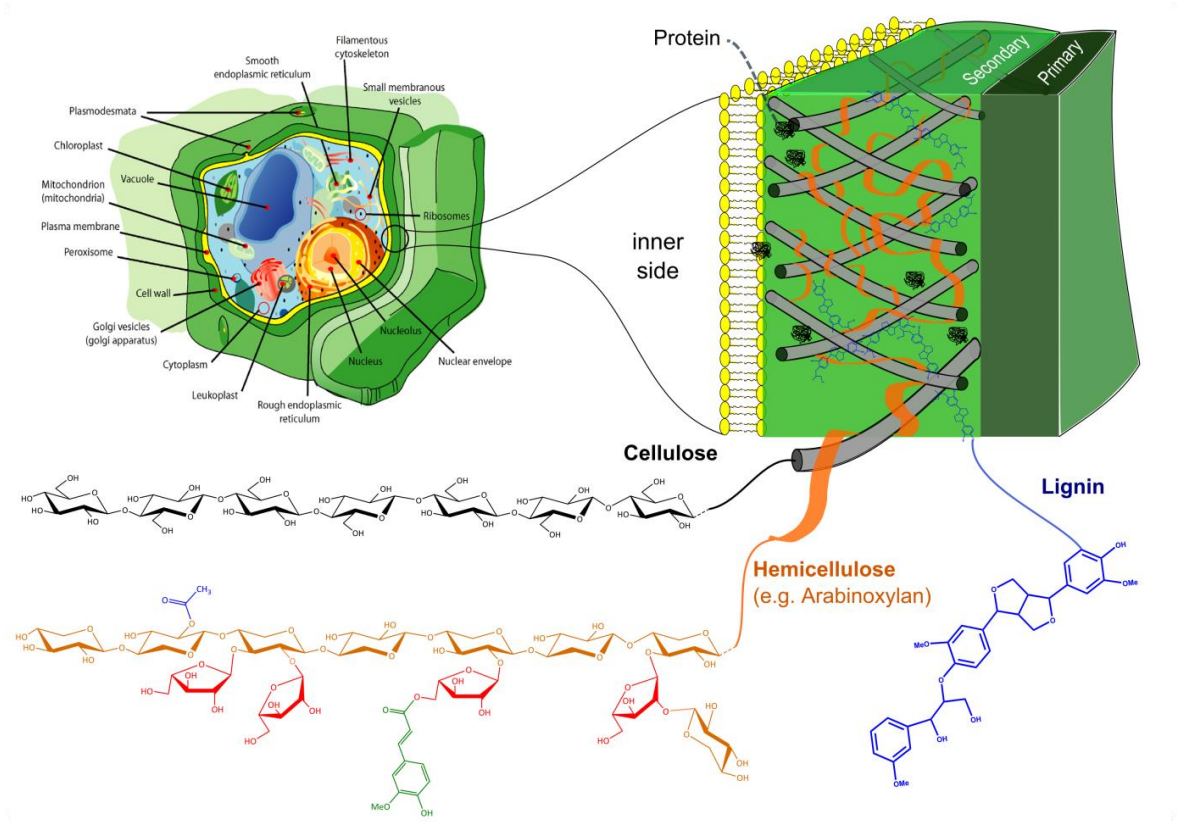


Fig. 1: Plant Cell wall Structure. Two types of wall are found within plant cells, the primary and secondary walls. The first one is found at the junction of cells at the outer edges of the secondary wall. The primary wall is made of 90% polysaccharides (celluloses, hemicelluloses and pectin) and 10 % of proteins.

Given the complex structural organization of plant cell walls, it is logical that a myriad of enzymatic activities is required to deconstruct plant cell walls (Gilbert, Stålbrand, & Brumer, 2008; Gilbert, 2010). In this respect, cellulases (i.e. enzymes that degrade cellulose) have been extensively studied for decades, although hemicellulases (the vast array of enzymes that degrade hemicelluloses) have received a little less attention, partly because of their greater diversity and also because of the primacy of cellulose as a renewable carbon repository. Nevertheless, since the need for renewable carbon is increasing, hemicelluloses are also important carbon repositories and thus the biotechnological tools required to use them are subject to development and implementation, notably in agro-industrial processes (Alvira, Ballesteros, & Negro, 2013; Banerjee, Scott-Craig, & Walton, 2010; Gírio et al., 2010). Indeed this is logical, because pentose-rich raw materials can also be used to manufacture biofuels, chemicals and bio-based polymers, substituting fossil feedstocks (Dumon et al., 2012), and providing a more favourable CO_2 balance.

On reaction conditions

First, by their very nature, enzymes can be considered as non-toxic catalysts, which is an advantage *per se*. Second, the operating conditions of these biocatalysts are compatible with the requirements of eco-friendly processes. Generally, enzymes are active in water, a "green solvent" *par excellence* and sometimes in supercritical fluids. Moreover, most enzymes operate in a moderate temperature range of 10-80 °C and at atmospheric pressure. Moreover, catalysts and more specifically biocatalysts allow accelerating a reaction leading to less energy-consuming processes compared to chemical ones. Indeed, for instance the enzymatic hydrolysis of glycosidic bonds carried out by glycoside hydrolases (GHs), provides an acceleration rate (k_{cat}/k_{uncat}) of up to 10^{17}-fold (Wolfenden et al., 1999).

On Target molecules

One of the main advantages of enzyme-catalyzed reactions is that these lead to less side products due to the high stereo-selectivity and specificity of the catalyst and provide purer end products that require less downstream purification, a feature that is often a major cost factor in industrial processes.

References

Alvira, P., Ballesteros, M. & Negro, M.J. (2013) Progress on enzymatic saccharification technologies for biofuels production. In: *Biofuel Technologies*. pp.145–169.

Badal C, S. (2000) α-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. *Biotechnol. Adv.*, 18 (5), pp.403–423.

Banerjee, G., Scott-Craig, J.S. & Walton, J.D. (2010) Improving enzymes for biomass conversion: a basic research perspective. *BioEnergy Research*, 3 (1), pp.82–92.

Bevan, M.W. & Franssen, M.C.R. (2006) Investing in green and white biotech. *Nat. Biotech.*, 24 (7), pp.765–767.

Bruntland commission (1987) *Our Common Future: World Commission on Environment and Development*.

Carpita, N.C. (1996) Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 47, pp.445–476.

Cosgrove, D.J. (2005) Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.*, 6 (11), pp.850–61.

Dumon, C., Song, L., Bozonnet, S., Fauré, R. & O'Donohue, M.J. (2012) Progress and future prospects for pentose-specific biocatalysts in biorefining. *Process Biochem.*, 47 (3), pp.346–357.

Gadde, B., Menke, C. & Wassmann, R. (2009) Rice straw as a renewable energy source in India, Thailand, and the Philippines: Overall potential and limitations for energy contribution and greenhouse gas mitigation. *Biomass Bioenergy*, 33 (11), pp.1532–1546.

Gallezot, P. (2012) Conversion of biomass to selected chemical products. *Chem. Soc. Rev.*, 41, pp.1538–1558.

Gilbert, H.J. (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol.*, 153 (2), pp.444–455.

Gilbert, H.J., Stålbrand, H. & Brumer, H. (2008) How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr. Opin. Plant Biol.*, 11 (3), pp.338–348.

Gírio, F.M., Fonseca, C., Carvalheiro, F., Duarte, L.C., Marques, S. & Bogel-Łukasik, R. (2010) Hemicelluloses for fuel ethanol: A review. *Bioresour. technol.*, 101 (13), pp.4775–4800.

Hermann, B.G., Blok, K. & Patel, M.K. (2007) Producing bio-based bulk Chemicals using industrial biotechnology saves energy and combats climate change. *Environ. Sci. Technol.*, 41 (22), pp.7915–7921.

Langeveld, J.W. a., Dixon, J. & Jaworski, J.F. (2010) Development perspectives of the biobased economy: a review. *Crop Sci.*, 50 (Supplement 1), pp.S142–S151.

Lichtenthaler, F.W. & Peters, S. (2004) Carbohydrates as green raw materials for the chemical industry. *C. R. Chim.*, 7 (2), pp.65–90.

O'Donohue, M.J. & Monsan, P. (2007) Adaptation des procédés, biotechnologies blanches. *Biofutur*, (282), pp.28–32.

Perlack, Robert D., Lynn L Wright, Anthony F Turhollow, Robin L Graham, Bryce J Stokes, D.C.E. (2005) *Biomass as feedstock for a bioenergy and bioproducts industry : the technical feasibility of a billion-ton annual supply - United State Department of Energy*.

Poliakoff, M., Fitzpatrick, J.M., Farren, T.R. & Anastas, P.T. (2002) Green Chemistry: science and politics of change. *Science*, 297 (5582), pp.807 –810.

Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R. & Tschaplinski, T. (2006) The path forward for biofuels and biomaterials. *Science*, 311 (5760), pp.484–489.

Salmen, L. & Olsson, A.M. (1998) Interaction between hemicelluloses, lignin and cellulose: Structure-property relationships. *J. Pulp Pap. Sci.*, 24 (3), pp.99–103.

Scheller, H.V. & Ulvskov, P. (2010) Hemicelluloses. *Annu. Rev. Plant Biol.*, 61, pp.263–289.

Wolfenden, R., Snider, M., Ridgway, C. & Miller, B. (1999) The temperature dependence of enzyme rate enhancements. *J. Am. Chem. Soc.*, 121 (32), pp.7419–7420.

II-2. Arabinoxylans: nature, functions and properties

1. Supramolecular structure

As previously mentionned, hemicelluloses are the second most abundant polysaccharide in the plant kingdom. Hemicelluloses comprise xyloglucans (XGs), xylans, mannans and galactomanans (food reserve polysaccharide) (Scheller & Ulvskov, 2010). XGs are the most abundant hemicelluloses in flowering plant primary walls (except the grasses). Regarding xylans, three types can be found in Nature: glucuronoarabinoxylans (GAXs) (softwoods, lignified tissues of grasses and annual plants), neutral arabinoxylans (cereals) and glucuronoxylans (hardwood). AXs, found in primary and secondary plant cell wall, are polymers constituted by a backbone of β-(1,4)-linked D-xylopyranoside units chain, each unit can be mono- or di-substituted by L-arabinofuranosyl moieties, with α-(1,2) and/or α-(1,3) glycosidic linkages (Fig.).

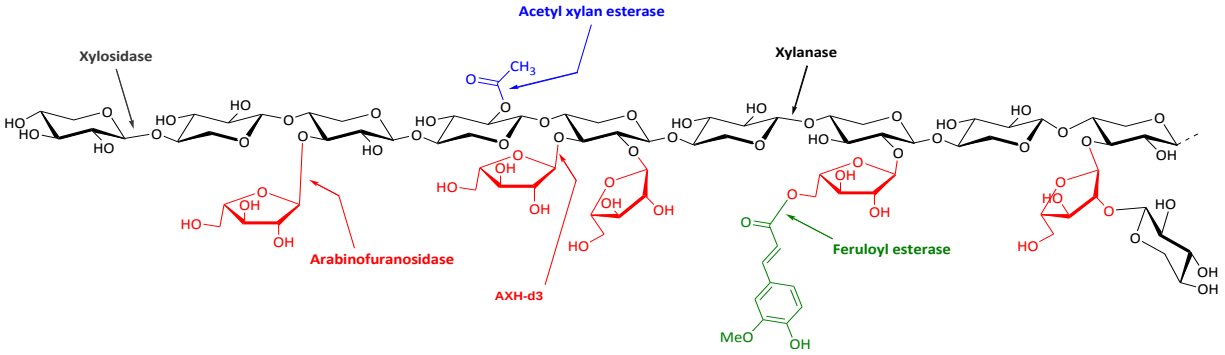


Fig. 1: Arabinoxylan structure ($XXA^{3}X^{2a}A^{2+3}XA^{5f2}XD^{2,3}$) and corresponding degrading enzymes. The latter can cleave L-arabinofuranosyl moieties of mono-substituted (GH 43, 51, 54 and 62) or di-substituted (GH43) D-xylosyl moieties as well hydrolyze linkages between D-xylosyl units of the backbone in *endo* (GH5, 8, 10 and 11) or *exo* (GH) manner. Esterases acting on xylan acetyl (CE1 to 7 and 12) or feruloyl (CE1) groups are also key accessory enzymes of AX decomposition.

The L-arabinose to D-xylose ratio varies from 0.2 to 0.9 depending on the plant origin and tissue. D-xyloses can as well be substituted less frequently on *O*-2 position by D-glucuronic acid, 4-*O*-methyl glucuronic acid or short oligomers consisting of L-arabinose, D-xylose, D-galactose, D-glucose or uronic acids. Substituents such as acetyl groups can also be found on *O*-2 and/or *O*-3 positions of D-xylosyl residues (Biely, 2012), which are thought to protect the polysaccharide from digestion by corresponding GHs and known to change its physicochemical properties. On the *O*-5 of L-arabinofuranosyl residues may be observed hydroxycinnamic acids, mainly ferulic acid, and to a lesser extent, dehydrodiferulic, *p*-coumaric or sinapic acids. These phenolic acids play a role in cross-linking between polymer chains, governing therefore the matrix reticulation. Besides, α-L-arabinofuranosyl moieties that substitute non-reducing ends D-xylosyl units (with (1,3) linkage) can be themselves by β-D-xylopyranosyl residues with (1,2) linkages, in AXs extracted from oat spelt, rice

husks, wheat straw, corn cobs and barley husks (Pastell et al., 2009). The presence of feruloyl groups on the *C*-5 position of L-Ara*f* residues is a unique structural feature of xylans from grasses. Besides, substitutions are thought to be not distributed randomly over the xylan chain but a pattern with up to 4 unsubstituted D-xyloses separated by 1 or 2 substituted moieties was proposed (Faik, 2010).

In order to describe these complex molecules, and facing a lack of homogeneity in writing standards, an efficient nomenclature system was created (Fauré et al. 2009, http://axonym.cermav.cnrs.fr). This rich-name nomenclature is a one-letter code system allowing an accurate description of diverse oligosaccharides derived from heteroxylans (e.g., X = unsubstituted xylose, A^2 or A^3 = D-xylose unit substituted by an α-L-arabinofuranose moiety with (1,2) or (1,3) linkage, respectively. Thereby, A^2X represents the trisaccharide α-L-Ara*f*-(1,2)-β-D-Xyl*p*-(1,4)-D-Xyl*p*).

Although the access to pure arabinoxylo-oligosaccharides (AXOS) is a tricky task, a significant quantity of NMR and mass spectroscopy data is available on a relatively large panel of AXOS, which were acquired from alkaline or enzymatically-treated (*endo*-xylanase action) hemicelluloses. Fully characterized AXOS with DP from 2 to 4 (D-xylosyl units) are listed hereinafter (Table 1). Xylan backbone with DP from 1 to 5, without any substitution were also described via NMR methods (Hoffmann et al., 1991) and a set of data on AXs with DP superior to 4 (from 5 to 7), mono and/or di-substituted is also available (R. A. Hoffmann et al., 1992; Kormelink et al., 1993).

Notwithstanding, are still missing in the literature the characterization of xylotriose mono-substituted with α-L-arabinofuranosyl on its central D-xylosyl unit with an (1,2) linkage (XA^2X) and on non-reducing (NR) end D-xylosyl unit with an (1,3) linkage (A^3XX). More complex AXOS such as $XA^{2+3}X$ and A^2A^3X are missing as well. However, α-L-arabinofuranosyl mono-substitutions with an (1,2) linkage on middle D-xylosyl units were recently described, although not on short XOS (DP 3) but on longer molecules, providing us hints on the chemical shift of the anomeric proton of the L-arabinofuranosyl unit which is around 5.28 ppm (Pitkänen et al., 2011; Anders et al., 2012; McKee et al., 2012).

Furthermore, for all aforementioned AXs, L-arabinofuranosyl substitutions occur on D-xylosyl (NR)-units of AXOS generated by xylanases action. Substitutions on reducing (Red) end D-xylosyl unit were recently described after digestion of AXs from rye by an arabinoxylanase from *Clostridium thermocellum* (*Ct*Xyl5A) (Correia et al., 2011). Indeed, *Ct*Xyl5A is an enzyme for which an L-arabinofuranosyl decoration of the D-xylosyl unit bound in -1 subsite would be a key specificity determinant, allowing therefore the generation of AXOS with reducing end D-xylose units harboring an L-arabinofuranosyl moiety (more than 99% of generated AXOS). Anomeric protons signals of L-arabinofuranosyl decorations (1→3 linkage) on D-xylosyl (Red)-unit were recorded at 5.34 and 5.39 ppm, for α and β anomers, respectively.

Table 1: AXOS motives characterized structures

Arabinoxylan motif	δ H-1 (A^x) (ppm) [a]	Reference
	DP ≤ 3	
A^2X	5.28	(Pastell et al., 2008)
A^3X	5.332-5.335[b]	(Gruppen et al., 1992; Vietor et al., 1994)
$A^{2+3}X$	[5.24/5.25]	(Pastell et al., 2008)
A^2XX	5.28 (5.26)[c]	(Vietor et al., 1994; Pastell et al., 2008)
A^3XX	nd	--
$A^{2+3}XX$	[5.238/5.246][d]	(Gruppen et al., 1992; Pastell et al., 2008; Vietor et al., 1994)
XA^2X	nd	--
XA^3X	5.401-5.396	(Gruppen et al., 1992; Vietor et al., 1994)
$XA^{2+3}X$	nd	--
A^2A^3X	nd	--
A^3A^3X	5.328/5.396-5.397	(Gruppen et al., 1992)
A^{2+3} A^3X	[5.244/5.244]/5.427-5.422	(Vietor et al., 1994)
$A^3A^{2+3}X$	5.331/[5.225/5.271]	(Gruppen et al., 1992)
	DP = 4	
XA^3XX	5.397	(Hoffmann et al., 1991)
XA^2XX	5.28[e]	(Ferré et al., 2000; McKee et al., 2012)
$XA^{2+3}XX$	[5.226/5.274]	(Hoffmann et al., 1991; Vietor et al., 1994)
XA^3A^3X	5.398/5.391-5.387	(Vietor et al., 1994)
$A^3A^{2+3}XX$	5.331/[5.225/5.271]	(Gruppen et al., 1992)
$A^{2+3}A^{2+3}XX$	[5.254/5.246]/[5.221/5.293]	(Gruppen et al., 1992)

[a] Chemical shift of the anomeric proton of the L-arabinofuranosyl substituent.
[b] The two different chemical shifts depend on the anomery of the D-xylose (Red) unit, α and β, respectively.
[c] The value between brackets indicates a value differing between given references.
[d] Chemical shifts are given for the anomeric proton of each L-arabinofuranosyl substituent from the (NR) to (Red) end of xylan backbone. When a D-xylosyl unit is di-substituted (put in square brackets) the substitution on *O*-2 position is given first followed by the one on *O*-3.
[e] Partial ^{1}H characterization
nd: no data available in the literature.

2. Arabinoxylans biosynthesis

Unlike cellulose microfibrils that are made at the plasma membrane to be directionally deposited directly in the cell wall (Somerville, 2006), hemicelluloses and pectic polysaccharides, forming the plant cell wall matrix in which the former microfibrils are embedded, are synthesized in the Golgi apparatus and then exported to the cell wall by exocytosis. Hydrolysis, transglycosylation and

hydrogen bond disruption or formation are the reactions, catalyzed by multi-enzyme complexes, occurring in the plant cell to explain the cell wall expansion and remodeling under turgor pressure (Sandhu et al., 2009). Although studies on plant cell wall formation understanding are relatively recent, several enzymatic activities involved in the plant cell wall biosynthesis were detected over the last decade. Regarding hemicellulose polysaccharide biosynthesis (Pauly et al., 2013), the genes encoding enzymes responsible for xyloglucan formation and remodeling, found in the primary cell wall, such as xyloglucan *endo*-transglycosylases/hydrolases (XETs/XEHs, belonging to the XTH "family") are well identified (Eklöf et al., 2010). On the other hand, studies on enzymes involved in AX biosynthesis remained really elusive till recently. Among revealed activities, a β-(1,4) xylosyltransferase from wheat was first described (Porchia et al., 2000) followed by a study on an arabinosyltransferase activity that was observed in Golgi membranes (Porchia et al., 2002), which highlighted its activity dependence on unsubstituted xylan chains that would act as acceptor substrates. However, no conclusive understanding on the biosynthetic mechanisms was withdrawn from this. The knowledge on the biosynthesis of plant cell wall hemicelluloses encountered a significant advance since 2007 thanks to the coupling of data obtained from comparative bioinformatics analyses (Mitchell et al., 2007) and several genetic studies carried out on *Arabidopsis,* revealing 3 key CAZy families GT43, 47 and 61. These new findings, reviewed in 2010 (Faik, 2010), demonstrated that GT43 are involved in xylan backbone biosynthesis (XylT), GT47 was assumed to be involved in D-glucuronic acid transfer (GlcAT), and to a lesser extent into L-arabinofuranose transfer (AraT), while GT61 function remained unknown. However, beyond genetic regulation understanding, clear-cut evidence of their biochemical function based on isolated and purified proteins is still required since current studies are always based on gene knock out followed by phenotype observation of the plant organism. Latterly, the activity of a purified complex of three GTs (GT43, 47 and 75) was assayed and revealed the three main activities required for GAX biosynthesis: XylT, AraT and GlcAT (Zeng et al., 2010). In the last few years was brought the proof of the biological function of a GT61 as a L-arabinofuranosyl transferase acting onto xylan in grasses and catalyzing the specific formation of α-L-(1,3) linkages. This was demonstrated by the deletion (GT61 gene knock out) or insertion (GT61 heterologous expression) of (1,3)-linked α-L-arabinofuranosyl in AXs of grasses (wheat, rice) or in *Arabidopsis* (naturally devoid of (1,3) substitutions), respectively (Anders et al., 2012). Similarly, the genes encoding for two arabinofuranosyl transferases from tomato were identified a few months ago as responsible for the arabinofuranosylation of xylose units in xyloglucan chains (Schultink et al., 2013).

Esterification of hemicelluloses

Hemicelluloses, and more specifically arabinoxylans, can be esterified via acetylation or feruloylation. The way feruloylation occurs is still a subject of debate with respect notably of cellular localization and substrate nature (Fig. 2), although many experiments would suggest that this process occurs mainly intra-protoplasmically rather than *in muro* (de O Buanafina, 2009). A feruloyl-transferase would use feruloyl-CoA as a donor substrate and transfer the feruloyl (Fe) moiety onto L-arabinofuranosyl units (Lenucci et al., 2009). However, there is no clear-cut information on the exact nature of the acceptor substrate *in planta*, *i.e.* does the transfer occurs directly on the AX polysaccharide chain or on shorter AXOS and these would be assembled later on? An *in vitro* assay revealed that an *O*-hydroxycinnamoyl transferase was able to catalyze this feruloylation reaction using an arabinoxylan trisaccharide (A^3X) as an acceptor (Yoshida-Shimokawa et al., 2001).

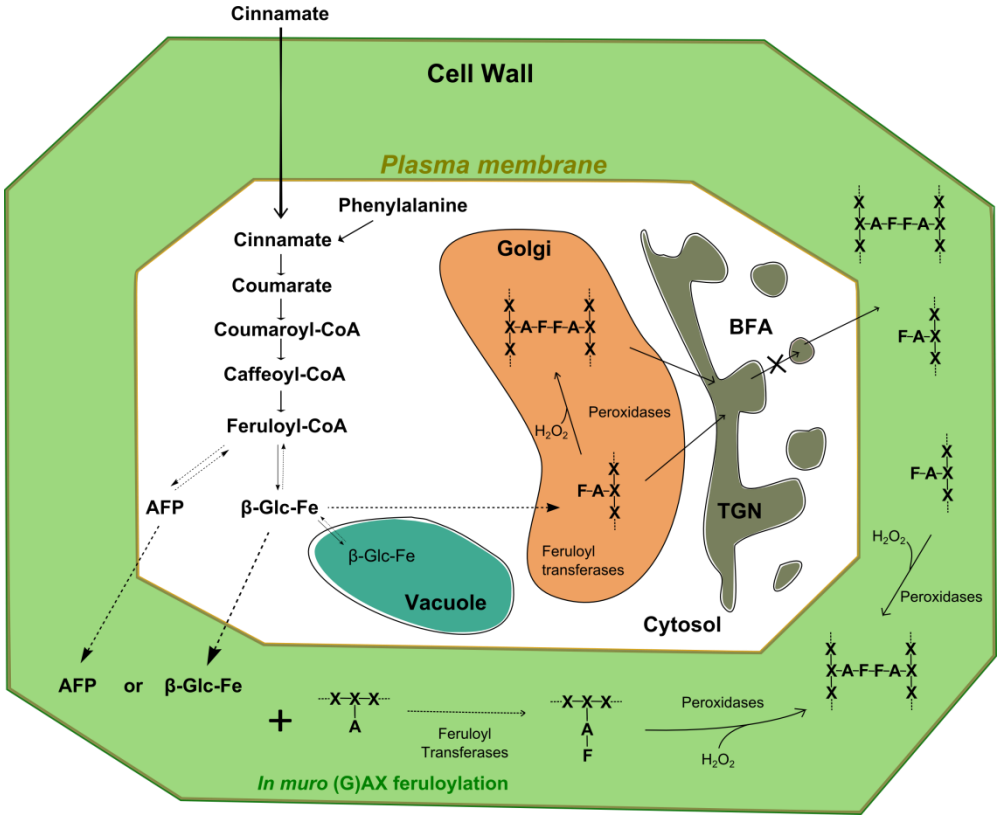


Fig. 2: Potential routes of (G)AX feruloylation and oxidative coupling. F, feruloyl group; TGN, trans-Golgi network; BFA, Brefeldin. A; AFP, Activated feruloyl precursor. Adapted from Lenucci et al., 2009.

Regarding the mode of activation of the donor substrate, it was also proposed that feruloyl-glucosyl ester could act as a donor. Alternatively, another hypothesis is that UDP-arabinosyl would be feruloylated leading to an intermediate serving as donor for a putative feruloyl arabinoside transferase that would use xylan backbone as an acceptor (de O Buanafina, 2009), providing therefore feruloylated AX. All these questions are good example of the importance of having

available well-defined AXOS to be able to study such phenomena.

Concerning the acetylation of polysaccharides, a protein system, analogous to the one found in fungi, bacteria and mammals, has been identified in plants. This system is made of *(i)* a multi-transmembrane domain, thought to serve as a translocation acetyl-carrier (across Golgi membrane) of acetyl-donor substrate (probably acetyl-CoA) (Gille et al., 2012), and *(ii)* a putative *O*-acetyltransferase catalyzing the esterification reaction. Besides, a series of experiments tend to show that acetylation would occur in the Golgi apparatus before secretion of polysaccharides toward the cell wall matrix but that this system is not the only option for non-cellulosic polysaccharide acetylation (presence of acetyl groups after genes deletion). Therefore, further studies are still necessary to ascertain the putative role of these proteins as well as the cellular localization of enzymatic events.

3. Arabinoxylans, many promises

Non digestible Oligosaccharides (NDOs), which are low molecular weight carbohydrates considered as dietary carbohydrates, became in the last few years "trendy" molecules since they present important physicochemical and physiological properties (Mussatto et al., 2007). Indeed, non-cariogenicity, low calorific value or even beneficial human microbiota stimulation are some of the properties credited to NDOs. For their diverse physico-chemical properties NDOs can also be used in non-food applications, such as drug delivery or even cosmetic stabilizer.

Arabinoxylo-oligosaccharides (AXOS), molecules we are interested in, can be considered as NDOs and their known properties are detailed hereinafter.

3.1. Arabinoxylans as therapeutics

3.1.1. Prebiotics

The notion of prebiotic, defined by Gibson and Roberfroid in 1995, is "a non-digestible food ingredient that beneficially affects the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon"[3].

AXOS can be considered as prebiotics since the human being lacks enzymes able to degrade them, which will therefore access the gastrointestinal tract as they have been ingested. Then, remained to show that among the complex microorganisms community only the beneficial for health bacteria (such as *Bifidobacteria* or *Lactobacilli*) were able to degrade these oligosaccharides, boosting in turn

[3] Regarding the recommended daily dose of prebiotics that should be taken, it depends on the initial community of benefic bacteria, the relationship between prebiotic effect and daily dose is not direct. Nevertheless, to discuss quantitatively the prebiotic effect a "prebiotic index" has been created and defined as "the increase in *bifidobacteria* expressed as the absolute number (N) of 'new' cfu/g of feces (E) divided by the daily dose (in grams) of prebiotic ingested (A)" (Roberfroid, 2007).

their growth (Aachary et al., 2011). This favored growth would be at the expense of putrefactive or pathogenic bacteria growth. It was demonstrated that AXOS allowed an increased selectivity during fermentation for *Bifidobacteria* compared to linear XOS structures (Van Laere et al., 2000) and, in 2004, an Australian study showed that AX-rich fiber diet could significantly improve glycaemic control in people suffering from Type II diabetes (Lu et al., 2000; Garcia et al., 2004). Notwithstanding, in 2007, were only considered as confirmed prebiotics: inulin, galacto-oligosaccharides (GOS) and fructo-oligosaccharides (FOS), since they were the best documented. Then, in 2011, AXOS were shown to meet the conditions to be classified as prebiotics (Broekaert et al., 2011). Indeed, it was demonstrated that an AX-rich diet could counteract high fat diet-induced gut dysbiosis via gut microbiota modulation, stimulating *Bacteroides-Prevotella* and *Roseburia* spp. (Neyrinck AM et al., 2011). Observed hypocholesterolemic, anti-inflammatory and anti-obesity effects attributed to AXs make these molecules promising nutrients for obesity prevention. Besides, a study on different *Bifidobacteria* strains showed that AXOS do present different prebiotic potentials according to the bacteria assayed and suggested also an important effect of the L-arabinofuranosyl substitution pattern (Pastell et al., 2009).

3.1.2. Anti-oxydative agent

In an earlier study performed *in vitro* on blood cells, MGN-3, which is an AXs extracted from enzymatically-treated rice bran and commercially known as Biobran (Daiwa-Pharm., Co., Ltd., Tokyo, Japan), revealed to have activity against HIV-1 (SF strain) (Mamdooh, 1998). MGN-3 would be involved in HIV-1 p24 antigen production inhibition, and therefore would block HIV replication. However, the mechanism by which MGN-3 is acting was not totally understood. A hypothesis was that this oligosaccharide would alter chemokine receptors (or even their production), which are surface glycoproteins used by HIV to gain entry into host cells (T lymphocytes, monocytes and macrophages). Besides, this work pointed out the fact that this plant-extracted drug did not cause side effects (at least after one month), which are one of the main problems of current anti-HIV treatments. Ten years later, the same team demonstrated that "MGN-3" AXs had oncolytic effect on solid Ehrlich carcinoma through oncostatic activity induction by modulating lipid peroxydation and increasing gluthatione content[4]. An observed enhancement of activity of the endogenous antioxidant scavenging enzymes, superoxide dismutase (SOD), glutathione peroxidase (GPx), catalase (CAT) and glutathione-S-transferase (GST) in blood, liver, and tumor tissue would be responsible for this (Noaman et al., 2008).

[4] One has to know that cancer growth is often related to oxy-radicals over-generation coupled to a loss of cellular redox balance, hydrogen peroxide produced contributes then to cancer cells mutation and subsequent proliferation.

3.1.3. Immunostimulating effects

A third property that may be credited to AXs or arabinan-containing polysaccharides is their immunostimulating effect. A study evaluated via immune-modulatory activity assays, *in vitro* and *in vivo*, the biological activity of arabinan-rich pectic polysaccharides, which are not AXs but contain mainly L-arabinosyl and to a lesser extent D-xylosyl moieties. It revealed that the crude extract stimulates expression of the lymphocyte activation markers CD69 and CD25 on the surface of B-cells and of CD69 on the surface of T-cells (Dourado et al., 2004). Such polysaccharides could therefore represent therapeutic opportunities in helping the immune system against pathogen threat. In this frame, the access to tailor-made compounds to be able to exactly determine the origin of such effects and the potential of L-arabinose and/or D-xylose-containing oligo- and polysaccharides is obviously necessary.

Along this line, two teams from France and Czech Republic worked on L-arabinofurano-oligosides chemo-enzymatic synthesis and realized *in vivo* assays with the resulting short furano-oligosides (tri, tetra and penta-L-Ara*f*). They showed that arabinofurano-oligosides could induce cytokine response in immunocompetent cells (murine macrophage cells) (Chlubnova et al., 2010). Besides, the response level depends on applied concentrations and also on the oligosaccharide length, the nature of glycosidic linkage having no significant impact in this study.

3.2. Arabinoxylans and biomaterials

Beyond physiological properties, AXs do also present physico-chemical properties which make these pentose-based oligosaccharides of utmost interest for bio-inspired material conception. Indeed, AXs display interesting features for food packaging since they possess good gas barrier properties. The (L-arabinose:D-xylose) ratio of AXs isolated from wheat bran revealed to have a significant impact on crystalline morphology, beta transitions and water sorption properties, rendering these macromolecules a “tunable film” according to desired applications (Saeed et al., 2011; Zhang et al., 2011). In the same line, it was demonstrated that the degree and pattern of substitution influences the interactions with cellulosic surfaces, leading to different xylan-cellulosic-based materials (Köhnke et al., 2011). Besides, the presence of acetyl and feruloyl groups were shown to have an influence on functional characteristics such as viscosity, foam stabilization and gelation capacity of AXs (Niño-Medina et al., 2009).

3.3. Arabinoxylans as a tool for enzyme and plant cell wall study

Tailor-made AXs, as a pure product of research, could be helpful for new researches. First, it would provide reference compounds for hemicellulases catalytic mechanism study. Second, well-defined AXs motifs would be valuable probes for immunochemical-based analysis of plant cell wall (Rémond et al., 2002). The understanding of plant cell wall biosynthesis is obviously really exciting on a fundamental point of view, and this knowledge could also be used by Green and White biotechnologies to engineer plants for the production of valuable products (Bevan et al., 2006). Indeed, the mastering of plant cell wall biosynthesis would allow getting plants with a controlled composition. For instance, AXs without acetyl substitutions would be of great interest for biomass-based biofuel industry since these molecules are *(i)* often fermentation inhibitors and *(ii)* involved in the resistance of plant cell walls to GH-mediated degradation. Another example could be to play on the xylan backbone substitution pattern to modify AXs solubility and therefore on their digestibility (Sandhu et al., 2009).

References

Aachary, A.A. & Prapulla, S.G. (2011) Xylooligosaccharides (XOS) as an emerging prebiotic: microbial synthesis, utilization, structural characterization, bioactive properties, and applications. *Compr. Rev. Food Sci. Food Saf.*, 10 (1), pp.2–16.

Anders, N., Wilkinson, M.D., Lovegrove, A., Freeman, J., Tryfona, T., Pellny, T.K., Weimar, T., Mortimer, J.C., Stott, K., Baker, J.M., Defoin-Platel, M., Shewry, P.R., Dupree, P. & Mitchell, R.A.C. (2012) Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. *Proc. Natl. Acad. Sci. U. S. A.*, 109 (3), pp.989–993.

Bevan, M.W. & Franssen, M.C.R. (2006) Investing in green and white biotech. *Nat. Biotech.*, 24 (7), pp.765–767.

Biely, P. (2012) Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnol. Adv.*, 30 (6), pp.1575–1588.

Broekaert, W.F., Courtin, C.M., Verbeke, K., Van de Wiele, T., Verstraete, W. & Delcour, J.A. (2011) Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides. *Crit. Rev. Food Sci. Nutr.*, 51 (2), pp.178–194.

Chlubnova, I., Filipp, D., Spiwok, V., Dvorakova, H., Daniellou, R., Nugier-Chauvin, C., Kralova, B. & Ferrieres, V. (2010) Enzymatic synthesis of oligo-D-galactofuranosides and L-arabinofuranosides: from molecular dynamics to immunological assays. *Org. Biomol. Chem.*, 8 (9), pp.2092–2102.

Correia, M.A.S., Mazumder, K., Brás, J.L.A., Firbank, S.J., Zhu, Y., Lewis, R.J., York, W.S., Fontes, C.M.G.A. & Gilbert, H.J. (2011) Structure and function of an arabinoxylan-specific xylanase. *J. Biol. Chem.*, 286 (25), pp.22510–22520.

Dourado, F., Madureira, P., Carvalho, V., Coelho, R., Coimbra, M.A., Vilanova, M., Mota, M. & Gama, F.M. (2004) Purification, structure and immunobiological activity of an arabinan-rich pectic polysaccharide from the cell walls of Prunus dulcis seeds. *Carbohydr. Res.*, 339 (15), pp.2555–2566.

Eklöf, J.M. & Brumer, H. (2010) The XTH gene family: an update on enzyme structure, function, and phylogeny in xyloglucan remodeling. *Plant Physiol.*, 153 (2), pp.456–466.

Faik, A. (2010) Xylan biosynthesis: news from the grass. *Plant Physiol.*, 153 (2), pp.396–402.

Fauré, R., Courtin, C.M., Delcour, J.A., Dumon, C., Faulds, C.B., Fincher, G.B., Fort, S., Fry, S.C., Halila, S., Kabel, M.A., Pouvreau, L., Quemener, B., Rivet, A., Saulnier, L., Schols, H.A., Driguez, H. & O'Donohue, M.J. (2009) A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls. *Aust. J. Chem.*, 62 (6), pp.533–537.

Ferré, H., Broberg, A., Duus, J.O. & Thomsen, K.K. (2000) A novel type of arabinoxylan arabinofuranohydrolase isolated from germinated barley analysis of substrate preference and specificity by nano-probe NMR. *Eur. J. Biochem.*, 267 (22), pp.6633–6641.

Garcia, A. L. Steiniger, J. Reich, S. C. Weickert, M. O. Harsch, I. Machowetz, A. Mohlig, M. Spranger, J. Rudovich, N. N. Meuser, F. Doerfer, J. Katz, N. Speth, M. Zunft, H. J. F. Pfeiffer, A. H. F. Koebnick, C. (2004) Arabinoxylan fibre consumption improved glucose metabolism, but did not affect serum adipokines in subjects with impaired glucose tolerance. *Horm. Metab. Res.*, 38 (11), pp.761–766.

Gille, S. & Pauly, M. (2012) O-acetylation of plant cell wall polysaccharides. *Frontiers in plant science*, 3 (12), pp.1–7.

Gruppen, H., Hoffmann, R.A., Kormelink, F.J., Voragen, A.G., Kamerling, J.P. & Vliegenthart, J.F. (1992) Characterisation by ^{1}H NMR spectroscopy of enzymically derived oligosaccharides from alkali-extractable wheat-flour arabinoxylan. *Carbohydr. Res.*, 233, pp.45–64.

Hoffman, R.A., Leeflang, B.R., de Barse, M.M.J., Kamerling, J.P. & Vliegenthart, J.F.G. (1991) Characterisation by H-NMR spectroscopy of oligosaccharides, derived from arabinoxylans of white endosperm of wheat, that contain the elements →4)[α-L-Ara*f*-(1→3)]-β-D-Xyl*p*-(1→ or →4)[α-L-Ara*f*-(1→2)][α-L-Ara*f*-(1→3)]-β-D-Xyl*p*-(1→. *Carbohydr. Res.*, 221, pp.63–81.

Hoffmann, R.A., Geijtenbeek, T., Kamerling, J.P. & Vliegenthart, J.F.G. (1992) arabinoxylan oligosaccharides : structures of hepta- to tetradeca-saccharides containing two or three branched xylose residues. *Carbohydr. Res.*, 223, pp.19–44.

Köhnke, T., Ostlund, A. & Brelid, H. (2011) Adsorption of arabinoxylan on cellulosic surfaces: influence of degree of substitution and substitution pattern on adsorption characteristics. *Biomacromolecules*, 12 (7), pp.2633–2641.

Kormelink, F.J., Hoffmann, R.A., Gruppen, H., Voragen, a G., Kamerling, J.P. & Vliegenthart, J.F. (1993) Characterisation by ^{1}H NMR spectroscopy of oligosaccharides derived from alkali-extractable wheat-flour arabinoxylan by digestion with endo-(1-->4)-β-D-xylanase III from *Aspergillus awamori*. *Carbohydr. Res.*, 249 (2), pp.369–382.

Van Laere, K.M.J., Hartemink, R., Bosveld, M., Schols, H.A. & Voragen, A.G.J. (2000) Fermentation of plant cell wall derived polysaccharides and their corresponding oligosaccharides by intestinal bacteria. *J. Agric. Food Chem.*, 48 (5), pp.1644–1652.

Lenucci, M.S., Piro, G. & Dalessandro, G. (2009) In muro feruloylation and oxidative coupling in monocots. *Plant Signal. Behav.*, 4 (3), pp.228–230.

Lu, ZX Walker, KZ Muir, JG O'Dea, K. (2004) Arabinoxylan fibre improves metabolic control in people with Type II diabetes. *Eur. J. Clin. Nutr.*, 58 (4), pp.621–628.

Mamdooh, G. (1998) Anti-HIV activity *in vitro* of MGN-3, an activated arabinoxylan from Rice Bran. *Biochem. Biophys. Res. Commun.*, 243 (1), pp.25–29.

McKee, L.S., Peña, M.J., Rogowski, A., Jackson, A., Lewis, R.J., York, W.S., Krogh, K.B.R.M., Viksø-Nielsen, A., Skjøt, M., Gilbert, H.J. & Marles-Wright, J. (2012) Introducing *endo*-xylanase activity into an *exo*-acting arabinofuranosidase that targets side chains. *Proc. Natl. Acad. Sci. U. S. A.*, 109 (17), pp.6537–6542.

Mitchell, R.A.C., Dupree, P. & Shewry, P.R. (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiol.*, 144 (1), pp.43–53.

Mussatto, S.I. & Mancilha, I.M. (2007) Non-digestible oligosaccharides: A review. *Carbohydr. Polym.*, 68 (3), pp.587–597.

Neyrinck, A.M., Possemiers, S., Druart, C., Van de Wiele, T., De Backer, F., Cani, P.D., Larondelle, Y. & Delzenne, N.M. (2011) Prebiotic effects of wheat arabinoxylan related to the increase in *bifidobacteria*, *roseburia* and *bacteroides/prevotella* in diet-induced obese mice. *PLoS One*, 6 (6), p.e20944.

Niño-Medina, G., Carvajal-Millán, E., Rascon-Chu, A., Marquez-Escalante, J.A., Guerrero, V. & Salas-Muñoz, E. (2009) Feruloylated arabinoxylans and arabinoxylan gels: structure, sources and applications. *Phytochem. Rev.*, 9 (1), pp.111–120.

Noaman, E., Badr El-Din, N.K., Bibars, M.A., Abou Mossallam, A.A. & Ghoneum, M. (2008) Antioxidant potential by arabinoxylan rice bran, MGN-3/biobran, represents a mechanism for its oncostatic effect against murine solid Ehrlich carcinoma. *Cancer Lett.*, 268 (2), pp.348–359.

De O Buanafina, M.M. (2009) Feruloylation in grasses: current and future perspectives. *Molecular plant*, 2 (5), pp.861–872.

Pastell, H., Tuomainen, P., Virkki, L. & Tenkanen, M. (2008) Step-wise enzymatic preparation and structural characterization of singly and doubly substituted arabinoxylo-oligosaccharides with non-reducing end terminal branches. *Carbohydr. Res.*, 343 (18), pp.3049–3057.

Pastell, H., Westermann, P., Meyer, A.S., Tuomainen, P. & Tenkanen, M. (2009) *In vitro* fermentation of arabinoxylan-derived carbohydrates by bifidobacteria and mixed fecal microbiota. *J. Agric. Food Chem.*, 57 (18), pp.8598–8606.

Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A. & Xiong, G. (2013) Hemicellulose biosynthesis. *Planta*, 238 (4), pp.627–642.

Pitkänen, L., Tuomainen, P., Virkki, L. & Tenkanen, M. (2011) Molecular characterization and solution properties of enzymatically tailored arabinoxylans. *Int. J. Biol. Macromol.*, 49 (5), pp.963–969.

Porchia, A.C. & Scheller, H.V. (2000) Arabinoxylan biosynthesis: Identification and partial characterization of β-1,4-xylosyltransferase from wheat. *Physiol. Plant.*, 110 (3), pp.350–356.

Porchia, A.C., Sørensen, S.O. & Scheller, H.V. (2002) Arabinoxylan biosynthesis in wheat. Characterization of arabinosyltransferase activity in Golgi membranes. *Plant Physiol.*, 130 (1), pp.432–441.

Rémond, C., Ferchichi, M., Aubry, N., Plantier-Royon, R., Portella, C. & O'Donohue, M.J. (2002) Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α-L-arabinofuranosidase. *Tetrahedron Lett.*, 43 (52), pp.9653–9655.

Saeed, F., Pasha, I., Anjum, F.M. & Sultan, M.T. (2011) Arabinoxylans and arabinogalactans: a comprehensive treatise. *Crit. Rev. Food Sci. Nutr.*, 51 (5), pp.467–476.

Sandhu, A.P.S., Randhawa, G.S. & Dhugga, K.S. (2009) Plant cell wall matrix polysaccharide biosynthesis. *Mol. plant*, 2 (5), pp.840–850.

Scheller, H.V. & Ulvskov, P. (2010) Hemicelluloses. *Annu. Rev. Plant Biol.*, 61, pp.263–289.

Schultink, A., Cheng, K., Park, Y.B., Cosgrove, D.J. & Pauly, M. (2013) The identification of two arabinosyltransferases from tomato reveals functional equivalency of xyloglucan side chain substituents. *Plant physiology*, 163 (1), pp.86–94.

Somerville, C. (2006) Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.*, 22, pp.53–78.

Vietor, R.J., Hoffmann, R.A., Angelino, S.A.G.., Voragen, A.G.J., Kamerling, J.P. & Vliegenthart, J.F.G. (1994) Structures of small oligomers liberated from barley arabinoxylans by endoxylanase from *Aspergillus awamori*. *Carbohydr. Res.*, 6215 (93), pp.245–255.

Yoshida-Shimokawa, T., Yoshida, S., Kakegawa, K. & Ishii, T. (2001) Enzymic feruloylation of arabinoxylan-trisaccharide by feruloyl-CoA:arabinoxylan-trisaccharide O -hydroxycinnamoyl transferase from *Oryza sativa*. *Planta*, 212 (3), pp.470–474.

Zeng, W., Jiang, N., Nadella, R., Killen, T.L., Nadella, V. & Faik, A. (2010) A glucurono(arabino)xylan synthase complex from wheat contains members of the GT43, GT47, and GT75 families and functions cooperatively. *Plant Physiol.*, 154 (1), pp.78–97.

Zhang, Y., Pitkänen, L., Douglade, J., Tenkanen, M., Remond, C. & Joly, C. (2011) Wheat bran arabinoxylans: chemical structure and film properties of three isolated fractions. *Carbohydr. Polym.*, 86 (2), pp.852–859.

II-3. Pentose-containing compounds synthesis

Pentoses can be subdivided into either aldopentoses (xylose, arabinose, ribose and lyxose) or ketopentoses (ribulose and xylulose). The occurrence of L-arabinose and D-xylose in hemicelluloses of plant cell walls makes them the most abundant pentoses on earth. However, though scientists have been interested in their transformation via microorganisms metabolism for more than one century (Hawkins, 1915), their use for the synthesis of new molecules or their conversion into enzymatic pathways is still lagging far behind hexoses exploitation. This chapter aims to list all available D-xylose/L-arabinose-based oligosaccharides, accessible via chemical or chemo-enzymatic synthesis routes.

1. Exclusive chemical synthesis

Exclusive chemical syntheses of D-xylose/L-arabinose-based oligosaccharides are really scarce and the few ones reported in the literature are listed hereinafter (Table 1). The reader can observe that the more complex oligosaccharide obtained by such means is the trisaccharide A^3X.

Table 1: D-xylose-or L-arabinose-based compounds obtained via chemical synthesis

Compounds	References
Methyl α-L-arabinofuranosyl-(1→3)-β-D-xylopyranosyl-(1→4)-D-xylopyranoside and methyl β-D-xylopyranosyl-(1→4)-[α-L-arabinofuranosyl-(1→3)]-D-xylopyranoside	(Hirsch et al., 1984)
α-L-arabinofuranosyl-(1→3)-β-D-xylopyranosyl-(1→4)-D-xylopyranose = A^3X	(Koto et al., 1985)
5-*O*-*trans*-feruloyl-α-L-arabinofuranoside	(Hatfield et al., 1991)
Benzyl 2,3,5-tri-*O*-benzoyl-α-L-arabinofuranosyl-(1→3)-2,4-di-*O*-benzoyl-α-D-xylopyranoside	(Helm et al., 1991)
Methyl α-L-arabinofuranosyl-(1→2)-α-L-arabinofuranoside, methyl α-L-arabinofuranosyl-(1→3)-α-L-arabinofuranoside, and methyl α-L-arabinofuranosyl-(1→5)-α-L-arabinofuranoside	(Kawabata et al., 1995)
Allyl 2,3,5-tri-*O*-benzoyl-α-L-arabinofuranosyl-α-D-xylopyranosides and β-D-xylopyranosides	(Utille et al., 2007)

Of note, are only referenced in Table 1 molecules involving glycosidic bond formation between two pentoses via a chemical way, excluding therefore alkyl or nucleotide pentoses synthesis that are beyond the scope of this review. As a remark, many studies already reported the use of D-xylose, L-arabinose for biodegradable non-ionic surfactants synthesis from xylan material, by telomerization of

diverse alkyl chains, such as butadiene or butyl, octyl and decyl chains (Bouxin et al., 2010; Ochs et al., 2011). Relative to glucose-based surfactants, pentose-based surfactants present much better oil solubilization properties and increase in degreasing and wetting (Bouquillon, 2011).

2. Chemo-enzymatic synthesis

2.1. Glycosynthases and pentoses

A way to circumvent difficulties encountered with the chemical synthesis of pentose/furanoses-based oligosaccharides is to consider the glycosynthase strategy previously applied to hexose-active enzymes. To achieve this, several retaining *endo*-1,4-β-xylanases from GH10 were successfully converted to the corresponding glycosynthases to afford oligosaccharides ranging from tetra- to several dozens of D-xylose units (Table 2). To circumvent the issue of expensive α-X_2-F donor substrate, a β-xylosidase (GH52) from *Geobacillus stearothermophilus* was also converted into a glycosynthase (XynB2-E335G) catalyzing the transfer of α-D-xylopyranosyl fluoride (α-X-F) onto various aryl acceptor sugars, the best yields being obtained for autocondensation (89%), *p*NP-Glc*p* (49%) and *p*NP-Xyl*p* (42%) (Ben-David et al., 2007). In the latter study, the GH52 XynB2-E335G was coupled to a glycosynthase variant of a GH10 xylanase from *G. stearothermophilus* XT6 (XT6-E265G) yielding *p*NP-XOS with DP ranging from 6 up to 100 xylose units directly from the more easily accessible α-X-F donor substrate. Besides, unfortunately, regarding the L-arabinofuranosyl transfer reaction, the glycosynthase strategy was hitherto never successfully applied.

Table 2: GH10 xylanases-derived glycosynthases

Enzyme	GH Family	Organism	Donor	Acceptor	Products	k_{cat}/K_M $(s.mM)^{-1}$	Ref.	Remark
CFXcd-E235G	10	*Cellulomonas fimi*	αX_2F	*p*NP-X_2	DP = 4-12	15	(Kim et al., 2006)	
XylB - E259G	10	*Thermotoga maritima*	αX_2F	X_2 or X_3	DP>5	0.0295 (F^-)	(Sugimura et al., 2006)	X_2 as acceptor prohibits polym
XynB - E293G	10	*Clostridium stercorarium*	αX_2F	X_3	DP : up to 22	0.026 (F^-)		X_2 not well recognized
XynA - E301G	10	*Bacillus halodurans*	αX_2F	X_3	DP>5	0.0096 (F^-)		X_2 not well recognized
Cex (CfXyn10, previously endo-glycanase) : E233G	10	*Cellulomonas fimi*	αX_2F	X_2 or X_3	DP>5	0.035 (F^-)		X_2 as acceptor prohibits polym
XT6 B2 E265G	10	*Geob. Stear.*	αX_2F	αX_2F	DP= 6 - 100	-	(Ben-David et al., 2007)	-

2.2. GH-catalyzed synthesis of furanose/pentose-based compounds

2.2.1. D-Xylose-based molecules

A wide range of acceptors molecules has already been used in transglycosylation reaction involving D-xylosyl moieties as donor. To our knowledge, D-xylosyl unit can be transferred onto other carbohydrates such as cellobioside, mannosides or psicose (Table 3) or on alkyl chains (Table 4).

Xylo-oligosaccharides (XOS)

Table 3: XOS obtained via chemo-enzymatic synthesis

Product	Enzyme	Reference
Xylosyl-cellobioside	*endo*-(1→4)-β-D-xylanase from the yeast *Cryptococcus aldidus*	(Biely et al., 1983)
Xylosyl-mannosides and D-Xyl*p*-(1,1')-D-Xyl*p*	β-xylosidase from *Aspergillus niger*	(Kizawa et al., 1991)
XOS (DP 2-7) with *p*NP-β-**D**-Xyl*p* as donor	β-D-xylosidase from *Aspergillus oryzae*	(Eneyskaya et al., 2003)
MU-β-D-XOS	β-D-xylosidase from *Aspergillus sp.*	(Eneyskaya et al., 2005)
Xylosyl-psicose (XPs)	*endo*-(1→4)-β-D-xylanase from *Aspergillus sojae*	(Oshima et al., 2006)
XOS with large variety of carbohydrate acceptors	two β-xylosidase (GH3), *Bxl*A and *Bxl*B from *Aspergillus nidulans*	(Dilokpimol et al., 2011)

Alkyl D-xylosides

Table 4: Alkyl D-xylosides chemo-enzymatic synthesis

Product	Enzyme	Reference
Alkyl β-D-xylobioside	β-xylosidase from *Trichoderma reesei*	(Drouet, P Zhang, M Legoy, 1994)
Octyl β-D-xylobioside	xylanase of *Aureaobasidium pullulans*	(Matsumura S., 1999)
β-XOS with different aglycones: *n*-hexanol, 9-fluorenemethanol, 1,4-butanediol	xylosidase from *Thermotoga neapolitana*	(Tramice et al., 2007)
Polyphenyl β-oligoxyloside (on divalent and trivalent phenolic hydroxyl group)	*endo*-xylanase (GH11) from *Bacillus* KT-12	(Chiku et al., 2008)
Alkyl D-xyloside (with *p*NP-Xyl*p*, X_2 and X_3)	β-D-xylosidase from *Bacillus halodurans*	(Muzard et al., 2009)
n-octyl X_2 and X_3	xylanase from *Thermotoga neapolitana*	(Mamo et al., 2010)
Pentyl and octyl D-xylosides from XOS	xylanases from *Thermobacillus xylanilyticus* or Novozymes (NS-50030)	(Ochs et al., 2011)

2.2.2. Furanose-based molecules

The number of L-arabinofuranose-containing molecules obtained through chemo-enzymatic synthesis is even lower than D-xylose-based ones (Table 5). Up to date, L-arabinofuranose can be transferred onto several alcohols, on *para*-nitrophenyl α-L-arabinofuranoside (autocondensation of *p*NP-α-L-Ara*f*), and on the synthetic acceptor benzyl α-D-xylopyranoside (Bn-α-D-Xyl*p*).

Table 5: Arabinose-based compounds obtained via chemo-enzymatic synthesis

Donor	Acceptor	Linkage	Enzyme	Reference
*p*NP-α-L-Ara*f*	MeOH, EtOH, *n*-propanol, *n*-butanol, allyl alcohol	α-L	*Tx*Abf*	(Rémond et al., 2002)
*p*NP-α-L-Ara*f*	Bn-α-D-Xyl*p*	α-L-(1,2)	*Tx*Abf	(Rémond et al., 2004)
arabinobioside	Glycerol	α-L-(1,1)	*Exo*-1,5-α-L-arabinanase from *Penicillium chrysogenum*	(Sakamoto et al., 2004)
*p*NP-α-L-Ara*f*	MeOH	α-L	*Tk*Abf*	(Wan C. et al., 2007)
*p*NP-α-L-Ara*f*	*p*NP-α-L-Ara*f*	α-L-(1,2) or α-L-(1,3) or α-L-(1,5)	*Ct*Abf *	(Chlubnova et al., 2010)

* α-L-arabinofuranosidases from: *Thermobacillus xylanilyticus (Tx*Abf*)*, *Trichoderma koningii* (*Tk*Abf), *Clostridium thermocellum* (*Ct*Abf)

Research activities on the transfer of furanose sugars are also developed by another group of scientists, more specifically on D-galactofuranoses (D-Gal*f*) using the GH51 α-L-arabinofuranosidase from *Clostridium thermocellum* (*Ct*Abf). Chemo-enzymatic syntheses were set up to prepare notably β-D-Gal*f*-(1,6)-β-D-Gal*f*, a biologically relevant motif, as well as di- to penta-galactofuranosides (1→2 or 1→3 or 1→6 linkages) (Chlubnova et al., 2010). Using *Ct*Abf as a scaffold protein the first furanothioglycoligase was created (acid/base mutant E173A, coupled to the use of a thioimidoyl donor and thio-acceptor) allowing the transfer of L-arabinofuranosyl moieties onto aryl groups with a *S*-linkage, although remaining unsuccessful onto sugar acceptors (Almendros et al., 2011). This new enzymatic tool was then employed for the conception of an L-arabinofuranosyl cluster (erythritol dendrimer) (Almendros et al., 2013). Very recently, the acceptor plasticity of *Ct*Abf was also investigated showing that D-Gal*f* moieties can be transferred onto a variety of pyranosidic sugars (*p*NP-D-glucopyranose, D-galactopyranose or D-mannopyranose with α or β-anomery for both, and *p*NP-α-L-rhamnopyranose), leading to mixtures of 1→2, 1→3, 1→4 and1→6 regioisomers in different proportions depending on the acceptor (Chlubnová et al., 2014).

References

Almendros, M., Danalev, D., François-Heude, M., Loyer, P., Legentil, L., Nugier-Chauvin, C., Daniellou, R. & Ferrières, V. (2011) Exploring the synthetic potency of the first furanothioglycoligase through original remote activation. *Org. Biomol. Chem.*, 9 (24), pp.8371–8378.

Almendros, M., François-heude, M. & Legentil, L. (2013) Chemo-enzymatic synthesis of an original arabinofuranosyl cluster : optimization of the enzymatic conditions. *Arkivoc*, 2013 (2), pp.123–132.

Ben-David, A., Bravman, T., Balazs, Y.S., Czjzek, M., Schomburg, D., Shoham, G. & Shoham, Y. (2007) Glycosynthase activity of *Geobacillus stearothermophilus* GH52 β-xylosidase: efficient synthesis of xylooligosaccharides from α-D-xylopyranosyl fluoride through a conjugated reaction. *ChemBioChem*, 8 (17), pp.2145–2151.

Biely, P. & Vršanská, M. (1983) Xylosyl transfer to cellobiose catalysed by an *endo*-(1→4)-β-D-xylanase of *Cryptococcus albidus*. *Carbohydr. Res.*, 123 (1), pp.97–107.

Bouquillon, S. (2011) D-Xylose and L-arabinose-based surfactants: synthesis, reactivity and physico-chemical properties. *Comptes Rendus Chimie*, 14 (7-8), pp.716–725.

Bouxin, F., Marinkovic, S., Le Bras, J. & Estrine, B. (2010) Direct conversion of xylan into alkyl pentosides. *Carbohydr. Res.*, 345, pp.2469–2473.

Chiku, K., Uzawa, J., Seki, H., Amachi, S., Fujii, T. & Shinoyama, H. (2008) Characterization of a novel polyphenol-specific oligoxyloside transfer reaction by a family 11 xylanase from *Bacillus* sp. KT12. *Biosci., Biotechnol. Biochem.*, 72 (9), pp.2285–2293.

Chlubnova, I., Filipp, D., Spiwok, V., Dvorakova, H., Daniellou, R., Nugier-Chauvin, C., Kralova, B. & Ferrieres, V. (2010) Enzymatic synthesis of oligo-D-galactofuranosides and L-arabinofuranosides: from molecular dynamics to immunological assays. *Org. Biomol. Chem.*, 8 (9), pp.2092–2102.

Chlubnová, I., Králová, B., Dvořáková, H., Hošek, P., Spiwok, V., Filipp, D., Nugier-Chauvin, C., Daniellou, R. & Ferrières, V. (2014) The versatile enzyme Ara*f*51 allowed efficient synthesis of rare pathogen-related β-D-galactofuranosyl-pyranoside disaccharides. *Org. Biomol. Chem.*, 12 (19), pp.3080–3089.

Dilokpimol, A., Nakai, H., Gotfredsen, C.H., Appeldoorn, M., Baumann, M.J., Nakai, N., Schols, H.A., Hachem, M.A. & Svensson, B. (2011) Enzymatic synthesis of β-xylosyl-oligosaccharides by transxylosylation using two β-xylosidases of glycoside hydrolase family 3 from *Aspergillus nidulans* FGSC A4. *Carbohydr. Res.*, 346 (3), pp.421–429.

Drouet, P Zhang, M Legoy, M. (1994) Enzymatic-synthesis of alkyl β-D-xylosides by transxylosylation and reverse hydrolysis. *Biotechnol. Bioeng.*, 43 (11), pp.1075–1080.

Eneyskaya, E. V, Ivanen, D.R., Shabalin, K.A., Kulminskaya, A.A., Backinowsky, L. V, Brumer III, H. & Neustroev, K.N. (2005) Chemo-enzymatic synthesis of 4-methylumbelliferyl β-(1->4)-D-xylooligosides: new substrates for β-D-xylanase assays. *Org. Biomol. Chem.*, 3 (1), pp.146–151.

Eneyskaya, E. V., Brumer, H., Backinowski, L. V., Ivanen, D.R., Kulminskaya, A.A., Shabalin, K.A. & Neusstroev, K.N. (2003) Enzymatic synthesis of β-xylanase substrates: transglycosylation reactions of the β-xylosidase from *Aspergillus* sp. *Carbohydr. Res.*, 338 (4), pp.313–325.

Hatfield, R., Helm, R. & Ralph, J. (1991) Synthesis of methyl 5-O-feruloyl-α-L-arabinofuranoside and its use as a substrate to assess feruloyl esterase activity. *Anal. Biochem.*, 194 (1), pp.25–33.

Hawkins, L.A. (1915) The utilization of certain pentoses and compounds of pentoses by *Glomerella cingulata*. *Am. J. Bot.*, 2 (8), pp.375–388.

Helm, R.F., Ralphlb, J., Drive, L. & Anderson, L. (1991) Regioselective protection strategies. *J. Org. Chem*, 56 (9), pp.7015–7021.

Hirsch, J. & Petrakova, E. (1984) Stereoselective synthesis and ^{13}C-NMR spectra of two isomeric methyl 3-glycoside of trisaccharides related To arabinoxylan methyl bromide under modified Koenigs-Knorr conditions. *Science*, 131, pp.219–226.

Kawabata, Y., Kaneko, S. & Kusakabe, I. (1995) Synthesis of regioisomeric methyl α-L-arabinofuranobiosides. *Carbohydr. Res.*, 267, pp.39–47.

Kim, Y.-W., Fox, D.T., Hekmat, O., Kantner, T., McIntosh, L.P., Warren, R.A.J. & Withers, S.G. (2006) Glycosynthase-based synthesis of xylo-oligosaccharides using an engineered retaining xylanase from *Cellulomonas fimi*. *Org. Biomol. Chem.*, 4 (10), pp.2025–2032.

Kizawa, H., Shinoyama, H. & Yasui, T. (1991) The synthesis of new xylosyloligosaccharides by transxylosylation with *Aspergillus niger* β-xylosidase. *Agric. Biol. Chem.*, 55 (3), pp.671–678.

Koto, S., Morishima, N., Takenaka, K., Uchida, C. & Zen, S. (1985) Pentoside synthesis by dehydrative glycosylation. Synthesis of *O*-α-L-arabinofuranosyl-(1→3)-O-β-D-xylopyranosyl-(1→4)-D-xylopyranose. *Bull. Chem. Soc. Jpn.*, 58 (5), pp.1464–1468.

Mamo, G., Kasture, S., Faryar, R., Hashim, S. & Hatti-Kaul, R. (2010) Surfactants from xylan: Production of n-octyl xylosides using a highly thermostable xylanase from *Thermotoga neapolitana*. *Process Biochem.*, 45, pp.700–705.

Matsumura S., K.S.& K.T. (1999) Preparation of octyl β-D-xylobioside and xyloside by xylanase-catalyzed direct transglycosylation reaction of xylan and octanol. *Biotechnol. Lett.*, 21 (1), pp.17–22.

Muzard, M., Aubry, N., Plantier-Royon, R., O'Donohue, M. & Rémond, C. (2009) Evaluation of the transglycosylation

activities of a GH 39 β-D-xylosidase for the synthesis of xylose-based glycosides. *J. Mol. Catal. B: Enzym.*, 58 (1-4), pp.1–5.

Ochs, M., Muzard, M., Plantier-Royon, R., Estrine, B. & Remond, C. (2011) Enzymatic synthesis of alkyl β-D-xylosides and oligoxylosides from xylans and from hydrothermally pretreated wheat bran. *Green Chem.*, 13 (9), pp.2380–2388.

Oshima, H., Kimura, I. & Izumori, K. (2006) Synthesis and structure analysis of novel disaccharides containing D-psicose produced by endo-1,4-β-D-xylanase from *Aspergillus sojae*. *J. Biosci. Bioeng.*, 101 (3), pp.280–283.

Rémond, C., Ferchichi, M., Aubry, N., Plantier-Royon, R., Portella, C. & O'Donohue, M.J. (2002) Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α-L-arabinofuranosidase. *Tetrahedron Lett.*, 43 (52), pp.9653–9655.

Rémond, C., Plantier-Royon, R., Aubry, N., Maes, E., Bliard, C. & O'Donohue, M.J. (2004) Synthesis of pentose-containing disaccharides using a thermostable α-L-arabinofuranosidase. *Carbohydr. Res.*, 339 (11), pp.2019–2025.

Sakamoto, T., Fujita, T. & Kawasaki, H. (2004) Transglycosylation catalyzed by a *Penicillium chrysogenum* exo-1,5-α-L-arabinanase. *Biochim. Biophys. Acta, Gen. Subj.*, 1674 (1), pp.85–90.

Sugimura, M., Nishimoto, M. & Kitaoka, M. (2006) Characterization of glycosynthase mutants derived from glycoside hydrolase family 10 xylanases. *Biosci., Biotechnol., Biochem.*, 70 (5), pp.1210–1217.

Tramice, A., Pagnotta, E., Romano, I., Gambacorta, A. & Trincone, A. (2007) Transglycosylation reactions using glycosyl hydrolases from *Thermotoga neapolitana*, a marine hydrogen-producing bacterium. *J. Mol. Catal. B: Enzym.*, 47 (1-2), pp.21–27.

Utille, J.-P. & Jeacomine, I. (2007) Synthesis of a library of allyl α-L-arabinofuranosyl-α- or β-D-xylopyranosides; route to higher oligomers. *Carbohydr. Res.*, 342 (17), pp.2649–2656.

Wan, C., Chen, W., Chen, C., Chang, M., Lo, L. & Li, Y. (2007) Mutagenesis and mechanistic study of a glycoside hydrolase family 54 α-L-arabinofuranosidase from *Trichoderma koningii*. *Biochem. J.*, 401, pp.551–558.

Part III

Story of the pentose-acting enzyme *Tx*Abf since 2000

Among the diversity of plant cell wall hemicelluloses-degrading enzymes was found the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*Tx*Abf) which has constituted the model protein of this thesis work. The knowledge acquired on this enzyme over the past decade is presented, from its discovery to the exploration of its structural and mechanistic features.

III-1. The GH51 α-L-arabinofuranosidases

α-L-arabinofuranosidases (Abfs) are hemicellulases hydrolyzing α-L-arabinofuranosyl substitutions found in hemicelluloses. Abfs, from bacteria, fungi or plants, are classified in GH families 3, 43, 51, 54 and 62 according to the CAZy classification (Lombard et al., 2014). Among them, enzymes acting on polymeric substrates, arabinoxylan arabinofuranohydrolases (AXH), can be further divided according to their substrate specificity, either acting on mono-substituted xylose units with *O*-2 or *O*-3 linkages (AXH-m) or on disubstituted units releasing only the α-(1,3) linked L-arabinofuranosyl moiety (AXH-d3) (Badal C, 2000). Up to date[5], the GH51 family counts 967 members, and among the 64 that were characterized 3 are described as *endo*-glucanases and 1 as a cellulase/xylanase. Among GH51 Abfs, only 6 proteins, from bacterial origin, were successfully crystallized: those from *Geobacillus stearothermophilus* (*Gs*Abf, Uniprot ID: B3EYN4) (Hovel et al., 2003), *Clostridium thermocellum* (*Ct*Abf, A3DIH0) (Taylor et al., 2006), *Bifidobacterium longum* (*Bl*Abf, Q841V6), *Thermobacillus xylanilyticus* (*Tx*Abf, O69262) (Paës et al., 2008), *Thermotoga petrophila* (*Tp*Abf, A51KC8) (Souza et al., 2011), and *Thermotoga maritima* (*Tm*Abf, Q9WYB7) (Im et al., 2012).

Besides, a phylogenetic analysis performed on GH51 Abfs pertaining to the different kingdoms revealed that they can be subdivided into three different groups, showing that *Tx*Abf belongs to group C whereas the five other crystallized GH51 Abfs belong to clade B (Lagaert et al., 2014). This may let us foresee that *Tx*Abf display original features relative to other Abfs currently under study.

References

Badal C, S. (2000) α-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. *Biotechnol. Adv.*, 18 (5), pp.403–423.

Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. & Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 α-L-arabinofuranosidase. *EMBO J.*, 22 (19), pp.4922–4932.

Im, D.-H., Kimura, K., Hayasaka, F., Tanaka, T., Noguchi, M., Kobayashi, A., Shoda, S., Miyazaki, K., Wakagi, T. & Fushinobu, S. (2012) Crystal structures of glycoside hydrolase family 51 α-L-arabinofuranosidase from *Thermotoga maritima*. *Biosci., Biotechnol. Biochem.*, 76 (2), pp.423–428.

Lagaert, S., Pollet, A., Courtin, C.M. & Volckaert, G. (2014) β-xylosidases and α-L-arabinofuranosidases: accessory enzymes for arabinoxylan degradation. *Biotechnol. Adv.*, 32 (2), pp.316–332.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, 42 (Database issue), pp.D490–D495.

Paës, G., Skov, L.K., O'Donohue, M.J., Rémond, C., Kastrup, J.S., Gajhede, M. & Mirza, O. (2008) The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry*, 47 (28), pp.7441–7451.

Souza, T., Santos, C.R., Souza, A.R., Oldiges, D.P., Ruller, R., Prade, R. a., Squina, F.M. & Murakami, M.T. (2011) Structure of a novel thermostable GH51 α-L-arabinofuranosidase from *Thermotoga petrophila* RKU-1. *Protein Sci.*, 20 (9), pp.1632–1637.

Taylor, E.J., Smith, N., Turkenburg, J., D'Souza, S., Gilbert, H. & Davies, G. (2006) Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Ara*f*51, from *Clostridium thermocellum*. *Biochem. J.*, 395, pp.31–37.

[5] 8th, July 2014

III-2. Main features of *Tx*Abf

*Tx*Abf is a GH51 Abf produced in nature by *Thermobacillus xylanilyticus*, an anaerobic, thermophilic and xylanolytic bacteria which was found in 2000 underneath a manure heap, in a farm in northern France (Touzel et al., 2000). It was then found that the Gram positive bacteria *Thermobacillus xylanilyticus* genome displays several hemicellulases-encoding genes, including *Tx*Abf one (Debeche et al., 2000), and the expression level of these (endoxylanase, xylosidase, arabinofuranosidase, esterase) was demonstrated to be substrate-dependent (Rakotoarivonina et al., 2012). The full genome of *Tx*Abf has been sequenced and the annotation that is currently ongoing (Caroline Rémond group, Reims, *Unpublished data*) reveals that other Abfs would be present.

1. Biochemical properties and activities

*Tx*Abf is naturally found under hexameric form, the molar mass of each monomer being equal to 56 kDa. Optimum pH and temperature are respectively 5.8 and 75 °C. *Tx*Abf does cleave (1,2) and (1,3) linkages between α-L-arabinofuranosyl moieties and D-xylopyranosyl residues constituting the AX backbone, its natural substrate. *Tx*Abf is also highly active on the glycosidic linkage between *para*-nitrophenyl (*p*NP) and α-L-arabinofuranosyl of *p*NP-α-L-Ara*f,* the practical synthetic chromogenic substrate used for all Abfs kinetics studies.

*Tx*Abf acts with a two-step displacement mechanism with retention of configuration, and, as many GHs, harbors two activities, hydrolysis and transglycosylation (Fig. 1). Among the GH51 Abfs, to date and to our knowledge only *Tx*Abf (Rémond et al., 2002, 2004), *Pc*Abf[6] (Sakamoto et al., 2004), *Ct*Abf (Chlubnova et al., 2010) and *Tm*Abf (Im et al., 2012) were assessed to perform transglycosylation.

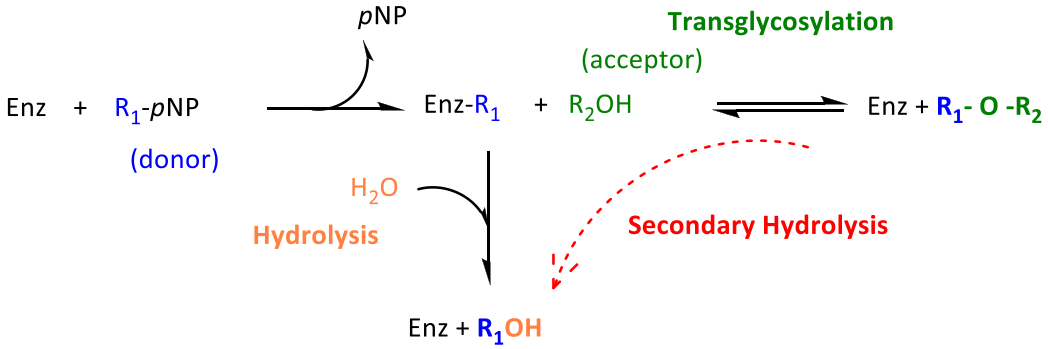


Fig. 1: *Tx*Abf-catalyzed reactions (R_1-LG = *p*NP-α-L-Ara*f* or *p*NP-β-D-Gal*f*, R_2 = Bn-α-D-Xyl*p*, X_2 or X_3 for instance).

Catalytic mechanism studies revealed that *Tx*Abf has drastically different affinities towards *p*NP-α-L-Ara*f* (**1**) relative to *p*NP-β-D-Gal*f* (**2**). The Michaelis-Menten constant (K_M) for the first substrate

[6] Abfs from *Penicillium chrysogenum* (Uniprot: **B5MGR0**)

lies within the millimolar range (0.72 mM, at 60 °C) whereas it is superior to 50 mM for the second one, when these two substrates differ via the presence of an extra hydroxymethyl at *C*-5 position for *p*NP-β-D-Gal*f* (Fig. 2).

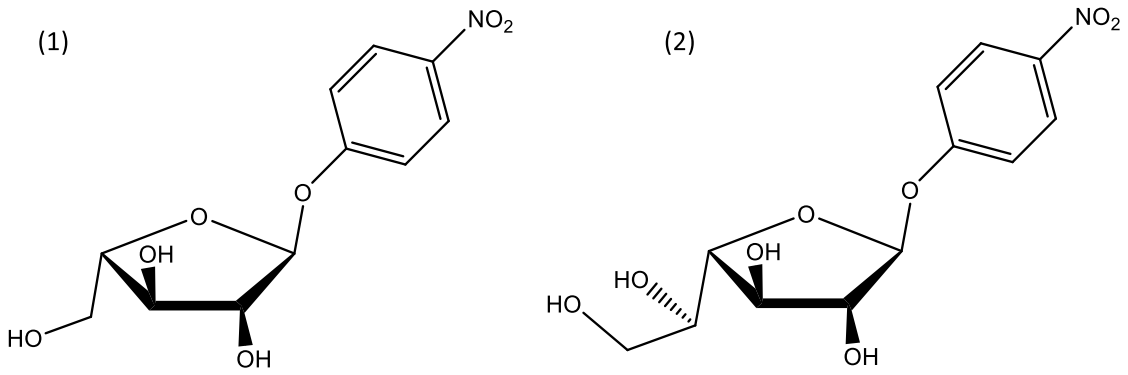


Fig. 2 : Donor substrates of interest for *Tx*Abf study, *p*NP-α-L-Ara*f* (1) and *p*NP-β-D-Gal*f* (2)

Specific activities (S.A.), and consequently performance constants (k_{cat}/K_M) are also dramatically affected with *p*NP-β-D-Gal*f*, respectively 40 and 6000-fold lower than with *p*NP-α-L-Ara*f* (Table 1).

Table 1: Kinetics parameters in hydrolytic mode for *Tx*Abf with *p*NP-α-L-Ara*f* and *p*NP-β-D-Gal*f* as donor substrates

Substrate	K_M (mM)	S.A. * (IU.mg^{-1})	k_{cat} (s^{-1})	k_{cat}/K_M ((s.mM)$^{-1}$)	References
*p*NP-α-L-Ara*f*	0.72 *± 0.04*	615 *± 51*	575 *± 48*	795 *± 67*	Arab-Jaziri et al., 2012
*p*NP-β-D-Gal*f*	> 50	-	-	0.557 *± 0.01*	

* determined at 60 °C in sodium acetate buffer (50 mM, pH 5.8) with 5 mM of *p*NP-furanoside.

Steric hindrances would be responsible for these different behaviors. Furthermore, *Tx*Abf activity was also evaluated with two other substrates derived from *p*NP-β-D-Gal*f*: *p*NP-β-D-fucofuranoside and *p*NP-6-deoxy-6-fluoro-β-D-Gal*f* (Euzen et al., 2005). These are two relevant substrates since their steric hindrance at *C*-5 position is intermediary to those of *p*NP-α-L-Ara*f* and *p*NP-β-D-Gal*f*. Affinity constants are really inferior to *p*NP-β-D-Gal*f* one, with 5.72 and 7.59 mM *vs* K_M > 50 mM, respectively. These results are in accordance with a destabilizing effect of the hydroxymethyl on the interaction network in the -1 subsite.

Along the story of *Tx*Abf, this was proved to perform transglycosylation one decade ago, in 2004, with the transfer an L-arabinofuranosyl moiety onto a synthetic D-xyloside acceptor (Rémond et al., 2004).

Besides, surprisingly, unlike with *p*NP-α-L-Ara*f* the use of *p*NP-β-D-Gal*f* as a donor substrate offers higher yields, 7 and 75% respectively, with benzyl α-D-xylopyranoside (Bn-α-D-Xyl*p*) as acceptor sugar (Table 2).

Table 2: *Tx*Abf-catalyzed transglycosylation reactions

Donor	Acceptor	Maximum Yield (%)	Linkage	References
*p*NP-β-D-Gal*f*	Bn-α-D-Xyl*p*	75	β-(1,2)	Rémond et al., 2005.
*p*NP-α-L-Ara*f*	Bn-α-D-Xyl*p*	7	α-(1,2)	Rémond et al., 2004.

NB: With *p*NP-α-L-Ara*f*, maximum yield is reached within 10 min with 0.11 IU. 3 h are necessary for *p*NP-β-D-Gal*f* with 11 IU. The number of IU corresponds to the hydrolytic activity of *Tx*Abf with *p*NP-α-L-Ara*f*.

2. Insights into the three-dimensional structure of *Tx*Abf

The 3D structure of *Tx*Abf obtained by X-ray crystallography in 2008 (PDB 2VRQ, Paës et al., 2008) allowed a deep exploration of secondary, tertiary and quaternary structures. It revealed two main domains: a TIM-barrel catalytic domain and a C-terminal domain with a jelly-roll folding whose function is still unknown (Fig. 3).

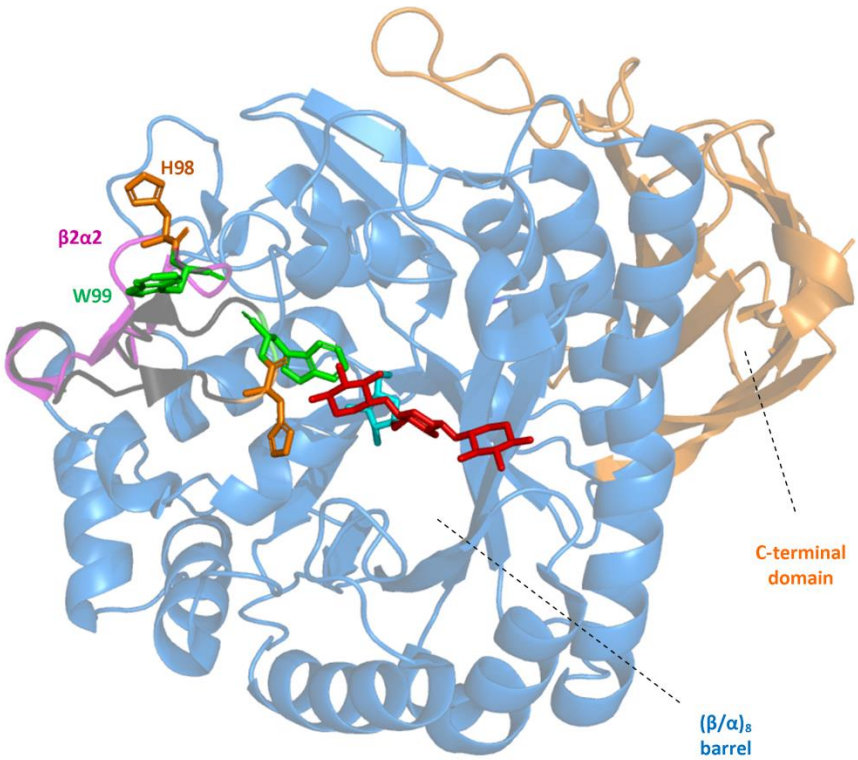


Fig. 3: Crystallographic structure of *Tx*Abf co-crystallized with XA^{3}XX (PDB ID: 2VRQ, Paës et al., 2008). Open (pink) and closed (black) conformations of the β2α2 loop are superimposed (PDB codes 2VRQ and 2VRK). Residues H98 (orange) and W99 (green) are shown as sticks for both conformations.

The *Tx*Abf active site is constituted of a pocket receiving the donor substrate glycon part (-1 subsite), the L-arabinofuranosyl unit for natural substrates (Fig. 4). Above this pocket is a groove in which acceptor sugars are located, D-xylopyranosyl units in the case of XA^3XX, in subsites +2', +1 and +2. This special architecture makes *Tx*Abf a debranching *exo*-enzyme.

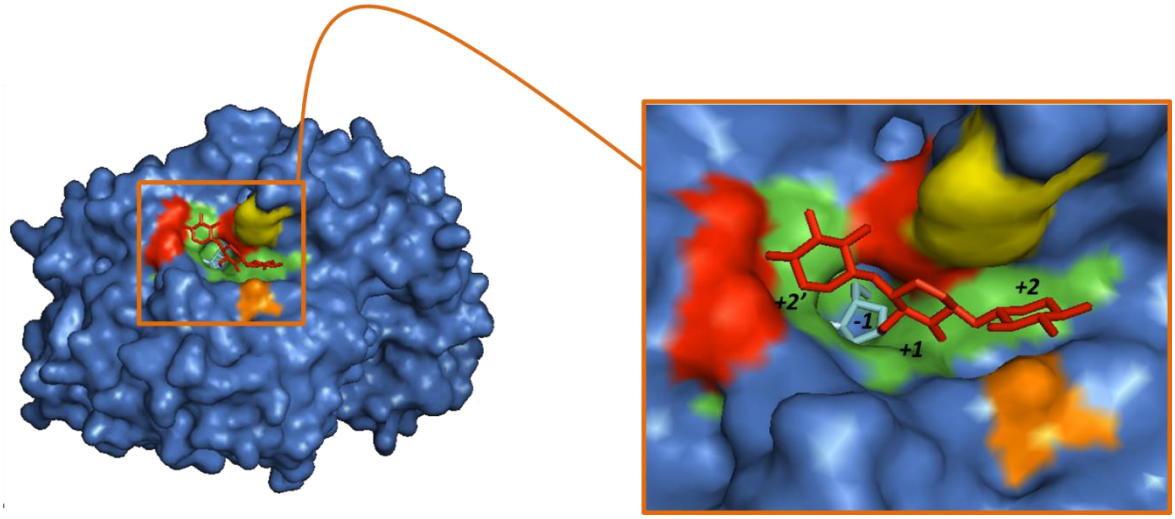


Fig. 4: *Tx*Abf active site topography co-crystallized with XA^3XX (Paës et al., 2008)

Regarding amino acids directly involved in the catalytic mechanism, residues E176 and E298 are playing the role of general acid/base and nucleophile, respectively. E176 is located between -1 and +1 subsites at the cleavage site and E298 in -1 donor subsite.

Crystallographic data revealed the presence of the β2α2 loop which can adopt a closed or open conformation (Fig. 3). Its functional importance was demonstrated by the deletion of 10 amino acids leading to a 72-fold decrease in performance constant towards *p*NP-α-L-Ara*f* (and 10-fold increase in K_M value) (Paës et al., 2008).

Besides, H98 and W99 amino acids, located at one end of the mobile β2α2 loop, and constitutive of the active site when this is in a closed conformation, were shown to be relevant residues for *Tx*Abf hydrolytic activity. W99 is crucial for donor substrate recognition and efficient hydrolysis, and molecular dynamics also showed that W99A mutation altered β2α2 loop motion (Arab-Jaziri et al., 2012) (cf Supporting Information A.1).

References

Arab-Jaziri, F., Bissaro, B., Barbe, S., Saurel, O., Débat, H., Dumon, C., Gervais, V., Milon, A., André, I., Fauré, R. & O'Donohue, M.J. (2012) Functional roles of H98 and W99 and β2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *FEBS J.*, 279 (19), pp.3598–3611.

Chlubnova, I., Filipp, D., Spiwok, V., Dvorakova, H., Daniellou, R., Nugier-Chauvin, C., Kralova, B. & Ferrieres, V. (2010) Enzymatic synthesis of oligo-D-galactofuranosides and L-arabinofuranosides: from molecular dynamics to immunological assays. *Org. Biomol. Chem.*, 8 (9), pp.2092–2102.

Debeche, T., Cummings, N., Connerton, I., Debeire, P. & O'Donohue, M.J. (2000) Genetic and biochemical characterization of a highly thermostable α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *Appl. Environ. Microbiol.*, 66 (4), pp.1734–1736.

Euzen, R., Lopez, G., Nugier-, Chauvin, C., Ferrières, V., Plusquellec, D., Rémond, C. & O'Donohue, M. (2005) A chemoenzymatic approach for the synthesis of unnatural disaccharides containing D-galacto or D-fucofuranosides. *Eur. J. Org. Chem.*, 2005 (22), pp.4860–4869.

Im, D.-H., Kimura, K., Hayasaka, F., Tanaka, T., Noguchi, M., Kobayashi, A., Shoda, S., Miyazaki, K., Wakagi, T. & Fushinobu, S. (2012) Crystal structures of glycoside hydrolase family 51 α-L-arabinofuranosidase from *Thermotoga maritima*. *Biosci., Biotechnol. Biochem.*, 76 (2), pp.423–428.

Paës, G., Skov, L.K., O'Donohue, M.J., Rémond, C., Kastrup, J.S., Gajhede, M. & Mirza, O. (2008) The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry*, 47 (28), pp.7441–7451.

Rakotoarivonina, H., Hermant, B., Monthe, N. & Rémond, C. (2012) The hemicellulolytic enzyme arsenal of *Thermobacillus xylanilyticus* depends on the composition of biomass used for growth. *Microb. Cell Fact.*, 11 (159), pp.1–12.

Rémond, C., Ferchichi, M., Aubry, N., Plantier-Royon, R., Portella, C. & O'Donohue, M.J. (2002) Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α-L-arabinofuranosidase. *Tetrahedron Lett.*, 43 (52), pp.9653–9655.

Rémond, C., Plantier-Royon, R., Aubry, N., Maes, E., Bliard, C. & O'Donohue, M.J. (2004) Synthesis of pentose-containing disaccharides using a thermostable α-L-arabinofuranosidase. *Carbohydr. Res.*, 339 (11), pp.2019–2025.

Sakamoto, T., Fujita, T. & Kawasaki, H. (2004) Transglycosylation catalyzed by a *Penicillium chrysogenum* exo-1,5-α-L-arabinanase. *Biochim. Biophys. Acta, Gen. Subj.*, 1674 (1), pp.85–90.

Touzel, J.P., Donohue, M.O., Debeire, P., Samain, E. & Breton, C. (2000) *Thermobacillus xylanilyticus* gen. nov., sp. nov., a new aerobic thermophilic xylan-degrading bacterium isolated from farm soil. *Int. J. Syst. Evol. Micr.*, 50, pp.315–320.

III-3. *Tx*Abf, towards a new bio-tool for glyco-synthesis

During Dr. Faten Arab's doctoral work, conducted in the O'Donohue's group from 2008 to 2012, the potential of *Tx*Abf as a tool for chemo-enzymatic synthesis specialized on pentoses/furanoses was investigated. First, the acceptor specificity of *Tx*Abf was studied through reactions of transgalactosylation catalyzed on a wide variety of acceptors: Bn-α-D-Xyl*p*, benzyl α-D-glucopyranoside (Bn-α-D-Glc*p*), benzyl 6-deoxy-α-D-glucopyranoside (Bn-α-D-6dGlc*p*), benzyl α-D-galactopyranoside (Bn-α-D-Gal*p*), benzyl α-L-rhamnopyranoside (Bn-α-L-Rha*p*), benzyl 2-*N*-acetamido-2-deoxy-α-D-glucopyranoside (Bn-α-D-Glc*pN*Ac), benzyl α-D-lyxopyranoside (Bn-α-D-Lyx*p*), benzyl α-D-mannopyranoside (Bn-α-D-Man*p*), and benzyl α-D-xylobioside (*Unpublished data*). This work confirmed the wide substrate specificity in +1 subsite and pointed out the importance of modifications at the *C*-2 position of the substrate acceptor which greatly affect the tranglycosylation abilities of *Tx*Abf. Moreover, transgalactofuranosylation kinetics revealed that the enzyme prefers non-encumbered acceptor substrates at the *C*-5 position, probably due to a better fit in +1 subsite. Complementarily to such a rational approach, a directed evolution strategy was applied to *Tx*Abf. Through the screening of a library of *Tx*Abf-encoding gene randomly mutagenized, mutants with greatly diminished hydrolytic activity were identified, while their transfer activity onto synthetic acceptor (Bn-α-D-Xyl*p*), evaluated in the course of the present thesis work, was conserved or slightly improved (Arab-Jaziri et al., 2013), cf Supporting information A.2). Along this line, a new library was screened using a two-step protocol that was adapted to *Tx*Abf, in collaboration with Charles Tellier and Michel Dion (University of Nantes, France) who developed this screening. The major aim was to

find mutants with improved transglycosylation ability onto natural acceptors (XOS) while displaying low hydrolytic potency (Arab-Jaziri et al., 2014, *Submitted article,* cf Supporting information A.3). To do so, the screening strategy relied on the measure of aglycon (indolyl) release with and without acceptor (Koné et al., 2009). Clones with a gain of intensity superior to 4 in presence of XOS acceptor were postulated to be potential positive candidates for the targeted transfer activity. Thereby, a set of 11 clones was retained after successive secondary screening steps consisting in: XOS-inhibition alleviation[7] study and TLC-based transglycosylation monitoring. At this stage, catalyzed reactions were further analyzed at the very beginning of my doctoral work through time-course NMR and mass spectroscopy, allowing the selection of 5 mutants with significantly improved transfer activity, up to 2-fold, albeit remaining modest for applied glyco-synthetic purposes (Arab-Jaziri et al., 2014, *Submitted article*).

It is noteworthy that both studies, based on random library screening, revealed the importance of 2nd shell amino acids as key positions to be mutated for transglycosylation improvement purposes. These were isolated from irrelevant mutations and their role was desiccated.

References

Arab-Jaziri, F., Bissaro, B., Dion, M., Saurel, O., Harrison, D., Ferreira, F., Milon, A. & Tellier, Charles, O'Donohue, Michael J, Fauré, R. (2013) Engineering transglycosidase activity into a GH51 α-L-arabinofuranosidase. *New Biotechnol.*, 30 (5), pp.536–544.

Koné, F.M.T., Le Béchec, M., Sine, J.-P., Dion, M. & Tellier, C. (2009) Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng. Des. Sel.*, 22 (1), pp.37–44.

III-4. Goals and Challenges of this thesis work

1. Scientific questions

As for any fundamental work a central scientific question drives or orientates the different axes explored or choices that are made. In relationship with the literature review realized in the first chapter, the present thesis work aimed at answering the following questioning:

How, for glycoside hydrolases, and more specifically for furanose-acting enzymes, can the hydrolysis/transglycosylation (H/T) ratio be modulated?

[7] *Tx*Abf is inhibited by XOS acceptors which makes even harder the detection of improved transglycosylating mutants since their detection relies on an increase of LG release (global activity) in presence of acceptor, the inhibition potential of which can hide an eventual increase of transglycosylation rate.

Behind this question, one may withdraw two sub-interrogations about the key molecular determinants responsible for:

- the enhancement of the transglycosylation activity *per se* (thermodynamic control, transfer rates between water and carbohydrate acceptors).

- a control and/or deletion of secondary hydrolysis (kinetic control, synthesis versus hydrolysis rates of the transglycosylation product).

- the regioselectivity of the transglycosylation reaction.

2. Developed Strategy

The developed strategy aims to reach two goals that could be considered as interdependent. The first aim is obviously to answer the hereinbefore asked questions about H/T ratio modulation, while the second one is to generate **furanose-specific glyco-synthetic enzymes** devoted to tailor-made AXOS conception. These two objectives are linked since the understanding of molecular determinants will need the access to enzymes displaying transglycosylation potency and, in turn, the acquisition of knowledge on transglycosylation influencing factors should provide the means to engineer better transglycosylating tools. However, the main difficulty of such a work relies in finding the evolution pathway that should lead to the hypothetic best transglycosylating mutant without stepping into a dead-end road. For the latter reasons, the strategy developed is an iterative process relying on the alternation of phases of random or irrational mutagenesis for diversity generation coupled to HTS, and phases of fundamental understanding of the behavior of newly generated biocatalysts. In this perspective, two main axes were developed:

I – Rational and Fundamental understanding of *Tx*Abf mutants

i) Finding the regioselectivity determinants of transglycosylation (Part I)
ii) Biochemical and biophysical characterization (Part II and IV)
iii) *In silico* approach to identify binding hotpsots (Part III)

II – Global Directed Evolution process

i) Random mutagenesis from relevant mutants scaffold and screening (Part II and IV)
ii) Ordered recombination (Part IV)

III – Prospective work: Implementation of chemo-enzymatic pathways (Part V)

Second Chapter

Results and Discussion

Part I

Acceptor anomeric control of the transfuranosylation regioselectivity in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*

(Article in preparation)

Acceptor anomeric control of the transfuranosylation regioselectivity in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*

Bastien Bissaro[1,2,3], Hélène Débat[4], Olivier Saurel[1,5], Alain Milon[1,5], Régis Fauré[1,2,3] and Michael J. O'Donohue[1,2,3,*]

[1] *Université de Toulouse; INSA,UPS,INP; 135 Avenue de Rangueil, F-31077 Toulouse, France*
[2] *INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France*
[3] *CNRS, LISBP UMR5504, F-31400 Toulouse, France*
[4] *Université Versailles St-Quentin, Institut de Génétique et Microbiologie, Faculté des Sciences Paris-Sud, F-91405 Orsay, France*
[5] *CNRS, IPBS UMR 5089, Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, BP 64182, Toulouse, France*

Abstract

The broad regioselectivity of *exo*-acting glycoside hydrolases is often related to a large acceptor subsite topology where the acceptor can adopt different binding modes or at least various positioning. The anomeric control onto the transglycosylation regioselectivity was clearly demonstrated for the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. Indeed, pyranosidic acceptor with α-D-anomery, such as benzyl α-D-xylopyranoside, leads to a strict regiospecific *O*-2 linkage formation whereas pyranosidic acceptors with β-D-anomery, such as benzyl β-D-xylopyranoside or xylo-oligosaccharides, lead to the formation of regioisomeric mixture (i.e. *O*-2 and *O*-3 glycosidic linkages). *In silico* docking analysis coupled with experimental approach through semi-rational mutagenesis suggest that this anomeric control might be related to an alternative positioning as an 'orientation switch of the aglycon' due to the original active site topology of this enzyme. Indeed the benzyl (or saccharidyl) 'aglycon' either binds to the subsite +2', resulting in regiospecific reaction, or occupies the subsite +2, leading to less tightly controlled regioselectivity. Furthermore, +2 acceptor subsite mutagenesis (through N216W substitution) created a regiospecific mutant for *O*-2 linkage formation with pyranosidic acceptor displaying β-D-anomery, validating the presence of the β-D-benzyl aglycon in +2 subsite and providing a valuable regiospecific glycosynthesis tool for the enzymatically-catalyzed transfuranosylation reaction.

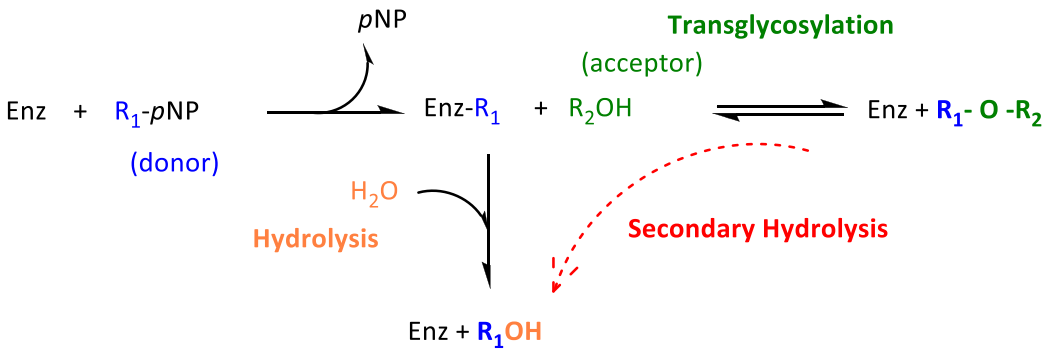
Correspondance: M. J. O'Donohue, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, 135 Avenue de Rangueil, 31077 Toulouse cedex 4, France; Fax +33 5 6155 9400; Tel: +33 5 6155 9428; E-mail: michael.odonohue@insa-toulouse.fr

Keywords: transglycosylation, regioselectivity, anomeric control, glycoside hydrolase, pentoses/furanoses

Abbreviations: Abfs, α-L-arabinofuranosidases; D-Xyl*p*, D-xylopyranosyl; Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p*, benzyl β-D-galactofuranosyl-(1,2)-α-D-xylopyranoside; Bn α-D-Xyl*p*, benzyl α-D-xylopyranoside; Bn β-D-Xyl*p*, benzyl β-D-xylopyranoside; FH, furanoside hydrolase; GH, glycoside hydrolase; L-Ara*f*, L-arabinofuranosyl; H/T, hydrolysis/transglycosylation partition; *p*NP, *para*-nitrophenol; *p*NP-α-L-Ara*f*, *para*-nitrophenyl α-L-arabinofuranoside; R_T, transfer rate; *Tx*Abf, α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*; X, donor conversion rate; Y, yield

1. Introduction

In the perspective of widening the diversity of available sugar-based structures, Carbohydrate Active Enzymes (CAZymes), and especially glycosides hydrolases (GHs), represent preferential targets due to their relative flexible specificity of substrate and regioselectivity, their ease of access and of production coupled to eco-friendly usage conditions. More importantly, GHs are biocatalysts which through their molecular mechanism might display a dual selectivity, an innate hydrolytic activity and a putative ability to transglycosylate. The main stake, in the quest to develop finely tuned sugar-transferring enzymes for glycosynthesis, is to be able to take advantage of this potential of transglycosylation and to improve it, while limiting hydrolyses (primary and secondary) (Scheme 1).



Scheme 1: Reactions catalyzed by *Tx*Abf: hydrolysis (primary and secondary) and transfuranosylation of exogenous acceptors. The transfuranosylation product can also be recognized as a substrate and then be degraded through secondary hydrolysis.

Regarding the target glycosides for chemo-enzymatic synthesis, hexose-acting enzymes have been extensively studied (Schmaltz et al., 2011) and encountered major improvements since the breakthrough of the 'glycosynthase concept' (Mackenzie et al., 1998; Malet & Planas, 1998) and artificially evolved GHs (Review). However, enzymatic-mediated synthesis of pentose/furanose-based carbohydrates is so far a highly under-developed field (Peltier et al., 2008; Dumon et al., 2012). Nevertheless, certain furanoside hydrolases (FHs), such as α-L-arabinofuranosidases (Abfs) of the CAZy family GH51 (Lombard et al., 2014), display the ability to catalyze glycosynthesis, though this activity is in competition with a dominant hydrolysis (Dumon et al., 2012; Rémond et al., 2005; Arab-Jaziri et al., 2013; Chlubnova et al., 2010; Chlubnová et al., 2014). Beyond the question of the hydrolysis/transglycosylation (H/T) partition the regioselectivity issue should also be tackled. Indeed, the wide substrate specificity of GHs, and especially concerning *exo*-type enzymes, is often related to loosened regioselective behavior that usually leads to the synthesis of a mixture of regioisomers when acting in transglycosylation mode (Ajisaka & Yamamoto, 2002; Dion, Osanjo, et al., 2001; Dion, Nisole, et al., 2001). Frequently acceptor specificity appears to be low, with the acceptor subsite(s) of glycosidases accommodating a wide range of sugars and non-sugar acceptor molecules.

Nevertheless, various studies proposed the use of a range of glycoside acceptors bearing anomeric substituents (Marton et al., 2008; Prade et al., 1998; Lopez & Fernandez-Mayoralas, 1994; Stick et al., 2004; Faijes et al., 2006; Teze et al., 2013) or involving an anomeric control (Nilsson, 1987; Crout et al., 1991; Crout & Vic, 1998; Zeng et al., 2000; MacManus & Vulfson, 2000; Giordano et al., 2005) to permit higher regioselectivity through modified interactions within the aglycon-binding subsite, especially in subsite +2 (Stick et al., 2004; Faijes et al., 2006). Moreover unraveling and understanding the structural features responsible for this is also capital for chemo-enzymatic synthesis purposes (Dion, Nisole, et al., 2001; Marton et al., 2008; Ngiwsara et al., 2012).

The glycosynthetic capacity of the GH51 Abf from *Thermobacillus xylanilyticus* (*Tx*Abf) and mutant thereof have been extensively studied (Rémond et al., 2002; Rémond et al., 2004; Rémond et al., 2005; Arab-Jaziri et al., 2013; Bissaro et al., 2014). The transglycosylation ability of *Tx*Abf on synthetic acceptors sugars such as benzyl α-D-xylopyranoside (Bn-α-D-Xyl*p*) is relatively low with its reference donor substrate, *para*-nitrophenyl α-L-arabinofuranoside (i.e. 16% yield of L-arabinofuranosyl transfer from *p*NP-α-L-Ara*f*) (Arab-Jaziri et al., 2013) and somewhat higher with it structural β-D-galactofuranoside analog. Unlike the five other crystallized GH51 Abfs, the acceptor binding site of *Tx*Abf appears as a narrow groove and transfuranosylation seems to be strictly regiospecific with acceptors such as α-D-anomer of glycopyranosides, leading to the formation of *O*-(1,2) linkage (Rémond et al., 2005; Chlubnová et al., 2014). It is noteworthy that the active site of *Tx*Abf displays an original architecture, since the catalytic pocket (donor subsite -1) of this *exo*-enzyme is located in the middle of the acceptor groove with two additional subsites (denoted +2 and +2') that flank the +1 subsite (Paës et al., 2008; Arab-Jaziri et al., 2012) (Fig. 1).

In the present study the attention was therefore focused on the understanding of the anomeric control by *Tx*Abf. Particularly, comparative experiments with the corresponding β-D-anomer glycopyranoside acceptors led to address the question of how the benzyl aglycon anomery drives the regioselectivity of the transfuranosylation reaction. In this respect, the global strategy consisted in employing site-directed mutagenesis on specific targeted amino acids constitutive of the active site to assess an alternative positioning of the acceptor according to its anomery (Fig. 1). The amino acid pair H98-W99 are residues located at the 'spatial end' of the mobile β2α2 loop, constitutive of -1 and +2' subsites and were demonstrated to be modulators of the hydrolytic activity (Arab-Jaziri et al., 2012) (Supporting information A1). W302, W248 and N216 are constitutive of subsites -1/+1, +1 and +2, respectively, and were thought to possibly alter directly the acceptor positioning (Paës et al., 2008). Furthermore, docking experiments were performed to provide a theoretical explanation to experimental results.

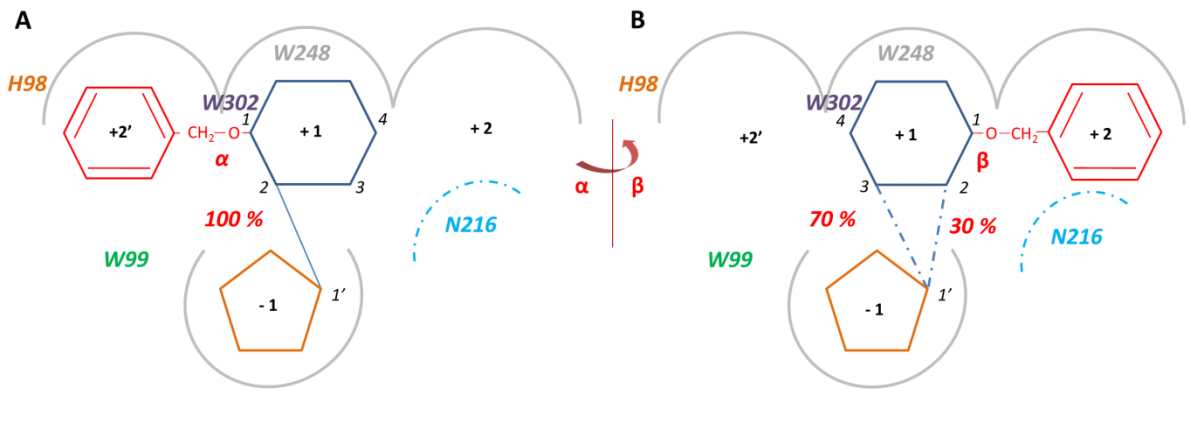


Fig. 1. Proposed anomery-induced'orientation switch' of the benzyl aglycon in acceptor subsites of *Tx*Abf. (A) Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p* or (B) Bn β-D-Gal*f*-(1,2)-β-D-Xyl*p* and Bn β-D-Gal*f*-(1,3)-β-D-Xyl*p*.

2. Results

2.1. Acceptor anomeric control and regioselectivity with *Tx*Abf

2.1.1. Experimental observations

A unique transgalactofuranosylation disaccharide, benzyl β-D-galactofuranosyl-(1,2)-α-D-xylopyranoside (Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p*, (Table SI 1), (Rémond et al., 2005)) was observed by TLC and ^{1}H-NMR analyses when using benzyl α-D-xylopyranoside (Bn-α-D-Xyl*p*) as an acceptor whereas two transglycosylation products (i.e. the *O*-2 and *O*-3 linkages with a ratio of 7:3) were obtained in the case of acceptors with β-D-anomeries, either the synthetic benzyl β-D-xylopyranoside (Bn-β-D-Xyl*p*) or the natural xylobiose and xylotriose substrates (β-D-anomery of their interglycosidic linkages). While the α-D-benzyl aglycon drives *O*-2 transfer regiospecificity, the β-D-anomery probably offers a greater flexibility for acceptor docking leading to a lack of transglycosylation regiospecificity. Thereby, it appears that *Tx*Abf transglycosylation is subject to an anomeric control and the original 'pocket and groove' shape of its active site might be responsible for this.

2.1.2. In silico *analysis of benzyl aglycon positioning into acceptor subsites*

Docking analysis demonstrated that in the case of the *O*-2 regiospecific transgalactofuranosylation the α-D-benzyl aglycon might be positioned into +2' subsite with a binding energy mean value of -5.79 kcal.mol^{-1} and showing staking-like interactions between the benzyl ring and H98 residue (Figs. 2A and 4). It is noteworthy that the benzyl group of the experimentally unproduced benzyl β-D-galactofuranosyl-(1,3)-α-D-xylopyranoside may be localized within the subsite +2 (binding energy of -3.33 kcal.mol^{-1}, Fig. 2B). Moreover, the β-D-benzyl aglycon is thought to preferentially take place within the subsite +2 whatever regioisomer (*O*-2 and *O*-3) is formed, with binding energies closer to the value calculated for the natural arabinoxylo-oligosaccharide XA^{3}X (-11.41, -12.75 and -16.89 kcal.mol^{-1}, respectively, Fig 2C to E).

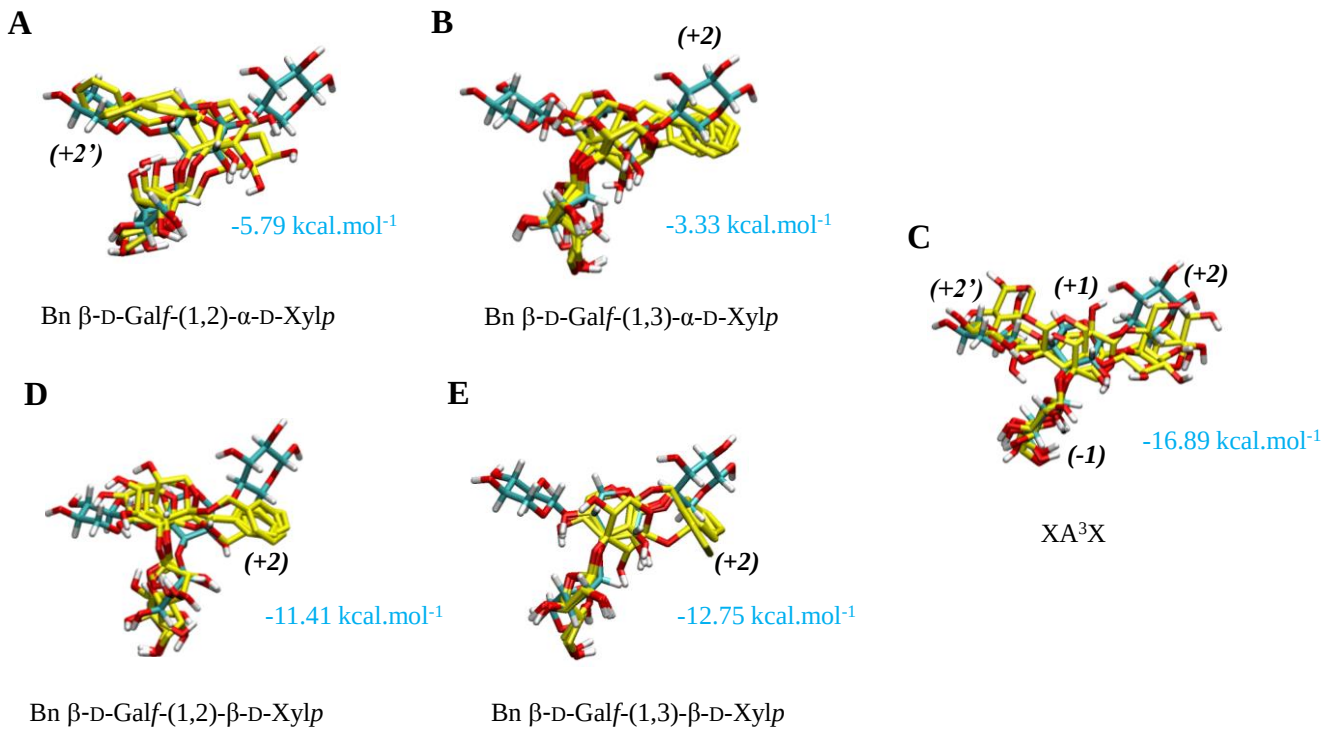


Fig. 2. Docking experimentations of *Tx*Abf wild-type with transglycosylation products obtained from the acceptor with α-D-anomery (A) Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p*, (B) Bn β-D-Gal*f*-(1,3)-α-D-Xyl*p*, or β-D-anomeries (C) the α-(1,3)-L-arabinofuranoxylo-oligosaccharides (denoted as XA^3X according to (Fauré et al., 2009)), (D) Bn β-D-Gal*f*-(1,2)-β-D-Xyl*p* and (E) Bn β-D-Gal*f*-(1,3)-β-D-Xyl*p*.

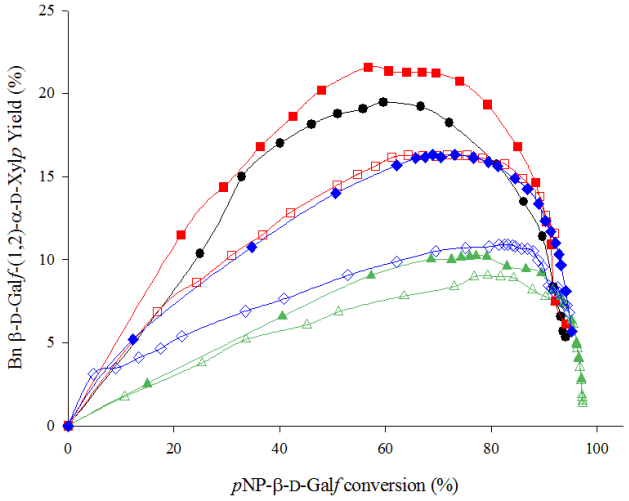
2.2. On the role of subsite +2' in anomeric control and transglycosylation: H98 and W99 key players

2.2.1. Experimental investigation of H98 and W99 role in transglycosylation regioselectivity

To investigate this anomeric control onto the acceptor positioning toward the +2' side, the two key residues H98 and W99 (Arab-Jaziri et al., 2012), located in this zone of the active site, were mutated (H98A, H98F, W99A, W99F, H98AW99A and H98FW99F) and the transglycosylation capacity of these mutants was evaluated. As a result, *Tx*Abf and H98A or H98F mutants catalyse the transfer reaction of D-galactofuranosyl moiety (D-Gal*f*) onto Bn-α-D-Xyl*p* at approximately the same rate (Fig. 3A) with equivalent maximum yields (16 and 22%) whereas W99A, W99F and H98AW99A mutations lead to 1.5 to 2.4-fold less transglycosylation product (9-11%) with kinetics much longer (Fig. 3B). Surprisingly, the double H98FW99F mutation allows the accumulation of more transglycosylation product (16% max yield) than the three latter ones. Visualizing the donor conversion-dependent plot of the transglycosylation yield, which constitutes a standardized view of the H/T partition, one can notice three main profiles: *(i)* *Tx*Abf and H98F for which 19 and 20% of D-Gal*f* went to transglycosylation product when 50% of *p*NP-β-D-Gal*f* was consumed, *(ii)* H98A and H98FW99F (15 and 14%, respectively), and *(iii)* W99A, W99F and H98AW99A (between 6 and 9%). These results

would tend to show that H98 and W99 mutations do impact transglycosylation and secondary hydrolysis rate since different transglycosylation profiles are obtained for a given global activity (donor consumption). Moreover, as a control experiment, to ensure that the diminished secondary hydrolysis was not due to enzyme denaturation, a thermostability study on *Tx*Abf and the six β2α2 loop mutants thereof was performed. Indeed, the different enzymatic behaviors were not due to enzyme inactivation, since all the enzymes remained sufficiently active under operating conditions (the remaining hydrolytic activity was superior to 85% after 6 h of incubation at 60 °C, data not shown).

A



B

Enzyme	Maximum yield (%)[a]	Time (min)
*Tx*Abf	19	41
H98A	16	60
H98F	22	41
W99A	9	130
W99F	10	107
H98AW99A	11	360-400[b]
H98FW99F	16	102

[a] Absolute error is estimated to be inferior to 2%.

[b] Time for maximum yield could not be determined accurately due to flat profile.

Fig. 3. Roles of H98 and W99 in *Tx*Abf-catalyzed transglycosylation (A) Transglycosylation product (Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p*) evolution, inferred from ^{1}H-NMR integration signals, as a function of donor conversion for wild-type *Tx*Abf and mutants thereof (2 µM each): wild-type (●), H98A (□), H98F (■), W99A(△), W99F (▲), H98AW99A (◇) and H98FW99F (◆). Assays were conducted at 60 °C in D_2O with *p*NP-β-D-Gal*f* and Bn-α-D-Xyl*p* with donor/acceptor ratio of 1:1 (5 mM each). Yields were calculated as a percentage of the initial donor quantity. (B) Maximal Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p* product yields and corresponding time.

In the perspective of probing the positioning of the β-D-benzyl unit in +2, the behavior of H98A (+2' subsite), which slightly decreased the transfer rate (Fig. 3) using Bn-α-D-Xyl*p* acceptor, was studied with Bn-β-D-Xyl*p* acceptor. It resulted that the transglycosylation yield was not significantly modified (Fig. SI 1). Nevertheless, the autocondensation activity is divided by more than 2-fold. These results would tend to favor the hypothesis of a positioning of the β-D-benzyl aglycon in +2 subsite, whereas autocondensation modification might signify that *p*NP-β-D-Gal*f* as an acceptor would lie in +1 and +2' subsites.

2.2.2. In silico *predictions on how H98 and W99 dictate the relative positions of catalytically-relevant atoms*

It appeared from the experimental approach developed hereinbefore that H98 and W99 do not

alter the regioselectivity but are clearly involved in the overall transfer efficiency of *Tx*Abf (i.e. transglycosylation and autocondensation). Since three main behaviors were observed, one representative of each was chosen, namely *Tx*Abf, H98A and H98AW99A, to provide theoretical explanations of these phenotypes through virtual docking of the transglycosylation product Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p* in their active site. For H98A, the product best positioning led to a binding energy mean value of -5.92 kcal.mol^{-1} (Fig. 4B), the α-D-benzyl aglycon being located in the +2' subsite as for the wild-type. These docking experimentations show that the anomeric carbon of β-D-Gal*f* unit is moving away from the catalytic nucleophile E298 (4.85 compared to 3.19 Å for the wild-type) as does the interglycosidic oxygen from the acid/base residue E176 (3.52 *vs* 2.73 Å, respectively). Alternatively, docking calculations suggested that the α-D-benzyl could be located in +1 subsite, the aromatic ring stacking against W248 side chain constitutive of +1 wall, with a calculated mean value for binding energy of -6.48 kcal.mol^{-1} (Fig. 4C).

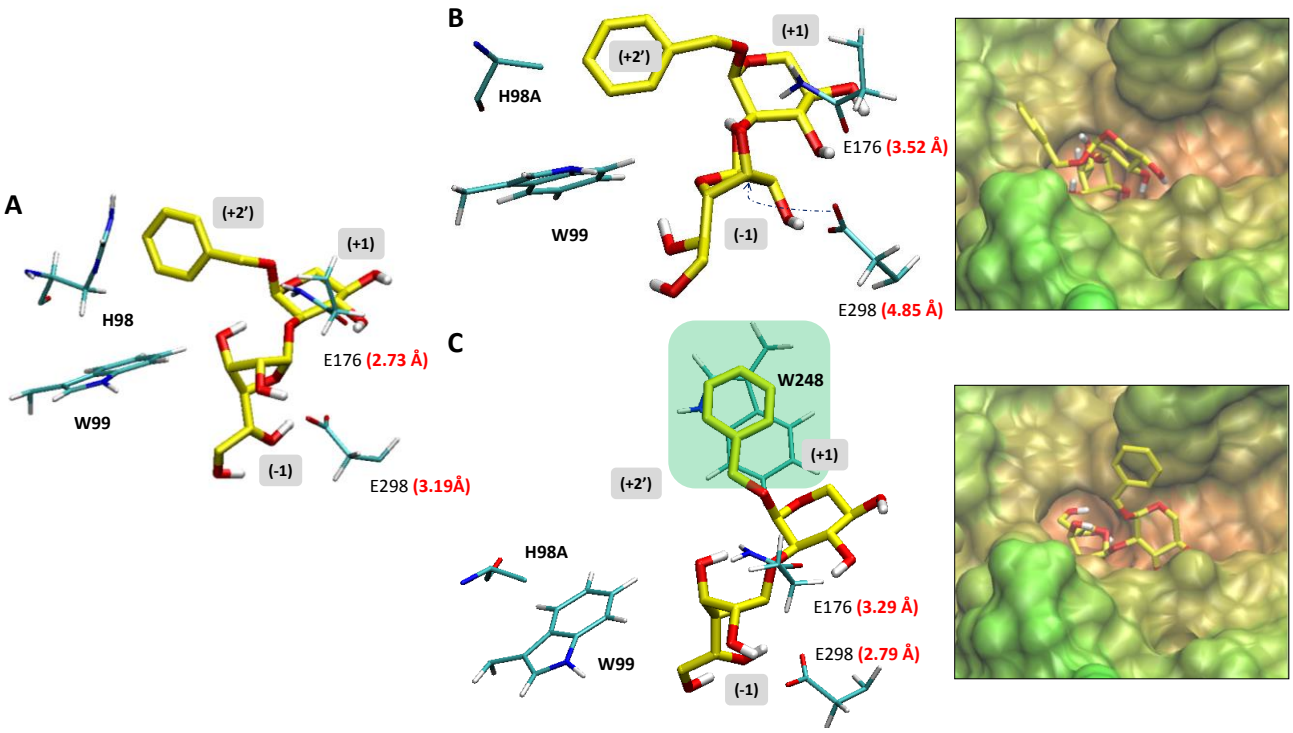


Fig. 4. Best docking solutions of Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p* into active sites of (A) *Tx*Abf and (B-C) H98A mutant. Two alternative binding modes are suggested by docking analyses for H98A mutant, either (B) the α-D-benzyl is located in +2' subsite or (C) the benzyl ring stacks with W248 constitutive of +1 wall. A mean binding energy value of -5.92 and-6.48 kcal.mol^{-1} are calculated for case (B) and (C) respectively, compared to -5.79 kcal.mol^{-1} for *Tx*Abf. The anomeric carbon of the β-D-Gal*f* unit is moving away from the nucleophile E298 (4.85 *vs* 3.19 Å for H98A mutant compared to the wild-type) as does the interglycosidic oxygen from the acid/base E176 (3.52 *vs* 2.73 Å, respectively). Top views of (B) and (C) are provided on the right hand side.

Docking experimentations realized on H98AW99A mutant showed an important remodeling of the active site (Fig. SI 2) revealing a significant displacement of the β-D-Gal*f* unit into the -1 subsite moving away from the catalytic amino acids. The anomeric carbon is then located between 4 to 6 Å

from the nucleophile and the interglycosidic oxygen between 3 to 4.5 Å from the acid/base.

2.3. β-D-anomers positioning within +1 and +2 subsites

Unlike the α-D-anomer acceptor proposed to occupy subsites +1 and +2', the premise is that with a β-D-anomery the acceptor is likely to lie from the +1 toward the +2 subsite. On that account, N216 position, constitutive of subsite +2, was submitted to site-saturation mutagenesis. In addition, W248 and W302, thought to interact with the D-Xyl*p* unit located in +1 subsite, as suggested by a basic observation of the 3-D structure (Paës et al., 2008), were also submitted to site-saturation mutagenesis. Because this +1 D-Xyl*p* unit performs the nucleophilic attack onto the anomeric carbon of the enzyme-bound α-D-Gal*f* moiety, such mutagenesis work may provide additional knowledge on the control of regioselectivity during the deglycosylation step.

The analysis of the conservation of amino acids N216, W248 and W302 in a GH51 family subset composed of 488 sequences of bacterial origin (Fig. SI 3) reveals three different features: *(i)* N216 appears as semi-conserved residue, displaying 25% identity and substituted by serine, with related physicochemical properties, in 54% of sequences, *(ii)* W248 is a poorly conserved residue (17% identity) that depict natural amino acid biodiversity, and *(iii)* W312 that is a highly conserved position (98% identity).

As a preliminary screening, the crude extracts of the nineteen W248 mutants were assessed for their transgalactofuranosylation potency on Bn-β-D-Xyl*p* as acceptor. It was observed that W248 substitutions had variable effects but, globally, most of them led to a decrease in overall D-Gal*f* transfer yields, except W248F, I and K mutations that displayed similar global transfer yields (8, 14 and 11%) relative to *Tx*Abf (11 %) with unaltered regioselectivity (Fig. SI 4). Interestingly, W248S displayed a regiospecific transglycosylation potency for β-(1,2) linkage formation, though its yield was significantly decreased (4%). All other W248 mutants displayed too low transglycosylation yields to be retained for further studies. Considering this, the transglycosylation behavior of W248F, I, K and S mutants was validated after full purification (Fig. SI 4D-G). This confirmed the regiospecificity created by the insertion of W248S mutation, making it interesting for structure-function studies, although not suitable as such for chemo-enzymatic synthesis due to the low yield provided by this enzyme.

In addition to W248 position, N216 and W302 were also investigated through NNS libraries screening. The measure of activity ratio (denoted as R), for each individual mutant, in presence or absence of acceptor (XOS), is meant to reflect an impact on the rate-limiting deglycosylation step (Part IV), and allows therefore to assess indirectly these libraries with respect to an eventual alteration of acceptor recognition. As a result, the W302 NNS library screening did not reveal any

activity improvement or inhibition alleviation in presence of acceptor. This result could be related to the highly conserved nature of this residue that might be considered has a cold spot for the targeted activity. Likewise, an NNS library was designed for the N216 position. Among the few active mutants (32%, residual stop codon) one of them (N216W) displayed an increased ratio (R = 1.03, relative to 0.39 for *Tx*Abf) in the presence of XOS acceptor suggesting an alteration of acceptor recognition and depicting either inhibition alleviation or transglycosylation increase (Arab-Jaziri et al., 2014) (SI A3). To further analyze this result, the profile of N216W mutant was analyzed on pure xylobiose or xylotriose (contained in XOS) and revealed a regiospecific feature when *p*NP-α-L-Ara*f* was used as a donor (Fig. SI 5), though transglycosylation yields are slightly lower. Following the identification of N216W in convenient screening conditions (i.e. with "usual" *p*NP-α-L-Ara*f* donor substrate, relative to *p*NP-β-D-Gal*f*), its behavior was then assayed with *p*NP-β-D-Gal*f* as a donor and synthetic (Bn-β-D-Xyl*p*, Fig. SI 4C) or natural (xylobiose and xylotriose, Fig. 5A-B) acceptors. It was shown that N216W substitution also creates a regiospecificity for (1,2) linkage formation, without altering the transglycosylation yield of the conserved product.

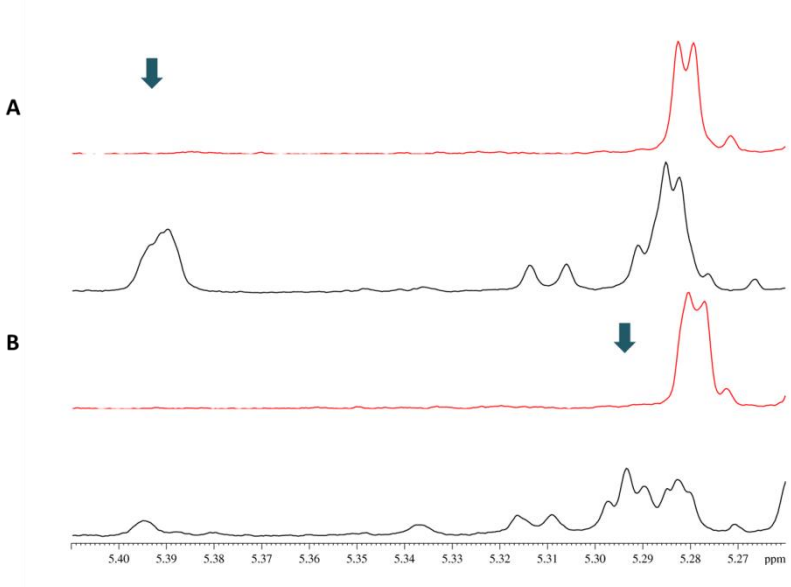


Fig. 5. ^{1}H-NMR spectra showing that N216W mutation (red line) introduces a regiospecific behavior as opposed to the wild-type *Tx*Abf (black line) for reactions with *p*NP-β-D-Gal*f* as donor and xylotriose (A) or xylobiose (B) as acceptors. N216W mutation prevents the formation (blue arrow) of (1,3) linkages: (A), putative X*M*^{3}X at 5.40 ppm and (B), putative *M*^{3}X at 5.29 ppm, with *M* defined as β-D-Gal*f* (Fauré et al., 2009), and assigned by analogy with the known chemical shifts of the L-Ara*f* unit of XA^{3}X or A^{3}X).

Furthermore, N216W capacity to transgalactofuranosylate the Bn-α-D-Xyl*p* was assayed. This revealed the presence of an additional transglycosylation product, that is likely to correspond to another regioisomer for which no standard is available (probably Bn α-D-Gal*f*-(1,3)-α-D-Xyl*p*) and reveals that the N216W-induced regiospecificity can be lost via the employment of the α-D-benzyl containing acceptor.

3. Discussion

3.1. On the acceptor α-D-anomery control on regioselectivity and role of H98 and W99

Anomeric control onto regioselectivity is a phenomenon that has already been reported among GHs and the present study provides an example for a GH51 furanose-acting enzyme. As a matter of fact, an important impact of the anomery was observed since the α-D-anomer of the acceptor led to a strict *O*-2 regiospecific transglycosylation catalyzed by *Tx*Abf whereas a mixture of (1,2)- and (1,3)-linked products is obtained using acceptors with β-D-anomery. The original contribution of this study lies in the fact that such observations were then integrated into enzyme structure-function study. Unlike the GH51 Abf from *Clostridium thermocellum* for which such drastic anomeric control was not noticed (Chlubnová et al., 2014), *Tx*Abf possesses an original active site topology relative to hitherto crystallized GH51 Abfs (Lagaert et al., 2011; Hovel et al., 2003; Taylor et al., 2006; Im et al., 2012; Souza et al., 2011) which might explain this feature. Indeed, the donor subsite is constituted of a pocket (-1 subsite) located beneath the middle of a groove which displays aromatic residues defining the acceptor subsites (Paës et al., 2008). Two +2 subsites, namely +2 and +2', flank the +1, and allow therefore a disaccharide acceptor to lie in +1 and +2 or +1 and +2' subsites. Furthermore, *Tx*Abf displays an unusual mobile β2α2 loop, essential for catalysis (Paës et al., 2008), since contributing to the dynamic formation of catalytically functional active site (Arab-Jaziri et al., 2012). Noticeably, H98 and W99 are located at one end of this loop and are key elements of the definition of the +2' subsite that is thought to be a unique feature of *Tx*Abf. It was previously noted that the employment of the synthetic Bn-α-D-Xyl*p* acceptor led to a regio-specific ((1,2)-linkage) transglycosylating behavior of *Tx*Abf (Rémond et al., 2005). This rather uncommon phenotype, for an *exo*-acting enzyme displaying a quite unconstrained acceptor subsites topology, was the opportunity to build a study aiming at establishing relationships between structure, regioselectivity and transglycosylation activity. In this frame, docking experiments revealed that the only-formed (1,2)-regioisomer obtained with this α-D-anomer acceptor would be the more favorable in terms of binding energy and might involve stacking interactions between the benzyl aglycon of Bn-α-D-Xyl*p* and the β2α2 loop-related H98. To further investigate this prediction, H98 mutation into alanine or phenylalanine revealed a slight impact on the transglycosylation potency of the enzyme, while the (1,2)-regiospecificity was kept. On the one hand, H98F led to a slightly increased transglycosylation yield, which could be consistent with equivalent or better stacking interactions between the phenyl ring of phenylalanine and benzyl group of the acceptor. Regarding H98A mutation, a slight decrease in transglycosylation yield was observed since this position is a major determinant of subsite +2', which is rather consistent with the location of the α-D-Bn into the +2' subsite. *In silico* predictions suggested that in the latter case the α-D-Bn could stack against the tryptophan W248, constituting the +1 subsite wall. This alternative

positioning could explain the limited impact of H98A mutation on the transfer activity. Interestingly, the role of H98 in acceptor docking was also visualized through a kind of compensatory effect of H98F mutation when W99 is mutated into phenylalanine since H98FW99F transglycosylation activity is almost restored to the wild-type level, in terms of yield, as opposed to the single W99F mutant. It is worth mentioning that H98-W99 pair was previously reported to have similar impacts on the catalytic efficiency (k_{cat}/K_M) on *p*NP-β-D-Gal*f* leading to approximately 50% of the wild-type value (0.557 $s^{-1}.mM^{-1}$) (Arab-Jaziri et al., 2012). In spite of this apparent similarity clear-cut differences do exist between these both residues. In contrast with H98 aforementioned features, the mutation of W99, which is a highly conserved residue among GH51 Abfs, leads to low transglycosylation yields, regardless of the mutation introduced (alanine or phenylalanine). This suggests that W99 allow a suitable donor orientation in -1 subsite as demonstrated previously on *p*NP-α-L-Ara*f* via STD NMR experiments (Arab-Jaziri et al., 2012) and consequently alter the interaction network when mutated. Such modifications would not be favorable for the transglycosylation event as well as for hydrolysis.

For the intermediary case, represented by H98A, theoretical calculations showed this mutation would make the H-1-(D-Gal*f*) unit of the transglycosylation product moving significantly away from the nucleophile catalyst. As a consequence, the nucleophilic attack of the anomeric carbon and the concomitant acid/base catalyst E176-mediated protonation of the interglycosidic oxygen are probably less efficient. However such observation is not translated by a modification of the H/T partition but more in a modification of the overall activity. Regarding the more unfavourable case, H98AW99A double mutant, the transglycosylation profile similarity with W99A confirms the preponderant functional role of W99. A lack of steric hindrance due to both H98A and W99A mutations must lead to a severe deconstruction of the interactions' network in the -1 subsite provoking a great flexibility of sugar units into the active site as suggested by docking experiments and explaining the low hydrolytic (Arab-Jaziri et al., 2012) and transglycosylating activities of these mutants. Furthermore, the predicted -1 topology deconstruction is also consistent with the low secondary hydrolysis rate displayed by this double mutant as observed from time-course transglycosylation yield monitoring (not shown).

By way of comparison with other GH51 Abfs, available co-crystal structures show that the catalytic nucleophile is located at 3.6 and 3.2 Å of the anomeric carbon of A^3X for *Ct*Araf51 (Taylor et al., 2006) and A^3 for *Gs*Abf (Hovel et al., 2003), respectively. These values are in the same range than the measured distance (3.18 Å) between *Tx*Abf-E298 and the C-1(α-L-Ara*f*) of XA^3XX (Paës et al., 2008) or than the predicted value (3.19 Å) between *Tx*Abf-E298 and the C-1 (β-D-Gal*f*) of Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p*. This supports the fact that a large distance of 4.85 Å (for *Tx*Abf-H98A) is effectively an inadequate value for nucleophilic attack to occur properly in Abfs enzymes.

Overall, it appears from this analysis that the H98-W99 pair might play an essential role in transglycosylation efficiency of acceptor with α-D-anomery, being putatively assigned to play the role of the 'sugar tongs' (Fieulaine et al., 2005) or 'gate keepers' (Yin et al., 2009). However, these functionally-relevant residues are not a determinant of regioselectivity. This may highlight the fact that both features are driven in part by energy binding of the acceptor but that positioning of acceptor through steric hindrances, introduced by α-D-benzyl aglycon in the present case, represents also a main lever of regioselectivity.

3.2. Acceptors with β-D-anomery offer relaxed regioselectivity

Unlike transglycosylation reactions carried out with Bn-α-D-Xyl*p*, it is noteworthy that for all acceptors with β-D-anomery, Bn-β-D-Xyl*p* and xylo-oligosaccharides, the same regioselectivity pattern was obtained, with approximately (depending on the reaction time) a regioisomer ratio (*O*-2:*O*-3) of 2:1, when *Tx*Abf was used as a biocatalyst. This should lead us to the conclusion that these acceptors are at least orientated in the same wise (i.e. +1 toward +2). The fact that the introduction of N216W mutation does not permit the formation of (1,3)-linkage clearly support that the +2 subsite is normally occupied in the case of *Tx*Abf when the acceptor docks for productive (1,3) glycosidic bond formation (Table 2). Regarding the (1,2)-linkage formation, a doubt is still present with respect to the (+1 toward +2) positioning of β-D-acceptors since the β-D-benzyl aglycon of Bn-β-D-Xyl*p* or alternatively the reducing end unit of xylobiose could lie in +2' subsite and would not be impacted by N216W mutation. However, two experimental results are in disagreement with this. First, a (1,2)-linkage on the non-reducing end of xylotriose can only occur if the acceptor lies between +1 and +3 subsites, since there is no evidence of +3' subsite based on 3-D structure when the β2α2 loop adopt a closed conformation (i.e. catalytically-operational state). Second, we demonstrated that H98A (+2' subsite) did not modify significantly the partition between *O*-2 and *O*-3 regioisomers when Bn-β-D-Xyl*p* was used as acceptor, validating the hypothesis of a positioning of β-D-anomer acceptors in +1 toward +2 wise for both regioisomers formation. Consistently, docking experiments predict that the β-D-benzyl unit would be located in the +2 subsite, with binding energy mean values getting closer to that calculated in the case of the natural tetrasaccharide XA^3X.

Besides, it is worth noting that calculated absolute binding energies for transglycosylation products with β-D-anomery are significantly higher (2 to 4-fold) than those with α-D-anomery. It is tempting to relate this computational result to the observation that transglycosylation yields are globally lower for the former than for the latter. A better affinity for the transglycosylation product is likely to induce a higher secondary hydrolysis rate, decreasing in turn the apparent transglycosylation yield. Conversely, the comparison of predicted absolute energies between regioisomers would tend to show that these binding energies are not critical determinants of product partition. From this, it

can be postulated that the regioisomer partition would be defined at the deglycosylation step, during which the acceptor attacking hydroxyl may display different reactivity depending on its position on the sugar ring (i.e. *O*-2 > *O*-3).

Table 1. Comparison of acceptor anomery on regioselectivity and impact of N216W mutation

Enzyme	Donor	Acceptor	Regiospecificity?	Conclusion
*Tx*Abf	*p*NP-β-D-Gal*f*	Bn-α-D-Xyl*p*	Yes - (1,2)	β-linkage: more flexibility in acceptor subsite
		Bn-β-D-Xyl*p*	No - (1,2) and (1,3)[a]	
		xylobiose	No - (1,2) and (1,3)[a]	
		xylotriose	No - (1,2) and (1,3)[a]	
N216W (+2 subsite)		Bn-α-D-Xyl*p*	No	Impact on +1
		Bn-β-D-Xyl*p*	Yes	Loss of flexibility
		xylobiose	Yes - (1,2)[a]	
		xylotriose	Yes - (1,2)[a]	

[a] Two regioisomers are clearly observed but their respective identities remain to be assigned.

More generally, this study provides also additional structure-function knowledge related to molecular evolution. Indeed, N216W is a substitution that never occurs among GH51 Abf family, pointing out the fact that most of Abfs are likely to hydrolyse (1,3)-linkages of mono-arabinofuranosylated D-Xyl*p* moieties in arabinoxylo-oliogosaccharides, as a vital function for microorganisms producing these enzymes since this branching is the most frequent (Saulnier et al., 2007).

4. Conclusions

In the present work, we demonstrated that the regioselectivity of the GH51 *Tx*Abf is subject to the acceptor anomeric control phenomenon. We revealed that the original active site topology of *Tx*Abf actually allows the acceptor to adopt an inverted positioning in acceptor subsites depending on its anomery. Noticeably, this 'orientation switch' was proposed to involve stacking interactions with aromatic residues (H98 or W248). Furthermore, in the course of this study H98 and W99 residues located at one end of the functional β2α2 mobile loop proved to be involved in the transglycosylation potency of *Tx*Abf, although not affecting the α-benzyl-driven anomeric control of

regioselectivity. This might underline the dissociated nature of both features. Besides, thanks to acceptor subsites engineering, and somewhat as a result of serendipity, this study provided us a new regiospecific biocatalyst for selective branching at *O*-2 position on xylo-oligosaccharides, which might be applied to other Abfs, being given the 'zero-occurrence' of N216W substitution. Since then, it is worth mentioning that N216W (Parts III and IV) was coupled to improved transglycosylating mutants issued from a directed evolution process allowing efficient regiospecific furanosyl transfer, leading to tailor-made D-galactofurano- or L-arabinofuranoxylo-oligosaccharides.

5. Experimental section

5.1. Substrates and chemicals

Unless otherwise stated, routine experimental work was performed using chemicals purchased mainly from Sigma-Aldrich (Illkirch, France) and molecular biology reagents purchased from New England Biolabs (Evry, France). The donor substrate *p*NP-α-L-Ara*f* was purchased from CarboSynth (Compton, U.K.). Xylotriose and XOS, used as acceptors, were purchased from Wako Chemicals GmbH (Neuss, Germany). Xylobiose was prepared in-house as previously described (Vrsanská et al., 2008). The *para*-nitrophenyl β-D-galactofuranoside (*p*NP-β-D-Gal*f*), benzyl α-D-xylopyranoside (Bn-α-D-Xyl*p*) and benzyl β-D-xylopyranoside (Bn-β-D-Xyl*p*) are synthetic substrates, prepared in-house using published protocols (Ballou et al., 1951; Rémond et al., 2004; Rémond et al., 2005).

5.2. Mutagenesis, protein expression, and purification

As previously described (Arab-Jaziri et al., 2012) *Tx*Abf mutants H98A, H98F, W99A, W99F, H98AW99A and H98FW99F were obtained via *in vitro* mutagenesis using the QuikChange Site-Directed Mutagenesis Kit (Stratagen). Regarding NNS libraries construction, codons corresponding to positions W302 and N216 were first transformed into STOP codon (TAG) serving as DNA template for the NNS mutagenesis. By this way clones transformed with residual matrix DNA will express an inactive form of *Tx*Abf and will be easily discarded during the screening step. For this the following primers (5'-3') were used:

W302Stop	C GAA TGG GGC ACG **TAG** TAT GAC GTC GAA CC
W302X	C GAA TGG GGC ACG NNS TAT GAC GTC GAA CC
N216Stop	T GCG TGC GGC GCG **TAG** ACG GCC GAC TAC CA
N216X	T GCG TGC GGC GCG NNS ACG GCC GAC TAC CA

The tryptophan residue present at position 248 (NNN = TGG) in the *Tx*Abf-encoding sequence (GenBank accession number CAA76421.2) was replaced by the 19 other amino acids using the following primers (underline letters represents mutated bases in NNN = TGG codon) for W248A (GCG), W248C (TGC), W248D (GAT), W248E (GAG), W248F (TTC), W248G (GGG), W248H (CAT),

W248I (ATC), W248K (AAA), W248L (CTG), W248M (ATG), W248P (CCG), W248Q (CAG), W248R (CGC), W248S (TCG), W248T (ACG), W248V (GTG), W248Y (TAT):

W248X 5'- CC GTT CCG GGC CCC **NNN** GAG AAG AAA GGA CC -3'

To produce recombinant *Tx*Abf-$(His)_6$ and mutants thereof, a standard protocol for the IPTG-induced heterologous expression of proteins in *E. coli* (BL21DE3) and subsequent extraction and purification was employed (Bissaro et al., 2014). After purification, the samples of *Tx*Abf-$(His)_6$ and mutants were exhaustively dialysed against Tris-HCl 20 mM pH 8 and were filtered and concentrated using and Amicon® Ultra 30 kDa device (Millipore). Absorption coefficients and molar masses were determined theoretically from the primary structure of each enzyme with ProtParam software from ExPASy website (http://www.expasy.org).

5.3. Hydrolytic activity assessment

Hydrolytic activities of *Tx*Abf and mutants thereof were determined as previously described (Bissaro et al., 2014), using a discontinuous enzyme assay, based on *p*NP release from the donor substrate *p*NP-α-L-Ara*f* or *p*NP-β-D-Gal*f*. The thermostability at 60 °C of *Tx*Abf and β2α2 loop mutants was determined through an assessment of their respective residual activity on *p*NP-α-L-Ara*f* at 60 °C, in sodium acetate buffer (50 mM, pH 5.8), by sampling incubated enzyme each hour over 6 hours, with an additional final point at t = 24h.

5.4. NNS libraries screening

For a one site-saturation NNS library, 96 clones have to be screened to cover 95% of the possible sequences according to equation $F = (1-e^{L/V})$, where F denotes the fractional completeness of the library, L the total number of clones of the library and V the number of possible sequences (Patrick et al., 2003). Indeed, since the NNS approach leads to 32 possible codons it follows that there are $V = 32^n$ possible sequences, with n = 1 position (W302X or N216X). Screening 95% of all the possible sequences implies that L = -V.ln(0.05) = 3.0 V = 3 x 32 = 96. Therefore, after transformation of *E.coli* BL21DE3 star with NNS libraries that were spread on selective (kanamycin, 40 $\mu g.mL^{-1}$) LB-agar plates, 95 clones were randomly selected and corresponding enzyme pre-purified crude extracts were prepared in microtiter plate format and submitted to automated kinetic assay as described elsewhere (Part IV). The activity assay consisted in measuring the consumption rate (*p*NP release) of the donor substrate *p*NP-α-L-Ara*f* (5 mM) in presence or absence of acceptor, a mixture of XOS (11.3 $g.L^{-1}$, equivalent to 10.8 mM of xylobiose), providing thus an activity ratio, $R = V_{i(acceptor+donor)}/V_{i(donor)}$. For this, reactions were carried out at 45 °C, in sodium acetate buffer (50 mM, pH 5.8) and enzyme crude extracts were diluted 100-fold before initiation of the reaction ($V_{added\ enzyme}$ = 1/10 of $V_{reaction}$).

5.5. Monitoring transglycosylation catalysed by *Tx*Abf and mutants

To investigate the ability of the different enzymes to perform transglycosylation, two techniques were employed. First, analytical thin-layer chromatography (TLC - silica gel $60F_{254}$ precoated plates, Merck) was used to monitor reactions performed at optimal temperature, *p*NP-β-D-Gal*f* (5 mM) and Bn-α-D-Xyl*p* (5 mM). Reactions were stopped when the donor was fully consumed. TLC was achieved using an ethyl acetate/acetic acid/water solution (7/2/2, v/v/v) as mobile phase and the migration profile was revealed using a UV lamp, followed by soaking in a orcinol solution (1 g in sulphuric acid/ethanol/water solution, 3/72.5/22.5, v/v/v) and image development at 300 °C. For finer analysis, reactions were monitored by collecting ^{1}H NMR spectra using a Bruker Avance 700 or 500 spectrometers. The reference used was water residual peak, calibrated at δ = 4.40 ppm for T = 333 K (Gottlieb et al., 1997). Transglycosylation reactions were performed in a 600 µL D_2O final volume with donor/acceptor ratio of 1:1 (5 mM each) and fixed amount of enzymes (30 µM, 15 µM or 250 nM for mutants 6, 1364 and 1753, respectively or 1 nM for *Tx*Abf wild-type). ^{1}H-NMR spectra were accumulated continuously in 4.75 min (32 scans) during 4 to 15 hours depending on mutants. For transglycosylation activities with Bn-β-D-Xyl*p* acceptor the time-course ^{1}H-NMR monitoring in D_2O was not feasible since the anomeric signals of transglycosylation products could not be isolated from the one of hydrolysis products (Table SI 1). Therefore, reactions were monitored offline by sampling 200 µL of 1 mL total reaction (5 times over 140 min), followed by enzyme inactivation ($N_{2,liq}$/ 95°C, 10 min/ $N_{2,liq}$), freeze drying and resuspension in 200 µL MeOD to realize ^{1}H-NMR analysis at 25 °C (Figs.2 and 5).

Data processing was performed as previously described (Bissaro et al., 2014). Briefly, the time-dependent evolution of donor (*p*NP-β-D-Gal*f*) and acceptor (Bn-α-D-Xyl*p*, Bn-β-D-Xyl*p*, xylobiose or xylotriose) concentrations were quantified by integrating the relative anomeric proton signals (Table SI 1). Molar balances, based on acceptor and donor consumption, were used to convert the transglycosylation product signal integral into concentration, and to evaluate concentrations of hydrolysis products, respectively. From this, the donor substrate conversion rate (X in %) was calculated as the mean of three distinct signals; those of the *ortho* and *meta* aromatic protons (*p*NP) and that of the anomeric proton H-1-(β-D-Gal*f*), with conversion accounting for both reaction outcomes (i.e. hydrolysis and transglycosylation), and identifying self-condensation. For all time-course NMR kinetics, the absolute error mean value on X ranged from 1 to 3%. The transglycosylation yield, $Y_{(product)}$, was determined by relative integration of product anomeric proton signals. Finally, $Y_{(product)}$ (%) was plotted against X (%) to provide the transfer rate (R_T) of the donor substrate to a given product (µmol of product/µmol of consumed substrate), a value which is independent of the reaction duration and thus suitable to provide a standardized comparison of the behavior of different

enzymes. Finally, initial rates were derived from the linear regions in which the variation of X was from 0-25% up to 0-40%, with root mean square deviations between 0.96 and 0.99.

5.6. Molecular modeling and docking

Energy minimization was performed using NAMD 2.5 program (Phillips et al., 2005). Results were visualized using VMD 2.9 software (Humphrey et al., 1996). AutoDockTools 1.5.2 (Sanner, 2005) was used to add polar hydrogens and assigned Gasteiger charges to the protein structure and ligands. Autodock 4 with Lamarckian genetic algorithm was used to simulate ligand-receptor docking (Morris et al., 1998). Docking parameters were chosen using population size of 250, number of energy evaluation of 25 000 000 and 250 runs. Docked conformations were clustered using external clustering script provided with AutoDockTools 1.5.2. Reported docking results correspond to the lowest relative binding energy (sum of intermolecular energy and torsional free-energy penalty in kcal.mol^{-1}) of the most populated cluster and RMSD of the ligand positions.

Acknowledgements

PhD fellowship of B. Bissaro was supported by the Institut National de la Recherche Agronomique (INRA) [CJS]. MetaToul (Metabolomics & Fluxomics Facitilies, Toulouse, France, www.metatoul.fr) and its staff members are gratefully acknowledged for technical support and access to NMR facility. MetaToul is part of the national infrastructure MetaboHUB (The French National infrastructure for metabolomics and fluxomics, www.metabohub.fr). MetaToul is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, the SICOVAL, the Infrastructures en Biologie Sante et Agronomie (IBiSa, France), the Centre National de la Recherche Scientifique (CNRS) and the Institut National de la Recherche Agronomique (INRA). NMR kinetics experiments were performed at the 'Integrated Screening Platform of Toulouse' (PICT, IPBS, CNRS – Université de Toulouse) using equipment funded by the French Research Ministry, CNRS, Université Paul Sabatier, the Région Midi-Pyrénées and European structural funds. We thank the ICEO facility dedicated to enzyme screening and discovery, and part of the Integrated Screening Platform of Toulouse (PICT, IBiSA) for providing access to the TECAN liquid handling equipment.

References

Ajisaka, K. & Yamamoto, Y. (2002) Control of the regioselectivity in the enzymatic syntheses of oligosaccharides using glycosidases. *Trends Glycosci. Glycotechnol.*, 14 (75), pp.1–11.

Arab-Jaziri, F., Bissaro, B., Barbe, S., Saurel, O., Débat, H., Dumon, C., Gervais, V., Milon, A., André, I., Fauré, R. & O'Donohue, M.J. (2012) Functional roles of H98 and W99 and β2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *FEBS J.*, 279 (19), pp.3598–3611.

Arab-Jaziri, F., Bissaro, B., Dion, M., Saurel, O., Harrison, D., Ferreira, F., Milon, A. & Tellier, Charles, O'Donohue, Michael J,

Fauré, R. (2013) Engineering transglycosidase activity into a GH51 α-L-arabinofuranosidase. *New Biotechnol.*, 30 (5), pp.536–544.

Arab-Jaziri, F., Bissaro, B., Tellier, C., Dion, M., Fauré, R. & O'Donohue, M.J. (2014) Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α-L-arabinofuranosidase using a directed evolution approach. *Submitted Results*.

Ballou, C.E., Roseman, S. & Link, K.P. (1951) Reductive cleavage of benzyl glycosides for relating anomeric configurations. preparation of some new benzyl pentosides. *J. Org. Chem.*, 73, pp.1140–1144.

Bissaro, B., Saurel, O., Arab-Jaziri, F., Saulnier, L., Milon, A., Tenkanen, M., Monsan, P., O'Donohue, M.J. & Fauré, R. (2014) Mutation of a pH-modulating residue in a GH51 α-L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products. *Biochim. Biophys. Acta Gen. Subj.*, 1840 (1), pp.626–636.

Chlubnova, I., Filipp, D., Spiwok, V., Dvorakova, H., Daniellou, R., Nugier-Chauvin, C., Kralova, B. & Ferrieres, V. (2010) Enzymatic synthesis of oligo-D-galactofuranosides and L-arabinofuranosides: from molecular dynamics to immunological assays. *Org. Biomol. Chem.*, 8 (9), pp.2092–2102.

Chlubnová, I., Králová, B., Dvořáková, H., Hošek, P., Spiwok, V., Filipp, D., Nugier-Chauvin, C., Daniellou, R. & Ferrières, V. (2014) The versatile enzyme Araf51 allowed efficient synthesis of rare pathogen-related β-D-galactofuranosyl-pyranoside disaccharides. *Org. Biomol. Chem.*, 12 (19), pp.3080–3089.

Crout, D.H.G., Howarth, O.W., Suddham, S., Swoboda, B.E.P., Critchley, P. & Gibson, W.T. (1991) Biotransformation in carbohydrate synthesis. N-acetylgalactosaminyl and N-acetylglucosaminyl transfer onto methyl α- and β-glucosides catalysed by the β-N-acetylhexosaminidase from *Aspergillus oryzae*. *J. Chem. Soc., Chem. Commun.*, (21), pp.1550–1551.

Crout, D.H.G. & Vic, G. (1998) Glycosidases and glycosyl transferases oligosaccharide synthesis. *Biocatal. Biotransform.*, 2, pp.98–111.

Dion, M., Nisole, A., Spangenberg, P., André, C., Glottin-Fleury, A., Mattes, R., Tellier, C. & Rabiller, C. (2001) Modulation of the regioselectivity of a *Bacillus* α-galactosidase by directed evolution. *Glycoconjugate J.*, 18 (3), pp.215–223.

Dion, M., Osanjo, G., André, C., Spangenberg, P., Rabiller, C. & Tellier, C. (2001) Identification by saturation mutagenesis of a single residue involved in the α-galactosidase AgaB regioselectivity. *Glycoconjugate J.*, 18 (6), pp.457–464.

Dumon, C., Song, L., Bozonnet, S., Fauré, R. & O'Donohue, M.J. (2012) Progress and future prospects for pentose-specific biocatalysts in biorefining. *Process Biochem.*, 47 (3), pp.346–357.

Faijes, M., Saura-Valls, M., Pérez, X., Conti, M. & Planas, A. (2006) Acceptor-dependent regioselectivity of glycosynthase reactions by *Streptomyces* E383A beta-glucosidase. *Carbohydr. Res.*, 341 (12), pp.2055–2065.

Fauré, R., Courtin, C.M., Delcour, J.A., Dumon, C., Faulds, C.B., Fincher, G.B., Fort, S., Fry, S.C., Halila, S., Kabel, M.A., Pouvreau, L., Quemener, B., Rivet, A., Saulnier, L., Schols, H.A., Driguez, H. & O'Donohue, M.J. (2009) A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls. *Aust. J. Chem.*, 62 (6), pp.533–537.

Fieulaine, S., Lunn, J.E., Borel, F. & Ferrer, J. (2005) The structure of a cyanobacterial sucrose-phosphatase reveals the sugar tongs that release free sucrose in the cell. *Plant Cell*, 17, pp.2049–2058.

Giordano, A., Tramice, A., Andreotti, G., Mollo, E. & Trincone, A. (2005) Enzymatic syntheses and selective hydrolysis of *O*-β-D-galactopyranosides using a marine mollusc β-galactosidase. *Bioorg. Med. Chem. Lett.*, 15 (1), pp.139–143.

Gottlieb, H.E., Kotlyar, V. & Nudelman, A. (1997) NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem.*, 3263 (3), pp.7512–7515.

Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. & Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 α-L-arabinofuranosidase. *EMBO J.*, 22 (19), pp.4922–4932.

Humphrey, W., Dalke, A. & Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graphics*, 14, pp.33–38.

Im, D.-H., Kimura, K., Hayasaka, F., Tanaka, T., Noguchi, M., Kobayashi, A., Shoda, S., Miyazaki, K., Wakagi, T. & Fushinobu, S. (2012) Crystal structures of glycoside hydrolase family 51 α-L-arabinofuranosidase from *Thermotoga maritima*. *Biosci., Biotechnol. Biochem.*, 76 (2), pp.423–428.

Lagaert, S., Schoepe, J., Delcour, J.A., Lavigne, R., Strelkov, S.V., Courtin, C.M., Mikkelsen, N.E., Sandgren, M. & Volckaert, G. (2011) Elucidation of the substrate specificity and protein structure of Abfb, a family 51 α-L-arabinofuranosidase from *Bifidobacterium longum*. *To be Published*, PDB 2Y2W.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, 42 (Database issue), pp.D490–D495.

Lopez, R. & Fernandez-Mayoralas, A. (1994) Enzymatic β-galactosidation of modified monosaccharides: study of the enzyme selectivity for the acceptor and its application to the synthesis of disaccharides. *J. Org. Chem*, 59 (4), pp.737–745.

Mackenzie, L.F., Wang, Q., Warren, R.A.J. & Withers, S.G. (1998) Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.*, 120 (22), pp.5583–5584.

MacManus, D.A. & Vulfson, E.N. (2000) Regioselectivity of enzymatic glycosylation of 6-*O*-acyl glycosides in supersaturated solutions. *Biotechnol. Bioeng.*, 69 (6), pp.585–590.

Malet, C. & Planas, A. (1998) From β-glucanase to β-glucansynthase: glycosyl transfer to alpha-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS letters*, 440 (1-2), pp.208–212.

Marton, Z., Tran, V., Tellier, C., Dion, M., Drone, J. & Rabiller, C. (2008) Engineering of glucoside acceptors for the regioselective synthesis of beta-(1-->3)-disaccharides with glycosynthases. *Carbohydr. Res.*, 343 (17), pp.2939–2946.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J. & Al, M.E.T. (1998) Automated

docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 19 (14), pp.1639–1662.

Ngiwsara, L., Iwai, G., Tagami, T., Sato, N., Nakai, H., Okuyama, M., Mori, H. & Kimura, A. (2012) Amino acids in conserved region II are crucial to substrate specificity, reaction velocity, and regioselectivity in the transglucosylation of Honey bee GH-13 α-glucosidases. *Biosci., Biotechnol., Biochem.*, 76 (10), pp.1967–1974.

Nilsson, K.G. (1987) A simple strategy for changing the regioselectivity of glycosidase-catalysed formation of disaccharides. *Carbohydr. Res.*, 167, pp.95–103.

Paës, G., Skov, L.K., O'Donohue, M.J., Rémond, C., Kastrup, J.S., Gajhede, M. & Mirza, O. (2008) The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry*, 47 (28), pp.7441–7451.

Patrick, W.M., Firth, A.E. & Blackburn, J.M. (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng. Des. Sel.*, 16 (6), pp.451–457.

Peltier, P., Euzen, R., Daniellou, R., Nugier-Chauvin, C. & Ferrières, V. (2008) Recent knowledge and innovations related to hexofuranosides: structure, synthesis and applications. *Carbohydr. Res.*, 343 (12), pp.1897–1923.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L. & Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26 (16), pp.1781–1802.

Prade, H., Mackenzie, L.F. & Withers, S.G. (1998) Enzymatic synthesis of disaccharides using *Agrobacterium sp.* β-glucosidase. *Carbohydr. Res.*, 305, pp.371–381.

Rémond, C., Ferchichi, M., Aubry, N., Plantier-Royon, R., Portella, C. & O'Donohue, M.J. (2002) Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α-L-arabinofuranosidase. *Tetrahedron Lett.*, 43 (52), pp.9653–9655.

Rémond, C., Plantier-Royon, R., Aubry, N., Maes, E., Bliard, C. & O'Donohue, M.J. (2004) Synthesis of pentose-containing disaccharides using a thermostable α-L-arabinofuranosidase. *Carbohydr. Res.*, 339 (11), pp.2019–2025.

Rémond, C., Plantier-Royon, R., Aubry, N. & O'Donohue, M.J. (2005) An original chemoenzymatic route for the synthesis of β-D-galactofuranosides using an α-L-arabinofuranosidase. *Carbohydr. Res.*, 340 (4), pp.637–644.

Sanner, M.F. (2005) A component-based software environment for visualizing large macromolecular assemblies. *Structure*, 13, pp.447–462.

Saulnier, L., Sado, P.-E., Branlard, G., Charmet, G. & Guillon, F. (2007) Wheat arabinoxylans: Exploiting variation in amount and composition to develop enhanced varieties. *J. Cereal Sci.*, 46 (3), pp.261–281.

Schmaltz, R.M., Hanson, S.R. & Wong, C.-H. (2011) Enzymes in the synthesis of glycoconjugates. *Chem. Rev.*, 111 (7), pp.4259–4307.

Souza, T., Santos, C.R., Souza, A.R., Oldiges, D.P., Ruller, R., Prade, R. a., Squina, F.M. & Murakami, M.T. (2011) Structure of a novel thermostable GH51 α-L-arabinofuranosidase from *Thermotoga petrophila* RKU-1. *Protein Sci.*, 20 (9), pp.1632–1637.

Stick, R. V., Stubbs, K. a. & Watts, A.G. (2004) Modifying the regioselectivity of glycosynthase reactions through changes in the acceptor. *Aust. J. Chem.*, 57 (8), pp.779–786.

Taylor, E.J., Smith, N., Turkenburg, J., D'Souza, S., Gilbert, H. & Davies, G. (2006) Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from *Clostridium thermocellum*. *Biochem. J.*, 395, pp.31–37.

Teze, D., Dion, M., Daligault, F., Tran, V., André-Miral, C. & Tellier, C. (2013) Alkoxyamino glycoside acceptors for the regioselective synthesis of oligosaccharides using glycosynthases and transglycosidases. *Bioorg. Med. Chem. Lett.*, 23 (2), pp.448–451.

Vrsanská, M., Nerinckx, W., Claeyssens, M. & Biely, P. (2008) An alternative approach for the synthesis of fluorogenic substrates of *endo*-β-(1→4)-xylanases and some applications. *Carbohydr. Res.*, 343 (3), pp.541–548.

Yin, J., Li, L., Shaw, N., Li, Y., Song, J.K., Zhang, W., Xia, C., Zhang, R., Joachimiak, A., Zhang, H.-C., Wang, L.-X., Liu, Z.-J. & Wang, P. (2009) Structural basis and catalytic mechanism for the dual functional endo-β-N-acetylglucosaminidase A. *PLoS One*, 4 (3), e4658.

Zeng, X., Yoshino, R., Murata, T., Ajisaka, K. & Usui, T. (2000) Regioselective synthesis of *p*-nitrophenyl glycosides of β-D-galactopyranosyl-disaccharides by transglycosylation with β-D-galactosidases. *Carbohydr. Res.*, 325 (2), pp.120–131.

Supporting Information of Part I: "Acceptor anomeric control of the regioselectivity and transfuranosylation in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*"

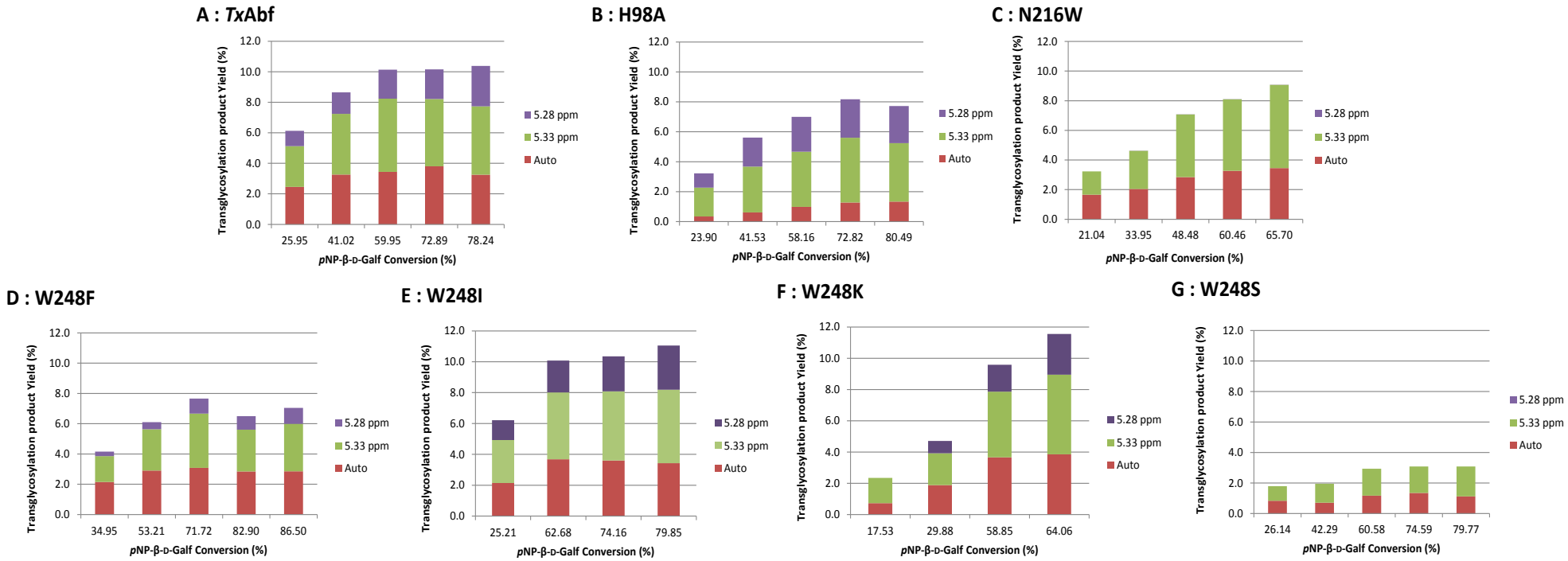


Fig. SI 1. Off-line time-course monitoring of transgalactofuranosylation and condensation products (5.33 and 5.28 ppm for (1,2) and (1,3) linkages) catalyzed by (A) *Tx*Abf, (B) H98A, (C) N216W, (D) W248F, (E) W248I, (F) W248K and (G) W248S. Reactions were performed with *p*NP-β-D-Gal*f* and Bn-β-D-Xyl*p* as donor and acceptor, respectively. Reactions were performed with donor/acceptor ratio of 1:1 (5 mM each) in buffered conditions (50 mM sodium acetate pH 5.8) at 60 °C. The reaction was followed off-line by sampling aliquot of 200 µL at t = 15, 30 60, 100 and 140 min, for subsequent ^{1}H-NMR analyses in MeOD at 25 °C.

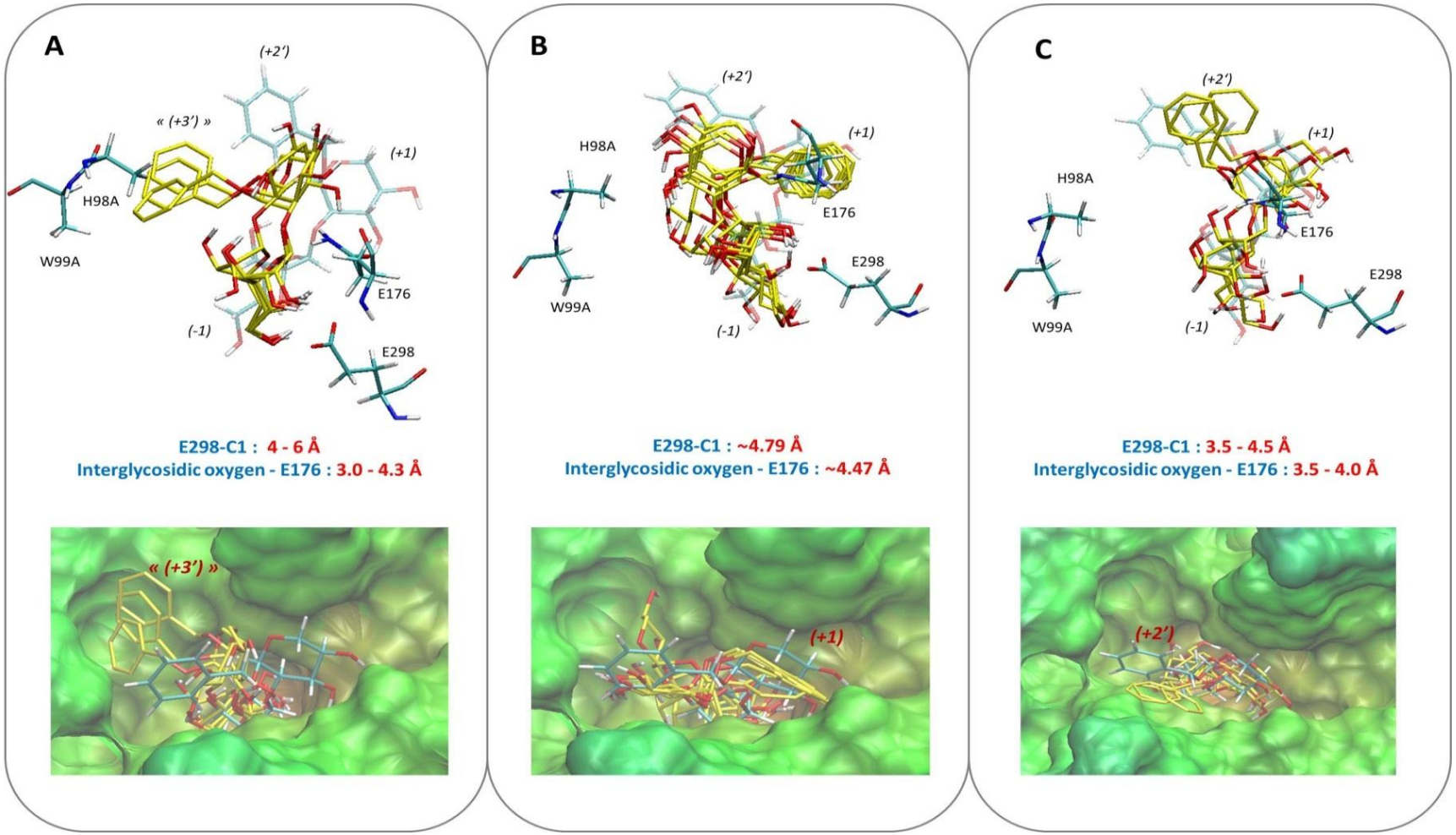
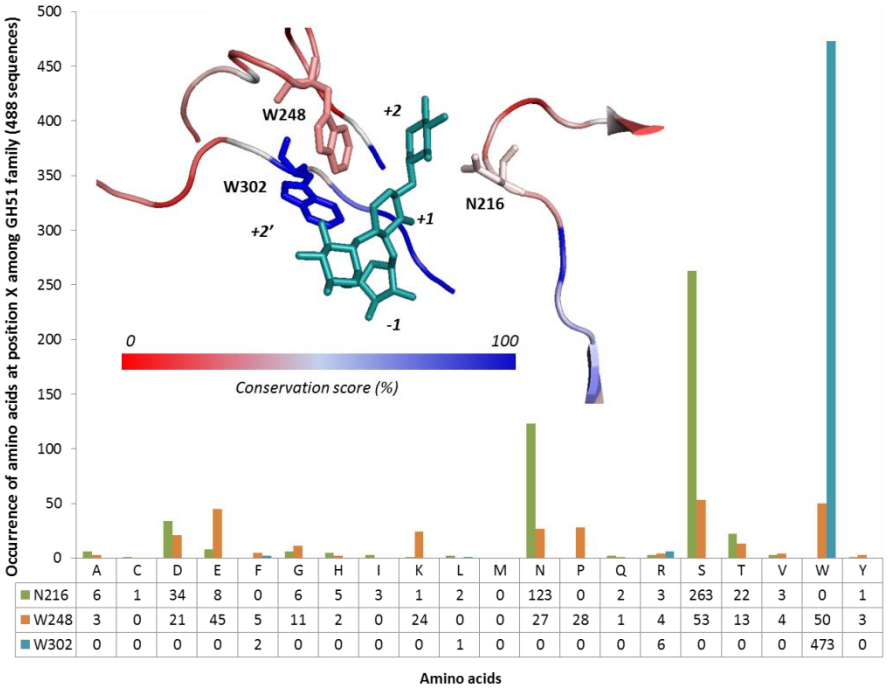


Fig. SI 2. Impact of H98AW99A mutation on the active site structure and Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p* docking in it (represented in yellow or blue when H98A or *Tx*Abf are used as receptor proteins, respectively). Three clusters were determined by Autodock 4 ranked as follow: (A) the benzyl unit is located in a pseudo +3' subsite and D-Xyl*p* moves toward the +2', (B) the benzyl unit undergoes a position switch and would be located in the +1 whereas D-Xyl*p* would be in the +2', and (C) the benzyl unit and D-Xyl*p* are respectively located in +2' and +1 subsites. This active site remodelling induce an important relocation of the β-D-Gal*f* unit in the -1 subsite which goes along a significant modification of distances between the β-D-Gal*f* anomeric carbon and the catalytic nucleophile (E298-C1) as well as between the interglycosidic oxygen and the acid/base E176, relative to values calculated with this same product for the wild-type *Tx*Abf, i.e. (E298-C1) = 3.19 Å and (interglycosidic oxygen-E176) = 2.73 Å.



	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
N216	6	1	34	8	0	6	5	3	1	2	0	123	0	2	3	263	22	3	0	1
W248	3	0	21	45	5	11	2	0	24	0	0	27	28	1	4	53	13	4	50	3
W302	0	0	0	0	2	0	0	0	0	1	0	0	0	0	6	0	0	0	473	0

Fig. SI 3. Amino acid substitutions at targeted positions (i.e. 216, 248 and 302 - *Tx*Abf numbering). The occurrence of alternative amino acid was determined by examining 488 GH51 Abf sequences.

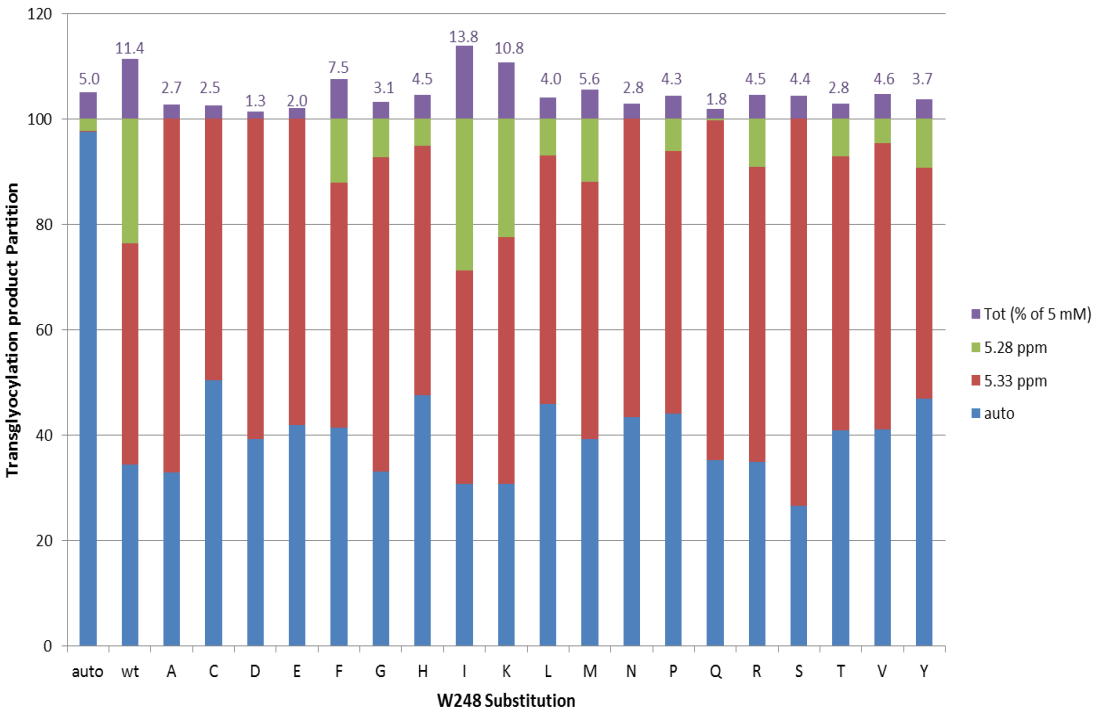


Fig. SI 4. Analysis of D-Gal*f* transfer partition by ^{1}H-NMR screening of the W248 site-saturation library. After calibration of transglycosylation reactions with 0.05 IU for each enzyme, reactions containing *p*NP-β-D-Gal*f* (5 mM) and Bn-β-D-Xyl*p* (5 mM) were sampled (200 µL) at t= 45 min and 90 min. After reaction quenching ($N_{2,\ liq.}$, 95°C, $N_{2,\ liq.}$) and lyophilisation, samples were dissolved in MeOD and analysed at 25 °C on 500 MHz spectrometer. Two transglycosylation products were observed at 5.28 and 5.33 ppm, putatively assigned to H-1 of Bn β-D-Gal*f*-(1,3)-β-D-Xyl*p* and Bn β-D-Gal*f*-(1,2)-β-D-Xyl*p*, respectively (of note their respective identities remain to be confirmed once their purification will have been achieved). Total synthesis yields are given on top of the histogram. The autocondensation reaction, realized as a control experiment where no transglycosylation product was observed, reveals that an error of 2% on integrals values has to be considered.

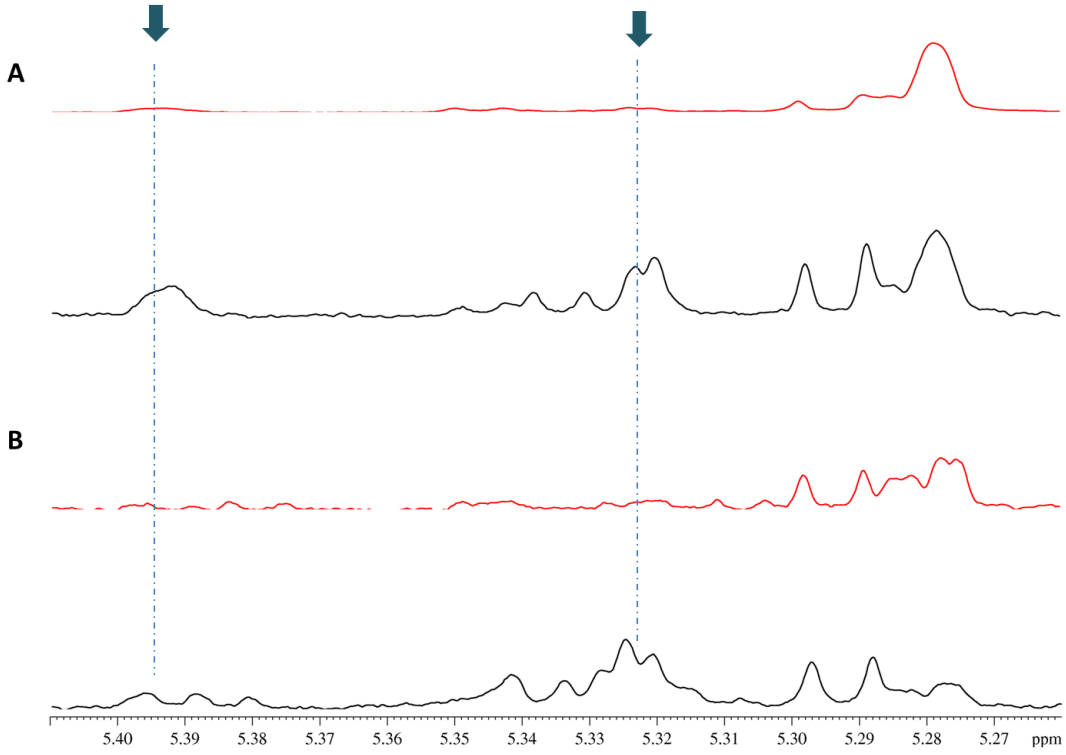


Fig. SI 5. N216W mutation (red line) makes *Tx*Abf a regiospecific enzyme as opposed to wild-type enzyme for reactions with *p*NP-α-L-Ara*f* as donor and xylotriose (A) or xylobiose (B) as acceptors. N216W mutation prevents the formation (blue arrow) of (1,3) linkages: (A) XA^{3}X at 5.40 ppm and A^{3}XX at 5.32 ppm or (B) A^{3}X at 5.32 ppm. Of note, H-1 of (1,2)-linked α-L-Ara*f* moieties display chemical shift at 5.27-5.28 ppm

Table SI 1. List of the proton signals used for the tracking of the different species/activities during the enzyme-catalyzed reactions

Activities/species monitoring	Protons	δ (ppm)[c] in D_2O (333 K)	δ (ppm)[c] in MeOD (298
Global activity	Linked *p*NP (*meta* and *ortho*)	8.26-7.25	8.21 and 7.22
	Released *p*NP (*meta* and *ortho*)	8.18-6.99	8.06 and 6.74
	H-1 of *p*NP-β-D-Gal*f*	5.82	5.66
Hydrolysis product[a]	H-1 of α-D-Gal*p*	5.25	5.14
	H-1 of β-D-Gal*f*	5.22	-
	H-1 of β-D-Gal*p*	4.57	-
Transglycosylation product[b]	H-1 of Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p*	5.09	*n.d.*
	H-1 of Bn β-D-Gal*f*-(1,2)-α-D-Xyl*p*	5.01	*n.d.*
	H-1 of Bn β-D-Gal*f*-(1,2)-β-D-Xyl*p*	5.28-5.22[d]	5.33 and 5.30[d]
	H-1 of Bn β-D-Gal*f*-(1,3)-β-D-Xyl*p*		
Acceptor	H-1 of Bn-α-D-Xyl*p*	4.97	*n.d.*
	H-1 of Bn-β-D-Xyl*p*	4.48	4.31

[a] Amongst the different D-Gal forms are not represented the linear one and the α-D-Gal*f* one since these are really minor species and thus not detected; the major one being the pyranose form with a β-anomery (β-D-Gal*p*).

[b] Autocondensation product: 5.89 and 5.86 ppm, in D_2O and MeOD, respectively.

[c] H-1 signals of transglycosylation products obtained with *p*NP-β-D-Gal*f* and Bn-β-D-Xyl*p* as acceptor are overlapping with D-Gal hydrolysis product when the analysis is performed in D_2O (whatever temperature is employed between 15 and 60 °C). Nevertheless, when the analysis is performed in MeOD as a solvent, at 25 °C, a better resolution is obtained, allowing to distinguish both regioisomers from hydrolysis product signals. This is why all experiments involving Bn-β-D-Xyl*p* as acceptor were monitored off-line with ^{1}H NMR analyses carried out in MeOD.

[d] The two observed regioisomers, putatively Bn β-D-Gal*f*-(1,2)-β-D-Xyl*p* and Bn β-D-Gal*f*-(1,3)-β-D-Xyl*p*, will be produced in batch reactions to be isolated and fully characterized in the next few weeks.

n.d.: not determined since these signals were not needed.

Part II

Finding the First Steps toward Transarabinofuranosylation

II-1. Identification of relevant 2nd Shell mutations

In the perspective of modulating the transglycosylation capability of *Tx*Abf at a molecular level, Dr. Faten Arab-Jaziri initiated a directed evolution process by creating random libraries of *Tx*Abf-encoding gene that were subsequently screened in the frame of a collaboration with Charles Tellier (Université de Nantes, France), which developed and implemented the method (Feng et al., 2005; Osanjo et al.,2007). For this, two approaches were sequentially developed; the first one consisted in screening clones only for their crippled hydrolytic activity, using the chromogenic substrate 5-bromo-4-chloro-3-indolyl α-L-arabinofuranoside (X-α-L-Ara*f*), leading to shades of blue upon hydrolysis. The light blue clones, weakly hydrolytic but not inactive (otherwise white), were selected and submitted to transglycosylation activity assessment. Among the 8 mutants that were retained (based on TLC analysis) I performed NMR-based time course monitoring to get an accurate view of the transglycosylation behavior of these. For this, a protocol and data analysis methodology had to be set up. Three mutants, denoted as 6, 1364 and 1753 were confirmed to transarabinofuranosylate the synthetic acceptor benzyl α-D-xylopyranoside (Bn-α-D-Xyl*p*) (Supporting information A2), and three putative culprit mutations: W178R, F26L and N344S were pointed out (Fig. 1). In a second place, to improve the screening methodology of transglycosylation activity a two-step procedure was implemented, also through this collaboration (Koné et al., 2009). Dr. F. Arab-Jaziri designed a second independent random library that was *in vivo* screened on solid medium, on a first plate, for diminished hydrolytic activity and, on a second plate, for a relative increase (rescue) of activity in presence natural acceptor, a mixture of xylo-oligosaccharides (XOS). The resulting 199 clones, out of 30,000, were analyzed through an *in vitro* secondary screening, in liquid medium, to assess the relative activity (ratio of released *p*NP without and with the presence of XOS acceptor) using enzyme crude extracts. Since the wild-type *Tx*Abf is inhibited by the presence of XOS, which also serves as acceptor at deglycosylation step, it makes more complicated the detection of XOS-activated clones (meant to reflect transglycosylation enhancement). This is why it turned out that no clear activation was observed but rather apparent inhibition alleviation (i.e. relative activities ~100% compared to ~38% for *Tx*Abf, in presence of 11.3 $g.mol^{-1}$ of XOS equivalent to 10.8 mM of X_2). Notwithstanding, following Faten work I carried out time-course NMR monitoring of the transglycosylation activity of the 9 most promising mutants leading to the validation of 5 of them, denoted as n° 1, 8, 169, 2032 and 2068. I isolated the mutations thought as relevant based on two criteria: *(i)* the occurrence of found substitutions among the GH51 Abfs diversity (if the substitution occurred at a significant level it was discarded) and *(ii)* the location within *Tx*Abf structure (only mutations within the 1st and 2nd shell were kept, i.e. at a maximum of 15 Å from the catalytic nucleophile E298). As a result, mutations V173D, Q316L, N344I, N344Y, I345L and L352M were retained (Fig. 1) and we showed that

mutants N344Y et L352M can constitute interesting targets for H/T modulation (Supporting Information A3, *Submitted Article*) that deserve deeper investigations, as described in the following sections of this manuscript. However, the transglycosylation potency of these mutants remained modest (< 20% of A^2XX, the main regiosiomer), making necessary a more sophisticated molecular evolution strategy to be able to design a *bona fide* 'non-Leloir' transarabinofuranosylase.

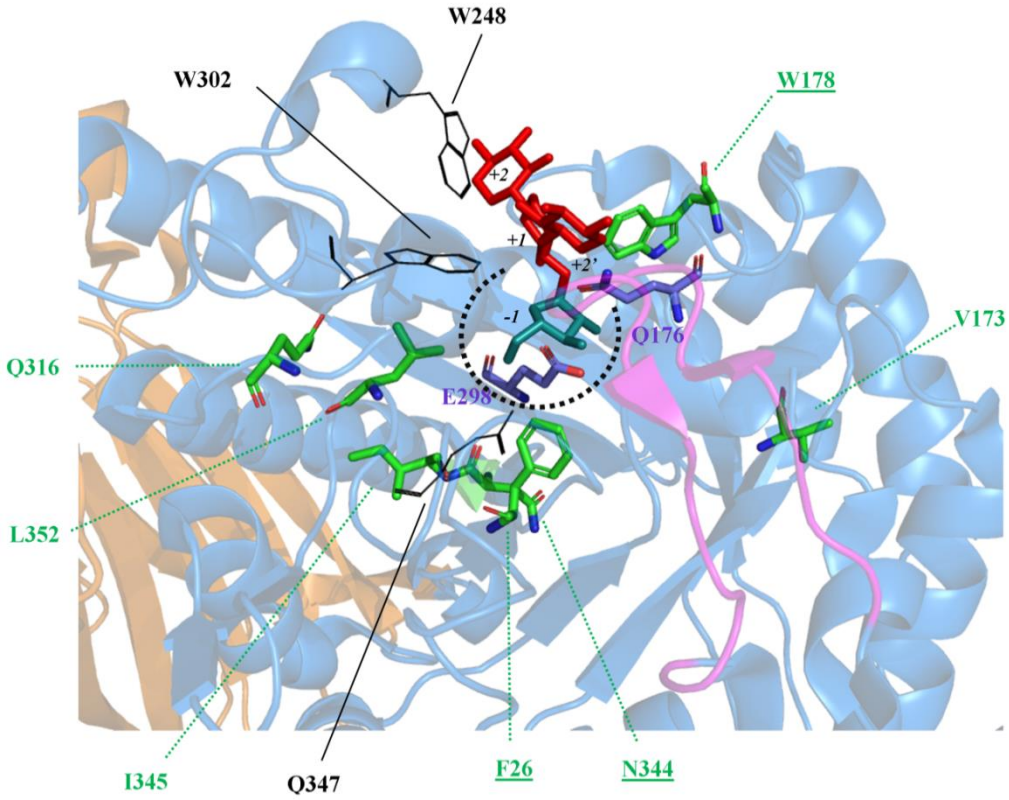


Fig. 1. Summary of or 1st and 2nd shell mutations identified in the course of the two independent rounds of random mutagenesis/screening. Mutations are represented as sticks (green) within *Tx*Abf structure (PDB 2VRQ). Underlined residues are those found in the course of the first random library screening. Amino acids involved in active site topolgy and that are likely to be impacted by mutations are represented as black lines, whereras catalytic nucleophile E298 and acid/base E176 (mutated into Q) are depicted as purple sticks. The dotted line allows to distinguish the pocket-like -1 subsite receiving the L-Ara*f* moiey of the substrate.

II-2. Mutation of a pH-modulating residue in a GH51 α-L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products

In the course of this thesis work, the position N344 was found to be mutated 5 times in different amino acids (N344I, K, S or Y) within two different libraries. As explained in the next section, this is in part why N344 role was investigated and gave rise to the present publication, available online since the 17th of October, 2013. Regarding the L352M mutation, which appeared *a posteriori* to be a more interesting protein scaffold for further molecular evolution, its mechanistic impact is described in parts III and IV.

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Mutation of a pH-modulating residue in a GH51 α-L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products☆

Bastien Bissaro [a,b,c], Olivier Saurel [a,d], Faten Arab-Jaziri [a,b,c], Luc Saulnier [e], Alain Milon [a,d], Maija Tenkanen [f], Pierre Monsan [a,b,c], Michael J. O'Donohue [a,b,c], Régis Fauré [a,b,c,*]

[a] Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France
[b] INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France
[c] CNRS, UMR5504, F-31400 Toulouse, France
[d] CNRS, IPBS UMR 5089, Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, BP 64182 Toulouse, France
[e] UR1268 Biopolymères Interactions Assemblages, INRA, 44300 Nantes, France
[f] Department of Food and Environmental Sciences, Faculty of Agriculture and Forestry, University of Helsinki, P.O. Box 27, FI-00014, Finland

ARTICLE INFO

Article history:
Received 10 June 2013
Received in revised form 23 September 2013
Accepted 4 October 2013
Available online 17 October 2013

Keywords: Pentoses/furanoses
Transglycosylation
pK_a modulation
pH-dependent inhibition
STD NMR

ABSTRACT

Background: The development of enzyme-mediated glycosynthesis using glycoside hydrolases is still an inexact science, because the underlying molecular determinants of transglycosylation are not well understood. In the framework of this challenge, this study focused on the family GH51 α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*, with the aim to understand why the mutation of position 344 provokes a significant modification of the transglycosylation/hydrolysis partition.

Methods: Detailed kinetic analysis (k_{cat}, K_M, pK_a determination and time-course NMR kinetics) and saturation transfer difference nuclear magnetic resonance spectroscopy was employed to determine the synthetic and hydrolytic ability modification induced by the redundant N344 mutation disclosed in libraries from directed evolution.

Results: The mutants N344P and N344Y displayed crippled hydrolytic abilities, and thus procured improved transglycosylation yields. This behavior was correlated with an increased pK_a of the catalytic nucleophile (E298), the pK_a of the acid/base catalyst remaining unaffected. Finally, mutations at position 344 provoked a pH-dependent product inhibition phenomenon, which is likely to be the result of a significant modification of the proton sharing network in the mutants.

Conclusions and general significance: Using a combination of biochemical and biophysical methods, we have studied *Tx*Abf-N344 mutants, thus revealing some fundamental details concerning pH modulation. Although these results concern a GH51 α-L-arabinofuranosidase, it is likely that the general lessons that can be drawn from them will be applicable to other glycoside hydrolases. Moreover, the effects of mutations at position 344 on the transglycosylation/hydrolysis partition provide clues as to how *Tx*Abf can be further engineered to obtain an efficient transfuranosidase.

Abbreviations: Abfs, α-L-arabinofuranosidases; AXOS, arabinoxylo-oligosaccharides; D-Xyl*p*, D-xylopyranosyl; FH, furanoside hydrolase; GH, glycoside hydrolase; KIE, kinetic isotope effects; L-Ara*f*, L-arabinofuranosyl; 4NTC-α-L-Ara*f*, 4-nitrochatecol α-L-arabinofuranoside; *m*NP, *meta*-nitrophenol; *o*NP, *ortho*-nitrophenol; *p*NP, *para*-nitrophenol; *p*NP-α-L-Ara*f*, *para*-nitrophenyl α-L-arabinofuranoside; R_T, transfer rate; STD NMR, Saturation Transfer Difference Nuclear Magnetic Resonance; T/H ratio, transglycosylation/hydrolysis ratio; *Tx*Abf, α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*; *Tx*Abf^t, inactivated form (E176A) of *Tx*Abf; X, donor conversion rate; Y, yield. The specific abbreviated names of different AXOS, such as A^{3}X and XA^{3}XX, were generated using the naming system developed by Fauré et al. [1].

* Corresponding author at: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, 135 Avenue de Rangueil, 31077 Toulouse cedex 4, France. Tel.: +33 5 6155 9488; fax: +33 5 6155 9400. *E-mail address:* regis.faure@insa-toulouse.fr (R. Fauré).

1. Introduction

So far, very few studies have dealt with the development of furanoside hydrolases (FHs) as glycosynthetic tools, with most work being performed on hexose and/or pyranose-acting enzymes [2–5]. This is unsurprising since, not only are FHs a relatively small subgroup of glycoside hydrolases (GHs), but also furanose chemistry is not a mainstream topic, in part because furanosides are often tricky to handle. Nevertheless, furanoses are quite widespread in Nature, being found in the cell wall polysaccharides of most plants and those of a variety of pathogenic microorganisms. Therefore, the development of chemoenzymatic tools and methods is a pertinent pursuit, notably in the context of the production of new furanose-based products for a whole range of applications, from tailored oligo- and polysaccharides and alkyl-polypentosides [6], to the synthesis of furanose-based glycomotifs for therapeutic purposes [7–9].

Whichever the GH under study, the key challenge associated with the use of these enzymes as glycosynthetic tools is competition with hydrolysis, which is an intrinsic and usually dominant activity of GHs. In this respect, thermodynamically-favorable hydrolysis competes with glycosynthesis at two stages. First, the two reactions compete for the donor substrate, with the outcome being either hydrolysis or glycosynthesis, and then, if synthesis occurs, hydrolysis intervenes again upon product accumulation, causing the decomposition of the transglycosylation product and thus a reduction of its final yield. To address this challenge, several approaches have been developed, with glycosynthases being a well-studied example. Glycosynthases are GHs that have been rendered catalytically-impotent through the elimination of the catalytic nucleophilic residue [10]. Only in the presence of highly-activated (generally 1-fluorosyl) donor glycosides, can glycosynthases mediate synthesis [4]. Moreover, once the donor glycoside is transferred onto an acceptor by a glycosynthase, the resulting product cannot be hydrolyzed. However, the fact that glycosynthases require glycosyl fluoride donors severely limits their application in the area of FHs, due to the extreme instability of furanosyl 1-fluorides.

Alternative strategies to create GH-based transglycosidases include directed evolution approaches that implement various screening protocols to select for GHs that better catalyze transglycosylation and/or perform less hydrolysis.

So far, although these experiments, which rely on random mutagenesis, have provided a variety of interesting observations, but taken together they have not yet provided any generic engineering rules for the improvement of the transglycosylation potency [11–23]. The modulation of the transglycosylation/hydrolysis (T/H) ratio via the modification of reactions conditions (e.g. increasing reactant concentrations [24,25], changing the operating temperature [26] or pH [27–30], or the use of a co-solvent [31]) is also an option that has been studied to promote GH-mediated glycosynthesis.

In retaining GHs, hydrolysis (and transglycosylation) operates via a double displacement mechanism that is catalyzed by two acidic residues (aspartate or glutamate), one acting as the general acid/base and the other as the nucleophile (Fig. 1). Therefore, since GH activity depends on the pK_a of the catalytic residues, 'pK_a cycling' during the reaction has been the subject of several careful analyses [32–38] and a recent report has provided significant insight into how residues that surround the catalytic dyad can influence this process [39].

Nevertheless, precise generic rules to identify the factors that determine the ionization states of the nucleophile and acid/base catalytic residues and, more generally, the catalytically-pertinent electrostatic interactions within GHs, remain elusive. Therefore, rational engineering aimed at modifying the optimum pH of GHs is not yet feasible. Moreover, it is noteworthy that the impact of modifications of the pK_a of catalytic residues in GHs on transglycosylation has never been addressed. The glycosynthetic properties of the GH51 α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*Tx*Abf)
have been extensively described, making this enzyme a paradigm for FH-mediated transglycosylation [40–44]. In recent work, we have adopted a directed evolution approach in an attempt to modulate the T/H ratio, using a high-throughput assay to select *Tx*Abf variants that display higher activity in the presence of artificial or natural acceptor sugars. This work has provided a series of mutants, some of which display significantly lower levels of primary and secondary hydrolysis when compared to the wild type enzyme [44,45]. Among the mutations that have been detected, we observed that the residue N344 was quite often subject to substitution. Therefore, in our present study, we have performed a thorough characterization of a library of N344 mutants, which has allowed us to offer a detailed description of the impact on hydrolytic and transglycosylation activities of mutations at this position

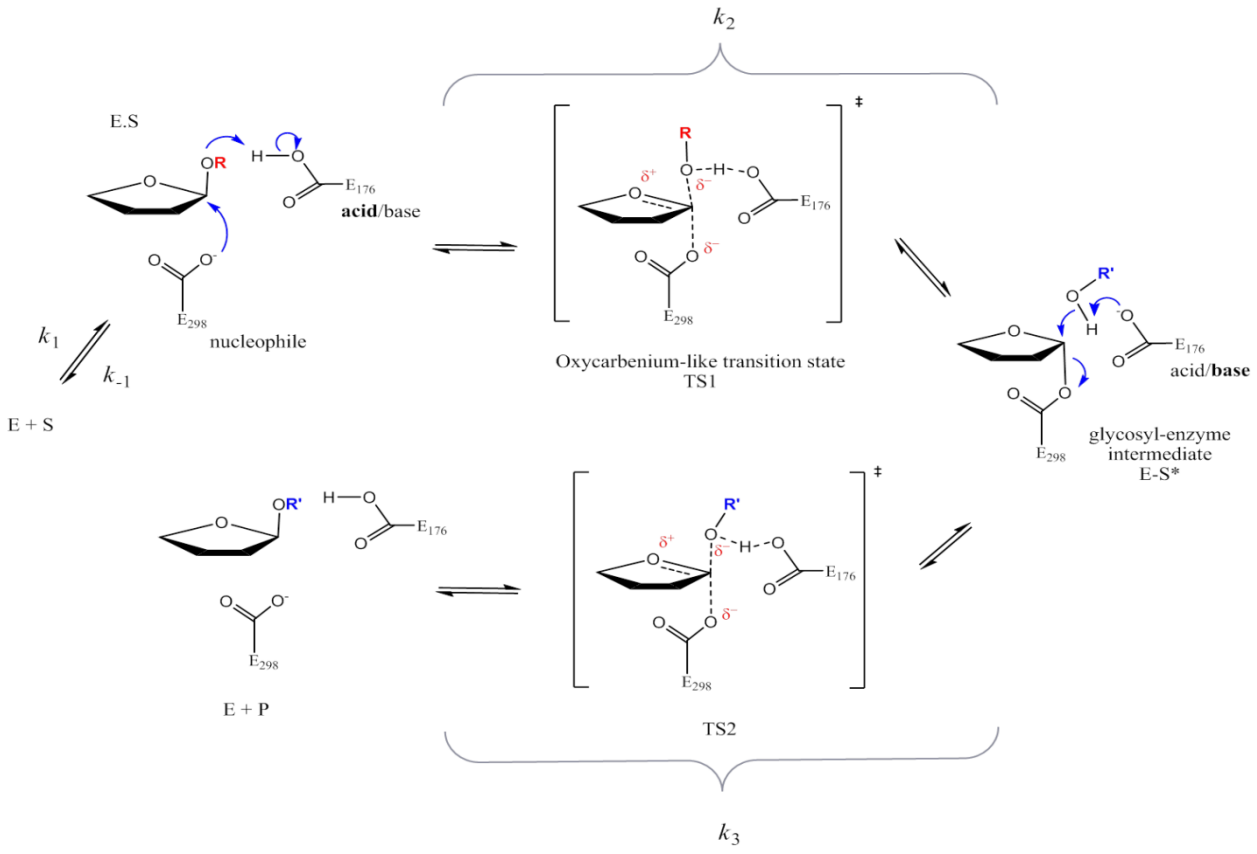


Fig. 1: Schematic representation of *Tx*Abf-catalyzed reactions through the double displacement mechanism of retaining GHs (R = leaving group and R' = xylobiosyl, xylotriosyl for transglycosylation or H for hydrolysis).

2. Materials and methods

2.1 Substrates and chemicals

Unless otherwise stated, routine experimental work was performed using chemicals purchased mainly from Sigma-Aldrich (Illkirch, France) and molecular biology reagents purchased from New England Biolabs (Evry, France). The substrate, *p*NP-α-L-Ara*f*, was prepared in-house according to established or adapted protocols [46,47]. The AXOS, A^3X and XA^3XX, were isolated and purified as previously described [48,49]. Xylotriose and XOS, used as acceptors, were purchased from Wako Chemicals GmbH (Neuss, Germany). Xylobiose was prepared in-house as previously described [50], using XOS as the starting material. ESI-HRMS: *m/z* calcd for $C_{10}H_{18}O_9Na$: $[M+Na]^+$ 305.0843; found: 305.0845 (1 ppm) [51].

2.2 Mutagenesis, protein expression, and purification

The plasmid construction pET24a-*Tx*Abf (original pET vectors from Novagen, Fontenay-sous-Bois, France) was used as template DNA for *in vitro* mutagenesis using the QuikChange Site-Directed Mutagenesis kit (Agilent, Courtaboeuf, France), following the manufacturer's instructions.

The asparagine residue present at position 344 (NNN = AAC) in the *Tx*Abf-encoding sequence (GenBank accession number CAA76421.2) was replaced by the 19 other amino acids using the following primers (underline letters represents mutated bases in NNN = AAC codon) for N344A (GCG), N344C (TGC), N344D (GAT), N344E (GAA), N344F (TTC), N344G (GGC), N344H (CAT), N344I (ATT), N344K (AAA), N344L (CTG), N344M (ATG), N344P (CCG), N344Q (CAG), N344R (CGT), N344S (AGC), N344T (ACC), N344V (GTG), N344W (TGG), N344Y (TAT):

N344X 5'- CGCGTCCGGATGGCG**NNN**ATCGCCCAGCTCG -3'

For STD experiments, mutations at position 344 were combined with the mutation E176A, which led to the production of inactivated enzymes, hereafter denoted by the symbol '‡' (e.g. $N344Y^‡$). This mutation was achieved using the following primer:

E176A 5'-GGCGTCGGCAAC**GCC**AACTGGGGCTGC-3'

The successful introduction of mutations was confirmed by DNA sequencing (GATC Biotech, Mulhouse, France). Expression and subsequent purification of wild-type and mutated recombinant *Tx*Abf, produced in *E. coli* BL21(DE3), were performed as previously described [52]. Briefly, enzyme crude extracts were obtained by sonication of cell pellets on ice (Bioblock Scientific vibracell 72434), employing cycles of 0.5 s 'on', 1.5 s 'off' during 8 min, with the probe operating at 30% of maximal power. The cell extract was then heat-treated (75 °C, 30 min) and clarified using centrifugation (30 min at 11,000 x g, 4 °C). Exploiting the presence of a C-terminal $(His)_6$ tag, *Tx*Abf and mutants thereof

were purified using a Cobalt resin (TALON® Metal Affinity Resin, Clontech Laboratories, Inc.) contained in a disposable column (Bio-Spin column, Bio-Rad). Elution of recombinant proteins from the resin was achieved under gravity feed using a gradient of imidazole in TALON buffer (one volume each of 5 and 10 mM imidazole, followed by three volumes of 100 mM imidazole and finally one volume of 200 mM imidazole). Fractions were then analyzed using SDS-PAGE (Bio-Rad) and those containing purified enzyme were pooled and dialyzed overnight at 4 °C against Tris-HCl buffer (20 mM, pH 8) and stored at 4 °C.

2.3 Kinetic Studies

Michaelis-Menten kinetic parameters K_M, k_{cat} and k_{cat}/K_M of *Tx*Abf and mutants thereof were determined using a discontinuous enzyme assay. Reactions were performed in duplicate at 45 °C in buffered conditions using *p*NP-α-L-Ara*f* (0.25-10 mM) and 0.1 mg.mL^{-1} BSA. For buffering, sodium acetate buffer (50 mM) was used in the pH range 3.5-5.8, while sodium phosphate buffer (50 mM) was used to cover the pH range 5.8-9.0, with a common point being measured in both buffers at pH 5.8 for each enzyme and each substrate concentration. The total reaction volume was 400 µL and reactions were performed over a 10-20 min period, removing samples at regular intervals. Hydrolysis reactions were terminated by addition of 250 µL of sodium carbonate (1 M) to 50 µL of reaction mixture. After, absorbance at 401 nm was measured using a VersaMax microplate spectrophotometer and the quantity of *p*NP released was calculated using an appropriate standard curve, prepared using pure *p*NP. Negative controls containing all of the reactants except the enzyme were used to correct for spontaneous hydrolysis of the substrates. One unit (IU) of enzyme specific activity corresponds to the amount of enzyme releasing one µmol of *p*NP per minute. Appropriate enzyme quantities (in the nM range) were used in combination with suitable substrate concentrations, such that less than 10% of the total substrate was hydrolyzed over the course of the measurement. The pH of each assay solution was measured at the end of the reaction. Experimental initial rates were fitted to the standard Michaelis-Menten expression using SigmaPlot 11.0 software (Systat software Inc, Ritme, Paris, France), and k_{cat}/K_M data were plotted as a function of pH and fitted to a bell-shaped activity profile as described in equation 1 [37]. This analysis provided the apparent pK_a values corresponding to the acid/base (pK_{a1}) and the nucleophile (pK_{a2}) catalytic amino acids, which were determined by nonlinear least-squares fitting on Excel.

$$\frac{k_{cat}}{K_M} = \left(\frac{k_{cat}}{K_M}\right)_{max} \cdot \left(\frac{1}{1+\frac{10^{-pH}}{10^{-pKa1}}+\frac{10^{-pKa2}}{10^{-pH}}}\right) \quad \text{(Eq. 1)}$$

2.4 Monitoring transglycosylation

2.4.1 Thin Layer Chromatography screening

Analytical thin-layer chromatography (TLC - silica gel 60F_{254} precoated plates, Merck) was used to qualitatively monitor reactions performed at optimal temperature, *p*NP-α-L-Ara*f* (5 mM) and Bn-α-D-Xyl*p* (5 mM). Reactions were stopped when the donor was fully consumed. TLC was achieved using an ethyl acetate/acetic acid/water (7:2:2, v/v/v) mobile phase and the migration profile was revealed using a UV lamp, following soaking in an orcinol solution (1 g in 1 L of sulphuric acid/ethanol/water solution, 3:72.5:22.5, v/v/v) and charring.

2.4.2 Time course NMR monitoring

Reactions were monitored by collecting ^{1}H NMR spectra using a Bruker Avance 500 spectrometer. First, the enzymatic solutions were diluted 10-fold in D_2O (99.90%) and then concentrated on an Amicon® Ultra filter (regenerated cellulose 10 kDa, Millipore), this operation being performed twice. The reference used was the residual water peak, calibrated at δ = 4.55 ppm at 45 °C [53]. Transglycosylation reactions were performed at 45 °C, a temperature arbitrarily chosen to reflect a compromise between activity and stability. Reactions were prepared in 600 µL of buffered D_2O (final volume), containing donor and acceptor at a ratio of either 1:1 (5 mM each) for Bn-α-D-Xyl*p*, or 1:2 (5 and 10 mM respectively) for xylobiose and xylotriose. Deuterated acetate (Euriso-Top, France) was used to prepare a buffer displaying a pD value of 5.9, while deuterated phosphate was used to attain a value of pD 7.1. Deuterated sodium phosphate was prepared in-house by dissolving sodium phosphate in D_2O, followed by lyophilisation. This two-step protocol was repeated three times to achieve sufficient deuteration. Values of pD were measured by determining pH using a glass pH electrode and then applying the equation pD = $pH_{electrode}$ + 0.41 [54].
The amount of *Tx*Abf or mutant enzyme added was adjusted for each enzyme to achieve 0.05 IU (activity measured on *p*NP-α-L-Ara*f*, in D_2O at 45 °C), thus final enzyme concentrations were in the range 8-25 µM. ^{1}H NMR spectra were accumulated semi-continuously over 4 to 15 h by accumulating a series of 5.52 min scan periods (i.e. 32 scans) that were interspersed by delay periods of 6 s. The exact full monitoring period was dependent on the enzyme employed. Each NMR spectrum was acquired using an excitation flip angle of 30° at a radiofrequency field of 29.7 kHz, and the residual water signal was pre-saturated during the repetition delay (with a radiofrequency field of 21 Hz). The following acquisition parameters were used: relaxation delay (6 s), dummy scans (4). For each reaction, before enzyme addition, an NMR spectrum of the reaction mixture was acquired, serving as the starting point of the reaction, from which integrals were then corrected according to the small dilution factor induced by the enzyme addition (< 5% of total volume).

2.4.3 NMR kinetics analysis

For data processing, the time-dependent evolution of donor (*p*NP-α-L-Ara*f*) and acceptor (Bn-α-D-Xyl*p*) concentrations were quantified by integrating the relative anomeric proton signal (5.85 and 4.98 ppm, respectively). Molar balances, based on acceptor and donor consumption, were used to convert the transglycosylation product signal integral (5.08 and 5.07 ppm for

H-1 of α-L-Ara*f* and H-1 of α-D-Xyl*p*, respectively) into concentration, and to evaluate hydrolysis product concentrations, respectively. Moreover, agreement between acceptor and transglycosylation product integrals was also verified. In the case of xylobiose and xylotriose, signal overlapping excluded the possibility of performing reliable integrations. Consequently, as an alternative, first donor consumption was evaluated using the signals of five different entities: H(*p*NP)$_{meta}$ or H(*p*NP)$_{ortho}$ in the free state, released after *p*NP-α-L-Ara*f* hydrolysis (8.14 and 6.86 ppm, respectively) and linked forms (8.28 and 7.25 ppm, respectively) as well as from H-1 of *p*NP-α-L-Ara*f* (5.85 ppm). Since the sum of *p*NP (free and linked) remains constant throughout the reaction, this information was used as an internal standard to correlate the integral value to the initial concentration, independently for each spectrum, as follows:

$I_{pNP,\ meta\ (free)} + I_{pNP,\ meta\ (linked)} = 2.I_{0,\ meta}$ (Eq. 2)

$I_{pNP,\ ortho\ (free)} + I_{pNP,\ ortho\ (linked)} = 2.I_{0,\ ortho}$ (Eq. 3)

, where I_0 corresponds to the integral of one proton at t_0 (initial concentration), and the sum of $I_{pNP,\ x}$ corresponds to $2.I_0$, since I_{pNP} integrates for two protons. A mean value for I_0 ($I_{0,\ mean}$) was calculated from $I_{0,\ meta}$ and $I_{0,\ ortho}$. Finally, the donor substrate conversion rate (X) can be calculated in three different ways from one spectrum as follows:

$X_{meta}\ (\%) = (I_{pNP,\ meta\ (free)}/2.I_{0,\ meta} + I_{auto}/I_{0,\ mean}) * 100$ (from Eq. 2)

$X_{ortho}\ (\%) = (I_{pNP,\ ortho\ (free)}/2.I_{0,\ ortho} + I_{auto}/I_{0,\ mean}) * 100$ (from Eq. 3)

$X_{H\text{-}1}\ (\%) = (1 - I_{H\text{-}1}/I_{0,\ mean}) * 100$

The donor conversion takes into account the consumption of the donor substrate toward hydrolysis and transglycosylation products, as well as self-condensation (I_{auto}). A mean conversion rate and the associated error are then deduced from the three previous values. For all time-course NMR kinetics, the absolute error mean value on X ranged between 1 and 3%.

Next, the transglycosylation yield was determined by integrating the signal of the anomeric proton of products (5.39, 5.32 and 5.27 ppm for XA^3X, A^3XX and A^2XX (or XA^3, A^3X and A^2X, respectively, when xylobiose was used as an acceptor) [48,55–59] and then dividing by $I_{0,\ mean}$, thus procuring a transglycosylation yield (Y): $Y_{(product)}\ (\%) = (I_{H\text{-}1(product)}/I_{0,\ mean}) * 100$. Finally, $Y_{(product)}$ (%) was plotted against X (%) to provide a transfer rate (R_T) of the donor substrate to a given product (µmoles of product/µmoles of consumed substrate), that is independent of the duration of the reaction (Supplementary Table 1 and Fig. 5). For each kinetics, the linear region for slope determination varied from X = 0-25% up to 0-40%, with root mean square deviations between 0.95 and 0.99.

2.5 STD NMR experiments

Samples were prepared in 600 µL of D_2O (5 mm NMR tubes) when *p*NP-α-L-Ara*f* (100 µM) was used as ligand, or in 170 µL of D_2O (3 mm NMR tubes) for A^3X (2 mM) or XA^3XX (2 mM), with a constant molar ratio of 1:100 (protein:ligand). STD NMR experiments were performed at 283 K with a Bruker Avance 600 spectrometer equipped with a cryoprobe TCi [60]. To achieve this, proteins were saturated on resonance at -0.4 ppm on methyl signals, and off resonance at 30 ppm, with selective Gaussian-shaped pulses of 50 ms duration, at a radio-frequency field of 86 Hz, with a 100 ms delay between each pulse. The total saturation time was 2 s. A WATERGATE sequence was used to suppress residual HOD signal. An identical experiment with no enzyme was used as a negative control in order to verify the selectivity of the saturation and the efficiency of the signal subtraction used to obtain the difference spectrum. Intensities of all STD effects (I_{STD}) were calculated by integration of the respective ^{1}H NMR signals and standardized with the reference signal I_0. The ratios of the intensities I_{STD}/I_0 were normalized using the largest STD effect (the H-3 proton of the L-Ara*f* unit was set to 100%) as a reference. In order to compare the impact of mutations on absolute STD effects, STD effects of protons exhibiting identical chemical shifts were standardized for each mutant, using the STD data obtained with *Tx*Abf‡. All data were acquired and processed using Topspin v2.1 software (Bruker).

3. Results

3.1 The impact of N344 mutations on transglycosylation

During the course of previous work involving the screening of randomly generated libraries of *Tx*Abf variants, five independent clones selected for their apparent ability to better perform transglycosylation were found to be mutated at position N344 [44,45]. The substitution N344Y occurred twice, while N344K, S and I were each found once. Interestingly, N344 is not directly located within the active site, but is rather found on strand β8 in the vicinity of the catalytic nucleophile E298 (β7 strand), within a zone that can be referred to as the second shell (Fig. 2A). Examination of the conservation of N344 within GH51 family revealed that N344 is not a highly conserved residue, though 270 sequences (55%) do display this side-chain at an equivalent position. The next most frequent occurrences are cysteine (147 sequences or 30%) and serine (9%), which like asparagine also display electronegative character (Supplementary Fig. SI 1). The localization of N344 homologues in the structures of other GH51 Abfs showed that homologous asparagine residues are identically positioned in two Abfs of *Thermotoga* sp. origin (N323 for *Tm*Abf and *Tp*Abf, PDB ID: 3UG5 and 3S2C respectively), while homologous residues in

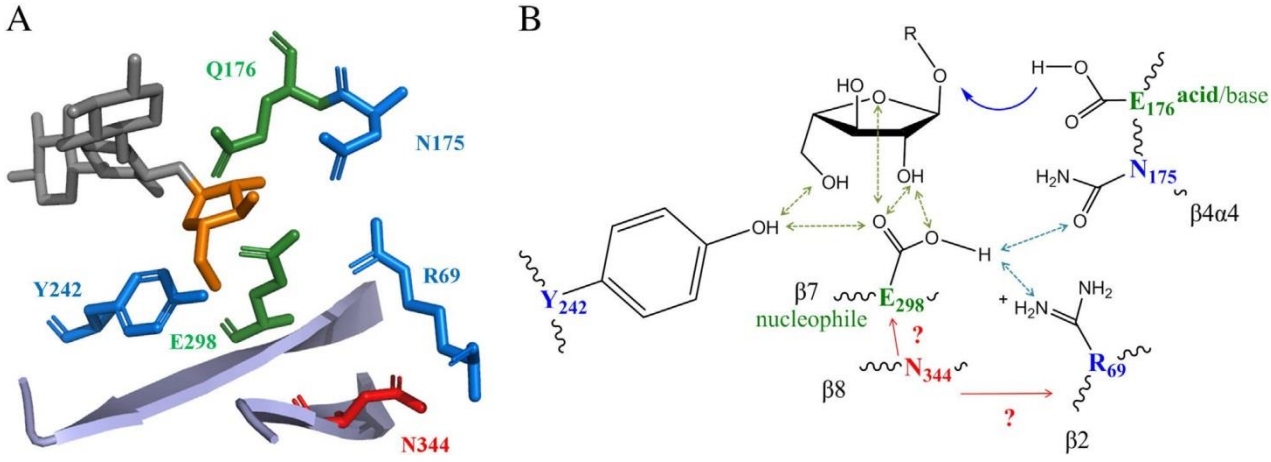


Fig. 2: View of the *Tx*Abf-XA^{3}XX complex active site. (A) N344 location relative to the inactivated *Tx*Abf-E176Q active site close to the catalytic nucleophile E298 and to the L-Ara*f* moiety in subsite −1 (PDB ID: 2VRQ) [66]; and (B) putative impact (red arrows) on hydrogen bonding (green and light blue) and electronic displacement system

Abfs from *Geobacillus stearothermophilus* (C348 in PDB ID: 1QW8), *Clostridium thermocellum* (C349 in PDB ID: 2C7F) or *Bifidobacterium longum* (S406 in PDB ID: 2Y2W) adopt similar, but not identical, orientations and occupy similar steric volumes. Therefore, accounting for these observations and to further investigate the precise role of N344, notably with regard to its potential role in determining the partition between hydrolysis and transglycosylation, we created a site-saturation mutant library, which was then submitted to a variety of analyses. To begin characterization of the nineteen position 344 mutants, a TLC-based assay was used to assess the ability of individual mutants to transfer an L-arabinofuranosyl (L-Ara*f*) donor moiety onto the acceptor, Bn-α-D-Xyl*p*. Indeed, it has been shown that *Tx*Abf can catalyze this reaction, forming Bn α-L-Ara*f*-(1,2)-α-D-Xyl*p* [40,44]. In our assay, six mutants (N344F, G, K, P, W and Y) displayed potentially higher yields than the reference reaction (13%), with three of these also displaying highly reduced secondary hydrolysis (N344G, P and Y). Therefore, using these preliminary results, the six improved mutants were selected for further analysis using time-course NMR, which confirmed our initial observations, revealing that the yield of transglycosylation product reached up to 20% in the case of N344K and 18% for N344Y or 15% for N344P and G (Supplementary Fig. SI 2). Moreover, whereas for N344K secondary hydrolysis was observed (similar to *Tx*Abf), in the case of N344P and Y, once synthesized, the amount of Bn α-L-Ara*f*-(1,2)-α-D-Xyl*p* remained stable for over 100 min, indicating that secondary hydrolysis was diminished to an undetectable level. The reaction catalyzed by the mutant N344G displayed a similar profile, although some secondary hydrolysis was evidenced. In this respect it is noteworthy that the thermostability of the six mutants was mostly unaffected, since all conserved > 85% of residual activity after incubation for 4 h at 45 °C, thus confirming that reductions in primary and secondary hydrolysis were not due to heat denaturation

Finally, for the different mutants it was noted that for the donor/acceptor pair, *p*NP-α-L-Ara*f*/Bn-α-D-Xyl*p*, the transglycosylation product yield reached a stable endpoint, but donor consumption continued, although this was never complete, with 15% initial *p*NP-α-L-Ara*f* remaining at the end of the reaction for mutants N344P and Y. Having revealed the improved transglycosylation ability of the six N344 mutants, N344G, P and Y were submitted to further tests in order to ascertain the ability of these mutants to catalyze the transfer of L-Ara*f* donor moieties onto xylobiose or xylotriose, taking into account the fact that when catalyzing such reactions parental *Tx*Abf procures quite low yields (10%) [45].

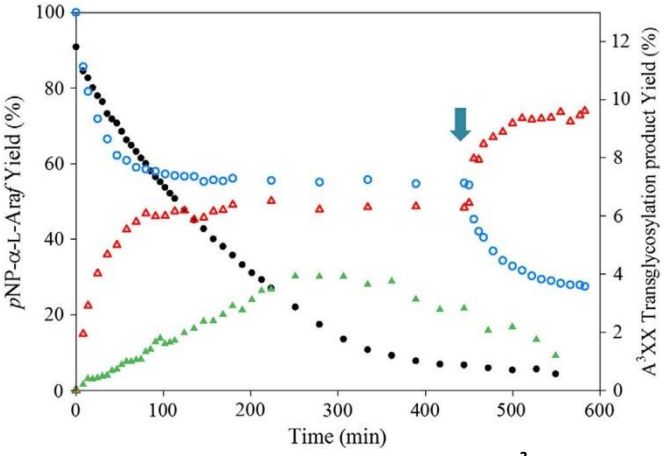


Fig. 3: Time-course NMR monitoring of A^{3}XX (5.32 ppm) transglycosylation product (▲ and △, for *Tx*Abf and N344Y respectively) obtained by transglycosylation reaction catalyzed by *Tx*Abf (8 nM) or N344Y (419 nM), with *p*NP-α-L-Ara*f* (5 mM) as donor (● or ○, respectively) and xylotriose (10 mM) as acceptor. An equal quantity of enzyme (419 nM) was added once steady state was reached (at t = 455 min, blue arrow). Assays were conducted in 25 mM deuterated acetate buffer (pD 5.9) at 45 °C.

Table 1: Steady-state kinetic parameters for the hydrolysis of *p*NP-α-L-Ara*f* by *Tx*Abf and N344P and Y mutants.

Enzyme	k_{cat} [a] (s^{-1})	K_M [a] (mM)	k_{cat}/K_M ($s^{-1}.mM^{-1}$)	Relative k_{cat}/K_M	$pK_{a\ E298}$ [b]	$pK_{a\ E176}$ [b]	pH_{opt} [c]
*Tx*Abf	431 ± 10	0.27 ± 0.03	1671	100	4.6	7.6	6.1
N344P	81 ± 3	0.98 ± 0.13	81	4.8	6.1	7.8	7.0
N344Y	17 ± 0.6	0.63 ± 0.09	28	1.7	6.4	7.6	7.03

[a] k_{cat} and K_M values were measured at 45 °C, pH 5.8 (for *Tx*Abf) and pH 7.0 (for mutants N344P and Y).
[b] Apparent p*K*a were determined by fitting experimental values to the bell-shaped model shown in Fig.4 (Eq. 1 from section 2.3). The error on pK_a values is estimated to be 0.1 pH unit.
[c] pH_{opt} was determined as the value of the pH when the derivative of the model, to which experimental points were fitted, is equal to zero.

Like the previous analyses, time-course NMR revealed that the reactions attained a plateau after 100 min (Fig. 3) and quantification of products showed that N344P and N344Y procured higher yields compared to *Tx*Abf (Supplementary Table SI 1). Closer analysis of the products produced in each reaction revealed that four regio-isomers were formed in different proportions (Supplementary Fig. SI 3).

Compared to reactions catalyzed by *Tx*Abf, those catalyzed by N344Y displayed increased yields (1.7-, 1.8- and 2.5-fold) of A^3XX (5.32 ppm), A^2XX (5.27 ppm) and XA^3X (5.39 ppm), respectively. It is noteworthy that because N344Y catalyzed the formation of XA^2X (5.28 ppm) [56,58] in quite negligible quantities, rather like parental *Tx*Abf, this particular reaction was not further analyzed.

Furthermore, N344Y did display higher transfer rate, R_T, with increases reaching up to 4.6-fold compared to *Tx*Abf. Regarding N344K-catalyzed reactions performed in the presence of xylobiose or xylotriose, these provided maximum transglycosylation yields (Y_{max}) of up to 14% for A^2XX as opposed to *Tx*Abf (6%), associated with a 4.7-fold increase in the R_T.

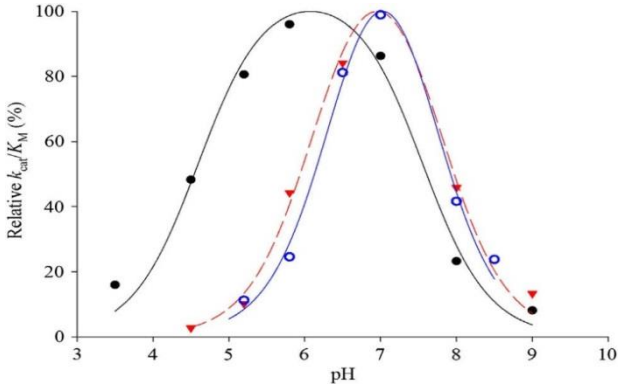


Fig. 4: pH dependency of k_{cat}/K_M for *Tx*Abf (●, solid black), N344P (▼, dashed red) and N344Y (○, solid blue). For each enzyme, the relative k_{cat}/K_M represents a percentage of the maximum $(k_{cat}/K_M)_{max}$ obtained from the theoretical model, for experimental (symbols) and model-fitted points (lines).

3.2 Kinetic analysis of the hydrolytic activities of N344 mutants

According to the double displacement mechanism applied to retaining GHs, transglycosylation occurs at the deglycosylation step, and when donor substrates with good leaving groups are employed (i.e. $k_2 >> k_3$), $k_{cat} = k_2.k_3/(k_2 + k_3) \approx k_3$ (Fig. 1) [61–63]. Therefore, to investigate how mutations at position 344 might have altered the deglycosylation step, the kinetic parameters describing the hydrolysis of *p*NP-α-L-Ara*f* catalyzed by N344P and N344Y were determined in reactions buffered at the optimum pH of *Tx*Abf (Table 1). As expected, this revealed that the k_{cat} values describing the two reactions were strongly decreased (81 and 96% for N344P and N344Y respectively), compared to the k_{cat} value of *Tx*Abf. Furthermore, the values of K_M were slightly raised and k_{cat}/K_M, or alternatively $k_1.k_2/(k_{-1} + k_2)$, reflecting the glycosylation step (enzyme-substrate association and glycosidic bond cleavage) was severely reduced (21 and 60-fold for N344P and Y respectively). Next, in order to better understand these results in terms of the electronic charge developed within the active site, the apparent pK_a values of the catalytic amino acids were determined by measuring the pH-dependency of the second order rate constant k_{cat}/K_M [37]. The data for *Tx*Abf and the two mutants were fitted to the bell-shaped model (Fig. 4), revealing summits at pH 6.1 (*Tx*Abf) and 7.0 (N344P and Y). Likewise, the data was found to be consistent with the presence of two ionizable groups, presumably corresponding to the catalytic acid/base (E176) and the nucleophile (E298), with apparent pK_a values for *Tx*Abf of 7.6 and 4.6, respectively. However, for the two mutants the apparent values of the lower pK_a were increased by >1 pH unit to 6.1 (N344P) and 6.4 (N344Y) respectively, while the more basic limbs of the pH-dependency profiles were mostly unaffected (7.8 and 7.6, respectively), with variations being considered to be within the error range (0.1 pH unit).

3.3 Kinetic isotope effects

To further investigate the reactions catalyzed by the mutants N344P and Y, solvent kinetic isotope effects (KIE) were determined and compared to those of *Tx*Abf. These experiments revealed that the substitution of water hydrogen atoms by deuterium led to an increase of the ratio k_H/k_D, by only 4% for N344P and 37% in the case of N344Y (Supplementary

Table SI 2). The wild-type solvent KIE value (1.88) reveals a modest impact of deuterium substitution, but is in agreement with other KIE values measured for other GHs [63,64].

3.4 pH optima for glycosynthesis

Taking into account the altered pH optimum of the mutants and the failure of external nucleophiles to rescue the reaction, the impact of the ionization state in the active site on transfer activity and secondary hydrolysis was investigated. To achieve this, time-course NMR was used to monitor both the consumption of *p*NP-α-L-Ara*f* and the production of transglycosylation products in comparable reactions catalyzed by N344P or Y, operating at pD 7.1 or pD 5.9. At pD 7.1, consumption of *p*NP-α-L-Ara*f* was complete within 100 min and evidence of secondary hydrolysis was detected, since the maximum yields of the transglycosylation products were not conserved over time (Fig. 5). In contrast, at pD 5.9, *p*NP-α-L-Ara*f* was incomplete and secondary hydrolysis abolished, thus once reached the maximum yields remained stable. Nevertheless, it is noteworthy that when an identical experiment was performed using *Tx*Abf operating at pD 4.6 (i.e. approximately at the pK_a of the nucleophile residue), although *p*NP-α-L-Ara*f* consumption was impaired, secondary hydrolysis was conserved (Supplementary Fig. SI 4), indicating that the hydrolytic behavior of N344P and N344Y cannot be solely attributed to changes in the ionization state in the active site.

3.5 pH-dependent inhibition mediated by the leaving group

As well as being pH dependent, the catalytic behavior of the mutants N344P or Y also appeared to be tightly correlated with *p*NP-α-L-Ara*f* decomposition, since maximum transglycosylation yield was reached at approximately the point at which donor consumption ceased. Regarding, the premature endpoint of the reaction, it seemed likely that this be the result of concentration-dependent product inhibition. Performing reactions in the presence of various concentrations of *p*NP quickly confirmed this hypothesis (Fig. 6A) revealing a competitive inhibition that could be partly alleviated at pD 5.9 by adding more enzyme (Fig. 6B). Being pH-dependent, inhibition was fully alleviated at pD 7.1, although this did not appear to be linked to the protonation state of *p*NP, or the relative positions of the hydroxyl and nitro groups. Indeed, the use of other related compounds displaying different pK_a values and structures, such as *o*NP, *m*NP and 4-nitrocatechol (4NTC), provided very similar results (Figs. 6B-D).

3.6 Enzyme-substrates/products recognition and interactions

Taking into account the fact that the R_T of L-Ara*f* are increased in reactions catalyzed by mutants N344P and Y, enzyme-substrate and enzyme-product interactions were studied to better understand their role in the T/H partition. Monitoring the release of *p*NP from *p*NP-α-L-Ara*f* in presence of 10 mM of xylobiose revealed that *Tx*Abf was inhibited, with a relative activity of 73%, compared to the same reaction performed in the absence of acceptor [45]. However, N344P and Y did not display this property since, in the presence of an identical concentration of xylobiose, relative activities were slightly increased to 105 and 117% respectively (Supplementary Table SI 3). Therefore, to determine whether the modified T/H ratios of reactions catalyzed by the N344 mutants modification entail alterations in donor and/or transglycosylation product recognition, Saturation Transfer Difference (STD) NMR was employed to probe substrate-associated proton:enzyme interactions. To achieve this, inactivated enzymes, *Tx*Abf‡, N344G‡, N344P‡ and N344Y‡, were prepared and used. Measurement of the residual activities of the various inactivated enzymes confirmed that, in the conditions of the STD NMR experiments, all were sufficiently inactive on *p*NP-α-L-Ara*f* (*Tx*Abf‡, 0.039 ± 0.003; N344G‡, 0.015 ± 0.001; N344P‡, 0.014 ± 0.002 and N344Y‡, 0.0030 ± 0.0003 IU.mg^{-1}), even though slight, but insignificant, modifications of the anomeric proton (H-1) signal were detected (Fig. 7A). Regarding the interaction of *p*NP-α-L-Ara*f* with the mutants, examination of the STD effects of the different protons, normalized using the STD effects measured for *Tx*Abf‡ (i.e. I_{STD}/I_0), revealed that interactions were globally less strong, consistent with K_M values (Fig. 7B and Table 1).

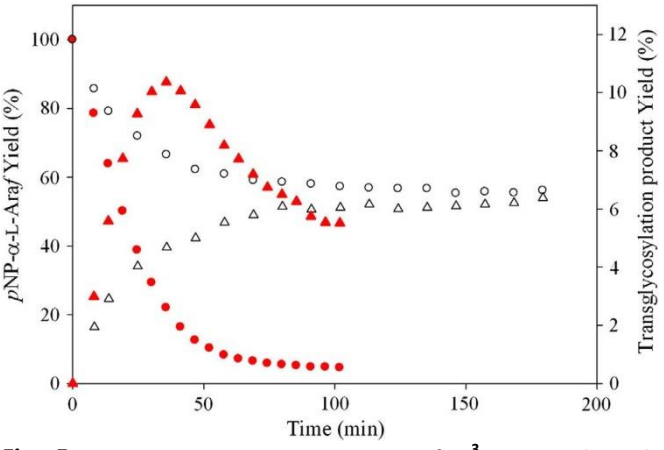


Fig. 5: Time-course NMR monitoring of A^{3}XX transglycosylation product (5.32 ppm) (△ and ▲ for pD 5.9 and 7.1, respectively) obtained by transglycosylation reaction catalyzed by N344Y (419 nM), with *p*NP-α-L-Ara*f* (5 mM) as donor (○ and ●, respectively) and xylotriose (10 mM) as acceptor. Assays were conducted in 25 mM deuterated sodium acetate (pD 5.9) or phosphate (pD 7.1) buffer, in D_2O, at 45 °C.

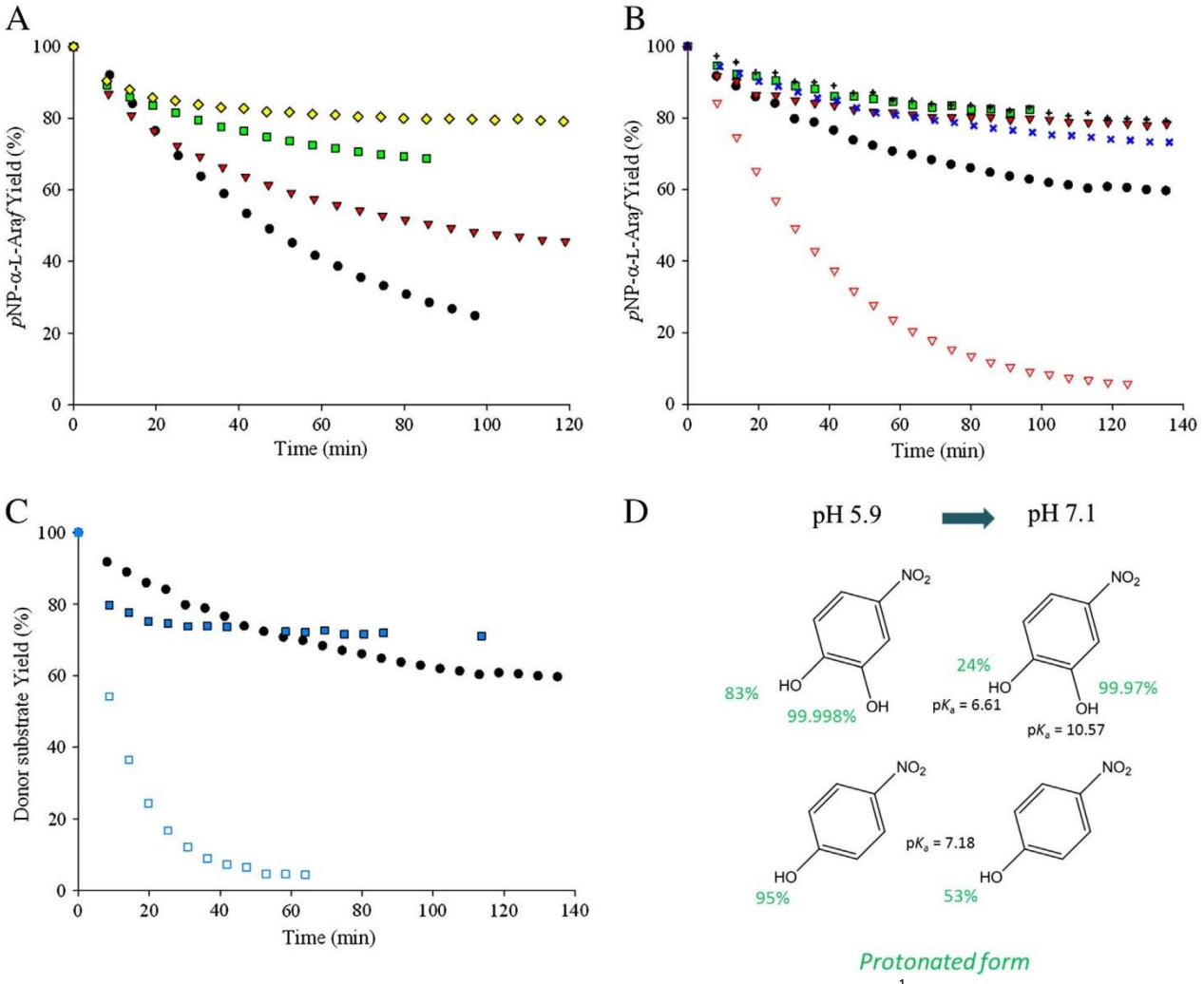


Fig. 6: Leaving group inhibition analysis of N344Y. The consumption of *p*NP-α-L-Ara*f*, monitored using ^{1}H NMR, was performed at pD 5.9 using different initial concentrations of *p*NP (0 mM, ●; 2 mM, ▼; 5mM, ■; and 7.5 mM, ◆) and (A) 250 nM of N344Y or (B) double N344Y concentration (500 nM), revealing weaker inhibition effect. (A) The protonated form of *p*NP is not responsible for inhibition as shown by the progress curve performed at pD 7.1 (▽) with 1.9 mM of protonated *p*NP (53% of 3.55 mM initial *p*NP concentration) as opposed to 1.9 mM protonated *p*NP at pD 5.9 (▼, 95% of 2 mM initial *p*NP concentration). The leaving group hydroxyl position does not drastically modifies the inhibition profile as highlighted by *p*NP-α-L-Ara*f* (5 mM) consumption in presence of 5 mM of *o*NP (x) or *m*NP (+). (C) An extra vicinal hydroxyl does not alleviate the inhibition as shown by the reaction carried out with 4NTC-α-L-Ara*f* (5 mM) as donor substrate at pD 5.9 (■) and does not prevent enzyme 'reactivation' at pD 7.1 (□), compared to 5 mM of *p*NP (●). pH dependence inhibition assays were conducted in 25 mM deuterated acetate (pD 5.9) or phosphate (pD 7.1) buffer, at 45 °C. (D) Percentage of leaving group protonated form.

Even more strikingly, when the STD effects characterizing the interaction of the different mutants with A^{3}X and XA^{3}XX were measured, the I_{STD}/I_0 values of those protons that could be clearly isolated within the spectra were significantly lowered, indicating that the affinities of the mutants for such compounds are largely inferior to those for *p*NP-α-L-Ara*f*. When using A^{3}X or XA^{3}XX, the enzyme:ligand ratio was maintained at 1:100 (i.e. the same as in experiments conducted using *p*NP-α-L-Ara*f*), but the concentrations of the enzyme and ligand were increased 20-fold to maintain a constant bound enzyme concentration and obtain equivalent STD effects. Consequently, one can postulate that the value of the dissociation constant (k_{-1}/k_1) might be higher for A^{3}X and XA^{3}XX when compared to the previously reported value for *p*NP-α-L-Ara*f* (K_d = 0.16 mM) [65].

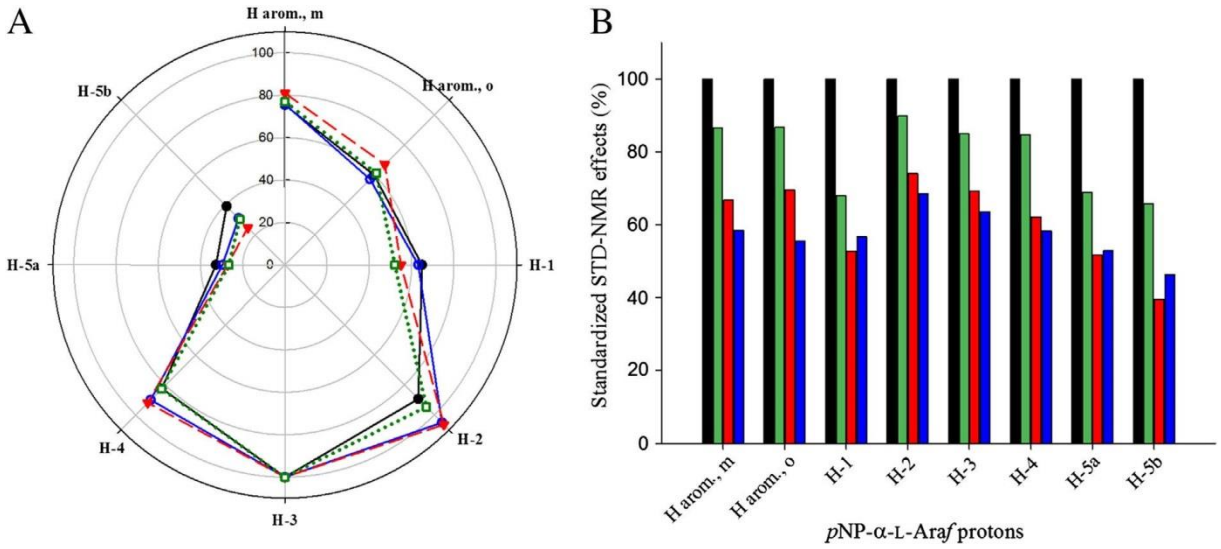


Fig. 7: STD NMR analysis of enzyme-ligand interactions. (A) STD NMR effect fingerprinting, expressed for each enzyme-ligand couple as relative percentages of the H-3 STD effect (100 %), between *p*NP-α-L-Ara*f* protons and the inactivated *Tx*Abf‡ (●, solid black), N344G‡ (□, dotted green), N344P‡ (▼, dashed red) and N344Y‡ (●, solid blue). (B) Standardized STD NMR effects expressed, for each proton, as relative percentages of the effect measured between *p*NP-α-L-Ara*f* protons and the inactivated *Tx*Abf‡ (black) for N344G‡ (green), N344P‡ (red) and N344Y‡ (blue).

Beyond the comparison of the absolute STD effects of mutants with *Tx*Abf‡, it was also considered interesting to perform a comparative analysis of binding patterns, using internal normalization of STD signals. Accordingly, considering the proton signals of the L-Ara*f* moiety of each AXOS (i.e. H-1, H-2/H-4 and H-3 for A^{3}X and H-1, H-2 and H-4 for XA^{3}XX), no major differences were observed between inactivated mutants and *Tx*Abf‡ (Supplementary Fig. SI 6). For example, considering the STD intensities of H-1 XA^{3}XX relative to H-2 XA^{3}XX (arbitrarily set to 100%) were 87, 91 and 93% for *Tx*Abf‡, N344P‡ and N344Y‡, respectively. Similarly, the STD intensities of H-1 A^{3}X relative to H-3 A^{3}X, these were 71, 73, 72 and 77% for *Tx*Abf‡, N344G‡, N344P‡ and N344Y‡, respectively. Significantly, these interactions are higher than those observed for H-1 *p*NP-α-L-Ara*f* relative to H-2 or H-3 *p*NP-α-L-Ara*f* (73 and 65% for *Tx*Abf‡, 55 and 52 for N344G‡, 52 and 55% for N344P‡ and 60 and 63% for N344Y‡, respectively). Consequently, although N344 mutants do not alter the overall STD fingerprints of the donor, or of compounds representative of transglycosylation products, it would appear that the interaction of the L-Ara*f* moiety with the enzymes is different depending on the nature of the groups occupying the acceptor subsite(s). Unfortunately, the complexity of the NMR spectra recorded for A^{3}X and XA^{3}XX precluded any STD analyses of interactions with the D-xylopyranosyl moieties.

4. Discussion

In previous work on *Tx*Abf, several active site residues have been shown to be important for hydrolytic activity [65,66]. However, to predict what the role of these, or that of other residues, could be in determining the T/H partition is not yet feasible, since our current understanding of this phenomenon is insufficient. Indeed, this remark holds true for all GHs, thus the option of using directed evolution to pinpoint pertinent residues for further investigation is an adequate, albeit opportunist, approach. In our previous work we have done just this, applying a screening strategy that is powerful enough to isolate mutants that have an apparent increased activity in the presence of acceptor sugars [45]. Having isolated mutants, the selection of clones for this study, specifically displaying substitutions at position 344, was based on a three-point argumentation: the repeated occurrence of mutations at position 344, even in two separate libraries, the relatively conserved nature of N344 in GH51 and, finally, the location of N344 in a second shell, close to the nucleophile catalytic residue E298. Taken separately these criteria may not have justified our study, but together they constitute a rather interesting starting point.

4.1 Why do mutations at position 344 alter the ionization state of the catalytic nucleophile?

Like many other GHs, *Tx*Abf catalyzes both the hydrolysis and the formation of glycosidic bonds via the well-known, two-step displacement mechanism described by D. Koshland [67]. This mechanism involves two amino acids bearing carboxylic acids, one acting as the general acid/base and the other as the nucleophile, the latter being deprotonated at the start of the reaction (Fig. 1). Accordingly, it is possible to affirm that the optimum pH for the activity of *Tx*Abf is related to two pK_a values, the higher being attributed to the acid/base (E176) and the lower one to the nucleophile (E298). Nevertheless, these pK_a values are actually apparent pK_a, since the active site can be

better considered as a 'polyprotic acid', because of the extensive hydrogen bonding network. The unusually high pK_a value for the acid/base amino acid in GHs is illustrative of this point, since free carboxylic acid functions display pK_a values < 5. However, in the case of the acid/base, its pK_a is strongly influenced by the nucleophilicity of the neighboring (within approximately 5 Å) catalytic nucleophile. In this study, the pK_a values measured for *Tx*Abf are quite similar to those obtained in others studies, conducted for example on a β-xylosidase (GH52) [64] or a xylanase (GH11) [37], and in general terms are consistent with the expected values for a GH. Importantly, these pK_a values were determined from experimental k_{cat}/K_M data points, thus they reflect the ionization states of the catalytic residues of unbound *Tx*Abf, meaning that the impact of mutations at position 344 is substrate independent. Accordingly, the modified optimum pH of the mutants N344P and Y is essentially due to an increased pK_a value for E298, which means that the fraction of appropriately charged E298 is decreased at pH 5.8. In order to understand how N344 modulates the 'nucleophilic strength' of E298, it is important to note that N344 is located on strand β8, which underlies the strand β7 bearing E298 (Fig. 2A). From this location it is highly probable that N344 participates in a hydrogen bonding network that includes E298.

4.2 Why do mutations at position 344 alter the turnover number?

In this work, we have shown that mutations at position 344 affected all of the Michaelis-Menten parameters. One explanation for these overarching effects could be that glycosidic bond cleavage step and leaving group departure (k_2) are affected, which is consistent with alteration of the ionization state of the catalytic residues and the observed drastic decrease of the k_{cat}/K_M values for N344P and Y. Nevertheless, the finding that solvent KIE values (i.e. H_2O versus D_2O activation/attack) for both N344 mutants and *Tx*Abf were > 1 suggests that deglycosylation (k_3) remains rate-limiting. Moreover, accounting for the fact that the solvent KIE values for N344P and *Tx*Abf were rather similar, it is logical to assume that the decreased k_{cat}/K_M value is due to a lowered k_2 value (leaving group departure), underlining the fact that the formation or disruption of the glycosyl-enzyme E298-C1(L-Ara*f*) bond has become a critical chemical step. Nevertheless, for N344Y a slight increase in the solvent KIE value was observed, which is possibly indicative of modifications to both k_2 and k_3.

4.3 The pH-dependent character of the catalytic properties of 344 mutants

Interestingly, when reactions catalyzed by N344P or Y mutants were performed at their optimum pH, both primary and secondary hydrolysis were restored, revealing that these enzymes are highly sensitive to the ionization state in the active site. This is apparently not the case for *Tx*Abf, which conserves hydrolytic activity, albeit no longer optimal, upon pH changes. This observation is probably indicative of a breakdown of the original hydrogen bond network in the mutants, which underpins normal, quite robust hydrolytic activity.

With regard to transglycosylation, when the optimum pH for the mutants is selected, the yield of transglycosylation products is increased. This can be partly explained by the fact that transglycosylation is under kinetic control, thus in the initial phase of the *Tx*Abf-catalyzed reaction, the concentration of transglycosylation products increases progressively within a general context of high donor concentration. However, at a certain point the donor concentration falls below a critical level, while transglycosylation simultaneously attains a critical high level, thus provoking secondary hydrolysis, which targets the transglycosylation product(s). In the case of the mutants, as mentioned above, the pH shift towards optimum pH partially restores hydrolysis, probably by providing adequate conditions for the acid/base to adopt its basic form, a requirement for the activation of incoming acceptors (water or sugar molecules). The pH shift also alleviates enzyme inhibition, which is somehow mediated by the *p*NP product, although the exact nature of the inhibition is unknown. Our results indicate that the protonated species is not the inhibitor and, because inhibition was observed at pH 5.8 (i.e. more than 1 log unit under the pK_a value of *p*NP), one can exclude *p*-nitrophenolate as the inhibitory species too. Furthermore, because the different nitrophenols (*p*NP, *o*NP and *m*NP), and even 4NTC, provoked similar inhibitory effects, it seems probable that this phenomenon is driven by an enzyme-product interaction that involves the phenyl ring, maybe via modified stacking interactions. In GHs that are subject to product inhibition, loop movements are sometimes involved [68,69]. Regarding *Tx*Abf, at least one significant loop movement has been shown to be involved in catalysis [65,66], thus pH-dependent protein dynamics could well be the basis for inhibition, but further work needs to be done to investigate this.

4.4 What are the underlying reasons for the improved transglycosylation ability of the mutants?

The use of xylobiose as acceptor appeared to slightly activate mutants N344P and Y, procuring higher activity compared to *Tx*Abf, which is actually inhibited (Supplementary Table 3). This observation is consistent with the fact that in reactions where the natural acceptor is used, the T/H ratio is more in favor of glycosynthesis. Nevertheless, STD NMR experiments did not provide any evidence that would support an alteration in donor positioning in subsite -1, the binding map being similar for *Tx*Abf and the mutants. Therefore, taking into account the k_{cat}/K_M drastic decrease, it is more likely that increased transglycosylation can be explained by thermodynamics, and more specifically regarding the transition state energy barriers associated with the formation of the glycosyl-enzyme intermediate (k_2) and/or deglycosylation step (k_3). In this case, we postulate that, for N344P and N344Y, an increase in the transition state energy barrier during the glycosylation step (ΔE_a can be approximated by -RT.ln $[(k_{cat}/K_M)_{mut} / (k_{cat}/K_M)_{wt}]$ = 1.9 and 2.6 kcal.mol^{-1}, respectively) must be mirrored by a similar

perturbation in the deglycosylation step. To explain this, we suppose that changes in the properties of the oxycarbenium intermediate transition state (TS2) in the mutants make water-mediated deglycosylation unfavorable and that, on the other hand, greater interactions provided by a sugar acceptor (relative to a single water molecule) lower the energy level of acceptor-mediated deglycosylation transition state, and increase the T/H ratio. In the mutants N344P and Y, acceptor-mediated activation (assessed by monitoring *p*NP release) is actually very low compared to the apparent increase in R_T. Therefore, it is likely that the mutations do not actually alleviate inhibition, but rather render acceptor-mediated deglycosylation advantageous to surmount the energy barrier.

In addition to the competition between water and non-water acceptor molecules, which determines the outcome of the deglycosylation step, the overall transglycosylation yield is also determined by the half-life of the product, which can be increased by reducing the enzyme's ability to hydrolyze it. To achieve this, it is important to take into account the fact that the first stage of the reaction uses an activated donor, bearing a *p*NP, which displays much higher leaving group ability than the xylobiosyl or xylotriosyl moieties present in the transglycosylation products. Considering both the k_{cat}/K_M and KIE values reported here, it appears that the mutants display diminished ability to cleave glycosidic bonds linking two sugar moieties.

5. Conclusions

In this work, we have shown that the mutants N344P and Y display a severely reduced ability to perform secondary hydrolysis, thus allowing the yield of transglycosylation products to attain a stable plateau over time. This indicates that the T/H equilibrium governing the mutant-catalyzed reactions might be modified. To our knowledge this report constitutes a first of a kind demonstration of a strategy that might be applicable to other GHs, providing that pK_a modulating residues, equivalent to N344, can be identified. Moreover, it is pertinent to wonder to what extent the mutants described in this study could be further improved. Given the fact that mutants N344P and Y lead to persistent transfuranosylation products compared to the wild-type, it might be feasible to use directed evolution to introduce other features that enhance the usefulness of *Tx*Abf as a tool for chemo-enzymatic syntheses.

Competing interests

The authors declare they have no competing interests.

Acknowledgements

PhD fellowship of B. Bissaro was supported by the Institut National de la Recherche Agronomique (INRA) [CJS]. MetaToul (Metabolomics & Fluxomics Facitilies, Toulouse, France, www.metatoul.fr) and its staff members are gratefully acknowledged for technical support and access to NMR facility. MetaToul is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, the SICOVAL, the Infrastructures en Biologie Sante et Agronomie (IBiSa, France), the Centre National de la Recherche Scientifique (CNRS) and the Institut National de la Recherche Agronomique (INRA). STD NMR experiments were performed at the 'Integrated Screening Platform of Toulouse' (PICT, IPBS, CNRS - Université de Toulouse) using equipment funded by the French Research Ministry, CNRS, Université Paul Sabatier, the Région Midi-Pyrénées and European structural funds.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2013.10.013. and as supporting information after the references list: Supporting Information of Part II

References

[1] R. Fauré, C.M. Courtin, J.A. Delcour, C. Dumon, C.B. Faulds, G.B. Fincher, S. Fort, S.C. Fry, S. Halila, M.A. Kabel, L. Pouvreau, B. Quemener, A. Rivet, L. Saulnier, H.A. Schols, H. Driguez, M.J. O'Donohue, A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls, Aust. J. Chem. 62 (2009) 533–537.

[2] J. Kadokawa, Precision polysaccharide synthesis catalyzed by enzymes, Chem. Rev. 111 (2011) 4308–4345.

[3] R. Kittl, S.G. Withers, New approaches to enzymatic glycoside synthesis through directed evolution, Carbohydr. Res. 345 (2010) 1272–1279.

[4] B. Rakić, S.G. Withers, Recent developments in glycoside synthesis with glyco- synthases and thioglycoligases, Aust. J. Chem. 62 (2009) 510–520.

[5] R.M. Schmaltz, S.R. Hanson, C.-H. Wong, Enzymes in the synthesis of glyco- conjugates, Chem. Rev. 111 (2011) 4259–4307.

[6] C. Dumon, L. Song, S. Bozonnet, R. Fauré, M.J. O'Donohue, Progress and future prospects for pentose-specific biocatalysts in biorefining, Process Biochem. 47 (2012) 346–357.

[7] B. Tefsen, A.F.J. Ram, I. van Die, F.H. Routier, Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact, Glycobiology 22 (2012) 456–469.

[8] M. Oppenheimer, A.L. Valenciano, P. Sobrado, Biosynthesis of galactofuranose in kinetoplastids: novel therapeutic targets for treating leishmaniasis and chagas' disease, Enzyme Res. 2011 (2011) 1–13.

[9] L.L. Pedersen, S.J. Turco, Cellular and molecular life sciences galactofuranose metabolism: a potential target for antimicrobial chemotherapy, Cell. Mol. Life Sci. 60 (2003) 259–266.

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

[11] N.N. Aronson, B.A. Halloran, M.F. Alexeyev, X.E. Zhou, Y. Wang, E.J. Meehan, L. Chen, Mutation of a conserved tryptophan in the chitin-binding cleft of *Serratia marcescens* chitinase A enhances transglycosylation, Biosci. Biotechnol. Biochem. 70 (2006) 243–251.

[12] H. Mori, K.S. Bak-Jensen, B. Svensson, Barley α-amylase Met53 situated at the high-affinity subsite−2 belongs to a substrate binding motif in the ß → α loop 2 of the catalytic (ß/α)8-barrel and is critical for activity and substrate specificity, Eur. J. Biochem. 269 (2002) 5377–5390.

[13] S. Armand, S.R. Andrews, S.J. Charnock, H.J. Gilbert, Influence of

the aglycone region of the substrate binding cleft of *Pseudomonas* xylanase 10A on catalysis, Biochemistry 40 (2001) 7404–7409.
[14] S.-Q. Fan, W. Huang, L.-X. Wang, Remarkable transglycosylation activity of glyco- synthase mutants of endo-D, an endo-β-*N*-acetylglucosaminidase from *Streptococcus pneumoniae*, J. Biol. Chem. 287 (2012) 11272–11281.
[15] H.-Y. Feng, J. Drone, L. Hoffmann, V. Tran, C. Tellier, C. Rabiller, M. Dion, Converting a β-glycosidase into a β-transglycosidase by directed evolution, J. Biol. Chem. 280 (2005) 37088–37097.
[16] A. Moreau, F. Shareck, D. Kluepfel, R. Morosoli, Alteration of the cleavage mode and of the transglycosylation reactions of the xylanase A of *Streptomyces lividans* 1326 by site-directed mutagenesis of the Asn173 residue, Eur. J. Biochem. 219 (1994) 261–266.
[17] T. Taira, M. Fujiwara, N. Dennhart, H. Hayashi, S. Onaga, T. Ohnuma, T. Letzel, S. Sakuda, T. Fukamizo, Transglycosylation reaction catalyzed by a class V chitinase from cycad, *Cycas revoluta*: A study involving site-directed mutagenesis, HPLC, and real-time ESI-MS, Biochim. Biophys. Acta Proteins Proteomics 1804 (2010) 668–675.
[18] M. Umekawa, W. Huang, B. Li, K. Fujita, H. Ashida, L.-X. Wang, K. Yamamoto, Mutants of *Mucor hiemalis* endo-β-*N*-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities, J. Biol. Chem. 283 (2008) 4469–4479.
[19] M.A. Frutuoso, S.R. Marana, A single amino acid residue determines the ratio of hydrolysis to transglycosylation catalyzed by β-glucosidases, Protein Pept. Lett. 20 (2013) 102–106.
[20] M. Hidaka, S. Fushinobu, Y. Honda, T. Wakagi, H. Shoun, M. Kitaoka, Structural explanation for the acquisition of glycosynthase activity, J. Biochem. 147 (2010) 237–244.
[21] Y. Honda, S. Fushinobu, M. Hidaka, T. Wakagi, H. Shoun, H. Taniguchi, M. Kitaoka, Alternative strategy for converting an inverting glycoside hydrolase into a glycosynthase, Glycobiology 18 (2008) 325–330.
[22] T. Kuriki, H. Kaneko, M. Yanase, H. Takata, J. Shimada, S. Handa, T. Takada, H. Umeyama, S. Okada, Controlling substrate preference and transglycosylation activity of neopullulanase by manipulating steric constraint and hydrophobicity in active center, J. Biol. Chem. 271 (1996) 17321–17329.
[23] R. Kuroki, L.H. Weaver, B.W. Matthews, Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 8949–8954.
[24] C. Giacomini, G. Irazoqui, P. Gonzalez, F. Batista-viera, B.M. Brena, Enzymatic synthesis of galactosyl–xylose by *Aspergillus oryzae* β-galactosidase, J. Mol. Catal. B Enzym. 20 (2002) 159–165.
[25] S. Malá, H. Dvoráková, R. Hrabal, B. Králová, Towards regioselective synthesis of oligosaccharides by use of α-glucosidases with different substrate specificity, Carbohydr. Res. 322 (1999) 209–218.
[26] M. Ribeirão, V.L. Pereira-Chioccola, D. Eichinger, M.M. Rodrigues, S. Schenkman, Temperature differences for *trans*-glycosylation and hydrolysis reaction reveal an acceptor binding site in the catalytic mechanism of *Trypanosoma cruzi trans*-sialidase, Glycobiology 7 (1997) 1237–1246.
[27] E. Bonnin, J. Vigouroux, Kinetic parameters of hydrolysis and transglycosylation catalyzed by an exo-β-(1,4)-galactanase, Enzyme Microb. Technol. 20 (1997) 516–522.
[28] G. Perugino, A. Trincone, A. Giordano, J. van der Oost, T. Kaper, M. Rossi, M. Moracci, Activity of hyperthermophilic glycosynthases is significantly enhanced at acidic pH, Biochemistry 42 (2003) 8484–8493.
[29] A. Trincone, A. Giordano, G. Perugino, M. Rossi, M. Moracci, Glycosynthase-catalysed syntheses at pH below neutrality, Bioorg. Med. Chem. Lett. 13 (2003) 4039–4042.
[30] M. Hrmova, T. Imai, S.J. Rutten, J.K. Fairweather, V. Bulone, H. Driguez, G.B. Fincher, Mutated barley (1,3)-β-D-glucan endohydrolases synthesize crystalline (1,3)-β-D- glucans, J. Biol. Chem. 277 (2002) 30102–30111.
[31] M. Pérez-Sánchez, Á. Cortés Cabrera, H. García-Martín, J.V. Sinisterra, J.I. García, M.J. Hernáiz, Improved synthesis of disaccharides with *Escherichia coli* β-galactosidase using bio-solvents derived from glycerol, Tetrahedron 67 (2011) 7708–7712.
[32] L.P. McIntosh, G. Hand, P.E. Johnson, M.D. Joshi, M. Ko, L.A. Plesniak, L. Ziser, W.W. Wakarchuk, S.G. Withers, The p*K*a of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A ^{13}C-NMR study of *Bacillus circulans* xylanase, Biochemistry 35 (1996) 9958–9966.
[33] C. Malet, A. Planas, Mechanism of *Bacillus* 1,3-1,4-β-D-glucan 4-glucanohydrolases: kinetics and pH studies with 4-methylumbelliferyl β-D-glucan oligosaccharides, Biochemistry 36 (1997) 13838–13848.
[34] K. Bartik, C. Redfield, C.M. Dobson, Measurement of the individual p*K*a values of acidic residues of hen and turkey lysozymes by two-dimensional ^{1}H NMR, Biophys. J. 66 (1994) 1180–1184.
[35] D.L. Zechel, S.G. Withers, Dissection of nucleophilic and acid–base catalysis in glycosidases, Curr. Opin. Chem. Biol. 5 (2001) 643–649.
[36] A. Vasella, G.J. Davies, M. Böhm, Glycosidase mechanisms, Curr. Opin. Chem. Biol. 6 (2002) 619–629.
[37] M.D. Joshi, G. Sidhu, J.E. Nielsen, G.D. Brayer, S.G. Withers, L.P. McIntosh, Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase, Biochemistry 40 (2001) 10115–10139.
[38] M.D. Joshi, G. Sidhu, I. Pot, G.D. Brayer, S.G. Withers, L.P. McIntosh, Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase, J. Mol. Biol. 299 (2000) 255–279.
[39] M.L. Ludwiczek, I. D'Angelo, G.N. Yalloway, J. Brockerman, M. Okon, J.E. Nielsen, N.C.J. Strynadka, S.G. Withers, L.P. McIntosh, Strategies for modulating the pH-dependent activity of a family 11 glycoside hydrolase, Biochemistry 52 (2013) 3138–3156.
[40] C. Rémond, R. Plantier-Royon, N. Aubry, E. Maes, C. Bliard, M.J. O'Donohue, Synthesis of pentose-containing disaccharides using a thermostable α-L-arabinofuranosidase, Carbohydr. Res. 339 (2004) 2019–2025.
[41] C. Rémond, R. Plantier-Royon, N. Aubry, M.J. O'Donohue, An original chemoenzymatic route for the synthesis of β-D-galactofuranosides using an α-L-arabinofuranosidase, Carbohydr. Res. 340 (2005) 637–644.
[42] R. Euzen, G. Lopez, C. Nugier-Chauvin, V. Ferrières, D. Plusquellec, C. Rémond, M. O'Donohue, A chemoenzymatic approach for the synthesis of unnatural disaccharides containing D-galacto or D-fucofuranosides, Eur. J. Org. Chem. 2005 (2005) 4860–4869.
[43] G. Lopez, C. Nugier-Chauvin, C. Rémond, M. O'Donohue, Investigation of the specificity of an α-L-arabinofuranosidase using C-2 and C-5 modified α-L-arabinofuranosides, Carbohydr. Res. 342 (2007) 2202–2211.
[44] F. Arab-Jaziri, B. Bissaro, M. Dion, O. Saurel, D. Harrison, F. Ferreira, A. Milon, C. Tellier, M.J. O'Donohue, R. Fauré, Engineering transglycosidase activity into a GH51 α-L-arabinofuranosidase, New Biotechnol. 30 (2013) 536–544.
[45] F. Arab-Jaziri, B. Bissaro, C. Tellier, M. Dion, R. Fauré, M.J. O'Donohue, Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α-L-arabinofuranosidase using a directed evolution approach, Unpublished results.
[46] C. Rémond, M. Ferchichi, N. Aubry, R. Plantier-Royon, C. Portella, M.J. O'Donohue, Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α-L-arabino-furanosidase, Tetrahedron Lett. 43 (2002) 9653–9655.
[47] L. Marmuse, M. Asther, E. Fabre, D. Navarro, L. Lesage-Meessen, M. Asther, M. O'Donohue, S. Fort, H. Driguez, New chromogenic substrates for feruloyl esterases, Org. Biomol. Chem. 6 (2008) 1208–1214.
[48] H. Gruppen, R.A. Hoffmann, F.J. Kormelink, A.G. Voragen, J.P. Kamerling, J.F. Vliegenthart, Characterisation by ^{1}H NMR spectroscopy of enzymically derived oligosaccharides from alkali-

extractable wheat-flour arabinoxylan, Carbohydr. Res. 233 (1992) 45–64.
[49] J.J. Ordaz-Ortiz, F. Guillon, O. Tranquet, G. Dervilly-Pinel, V. Tran, L. Saulnier, Specificity of monoclonal antibodies generated against arabinoxylans of cereal grains, Carbohydr. Polym. 57 (2004) 425–433.
[50] M. Vrsanská, W. Nerinckx, M. Claeyssens, P. Biely, An alternative approach for the synthesis of fluorogenic substrates of *endo*-ß-(1 → 4)-xylanases and some applications, Carbohydr. Res. 343 (2008) 541–548.
[51] R.A. Hoffman, B.R. Leeflang, M.M.J. de Barse, J.P. Kamerling, J.F.G. Vliegenthart, Characterisation by ^{1}H-NMR spectroscopy of oligosaccharides, derived from arabinoxylans of white endosperm of wheat, that contain the elements → 4) [α-L-Ara*f*(1 → 3)]-ß-D-Xyl*p* (1 → or → 4)[α-L-Ara*f*(1 → 2)][α-L-Ara*f*(1 → 3)]- ß-D-Xyl*p*(1→, Carbohydr. Res. 221 (1991) 63–81.
[52] T. Debeche, N. Cummings, I. Connerton, P. Debeire, M.J. O'Donohue, Genetic and biochemical characterization of a highly thermostable α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*, Appl. Environ. Microbiol. 66 (2000) 1734–1736.
[53] H.E. Gottlieb, V. Kotlyar, A. Nudelman, NMR chemical shifts of common laboratory solvents as trace impurities, J. Org. Chem. 3263 (1997) 7512–7515.
[54] P.K. Glasoe, F.A. Long, Use of glass electrodes to measure acidities in deuterium oxide, J. Phys. Chem. 64 (1960) 188–190.
[55] R.J. Vietor, R.A. Hoffmann, S.A.G.F. Angelino, A.G.J. Voragen, J.P. Kamerling,
J.F.G. Vliegenthart, Structures of small oligomers liberated from barley arabinoxylans by endoxylanase from *Aspergillus awamori*, Carbohydr. Res. 254 (1994) 245–255.
[56] H. Ferré, A. Broberg, J.O. Duus, K.K. Thomsen, A novel type of arabinoxylan arabino- furanohydrolase isolated from germinated barley analysis of substrate preference and specificity by nano-probe NMR, Eur. J. Biochem. 267 (2000) 6633–6641.
[57] H. Pastell, P. Tuomainen, L. Virkki, M. Tenkanen, Step-wise enzymatic preparation and structural characterization of singly and doubly substituted arabinoxylo- oligosaccharides with non-reducing end terminal branches, Carbohydr. Res. 343 (2008) 3049–3057.
[58] L. Pitkänen, P. Tuomainen, L. Virkki, M. Tenkanen, Molecular characterization and solution properties of enzymatically tailored arabinoxylans, Int. J. Biol. Macromol. 49 (2011) 963–969.
[59] M.A.S. Correia, K. Mazumder, J.L.A. Brás, S.J. Firbank, Y. Zhu, R.J. Lewis, W.S. York,
C.M.G.A. Fontes, H.J. Gilbert, Structure and function of an arabinoxylan-specific xylanase, J. Biol. Chem. 286 (2011) 22510–22520.
[60] M. Mayer, B. Meyer, Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor, J. Am. Chem. Soc. 123 (2001) 6108–6117.
[61] D. Shallom, V. Belakhov, D. Solomon, G. Shoham, T. Baasov, Y. Shoham, Detailed kinetic analysis and identification of the nucleophile in α-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase, J. Biol. Chem. 277 (2002) 43667–43673.
[62] D.L. Zechel, S.G. Withers, Glycosidase mechanisms: anatomy of a finely tuned catalyst, Acc. Chem. Res. 33 (2000) 11–18.
[63] M.L. Sinnott, Catalytic mechanism of enzymic glycosyl transfer, Chem. Rev. 90 (1990) 1171–1202.
[64] T. Bravman, G. Zolotnitsky, V. Belakhov, G. Shoham, B. Henrissat, T. Baasov, Y. Shoham, Detailed kinetic analysis of a family 52 glycoside hydrolase: a ß-xylosidase from *Geobacillus stearothermophilus*, Biochemistry 42 (2003) 10528–10536.
[65] F. Arab-Jaziri, B. Bissaro, S. Barbe, O. Saurel, H. Débat, C. Dumon, V. Gervais, A. Milon,
I. André, R. Fauré, M.J. O'Donohue, Functional roles of H98 and W99 and ß2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*, FEBS J. 279 (2012) 3598–3611.
[66] G. Paës, L.K. Skov, M.J. O'Donohue, C. Rémond, J.S. Kastrup, M. Gajhede, O. Mirza, The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit, Biochemistry 47 (2008) 7441–7451.
[67] D.E. Koshland Jr., Stereochemistry and the mechanism of enzymatic reactions, Biol.
Rev. Camb. Philos. Soc. 28 (1953) 416–436.
[68] L. Bu, M.F. Crowley, M.E. Himmel, G.T. Beckham, Computational investigation of pH dependence on loop flexibility and catalytic function in glycoside hydrolases, J. Biol. Chem. 288 (2013) 12175–12186.
[69] V. Nahoum, G. Roux, V. Anton, P. Rougé, A. Puigserver, H. Bischoff, B. Henrissat, F. Payan, Crystal structures of human pancreatic α-amylase in complex with carbohydrate and proteinaceous inhibitors, Biochem. J. 346 (2000) 201–208.

Supporting Information of Part II: "Mutation of a pH-modulating residue in a GH51 α-L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products".

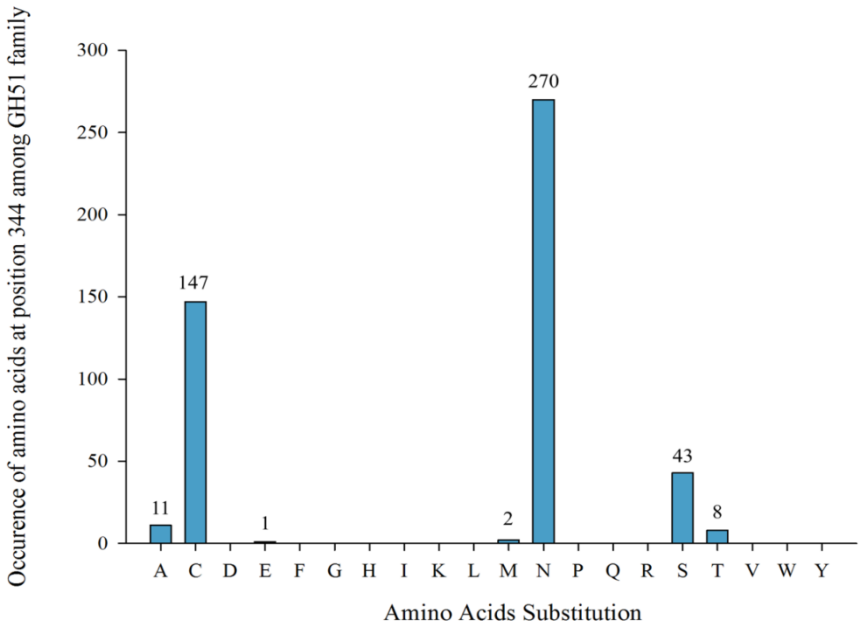


Fig. SI 1. Occurrence of substituting amino acids at position 344 (*Tx*Abf numbering) among GH51 family (488 sequences).

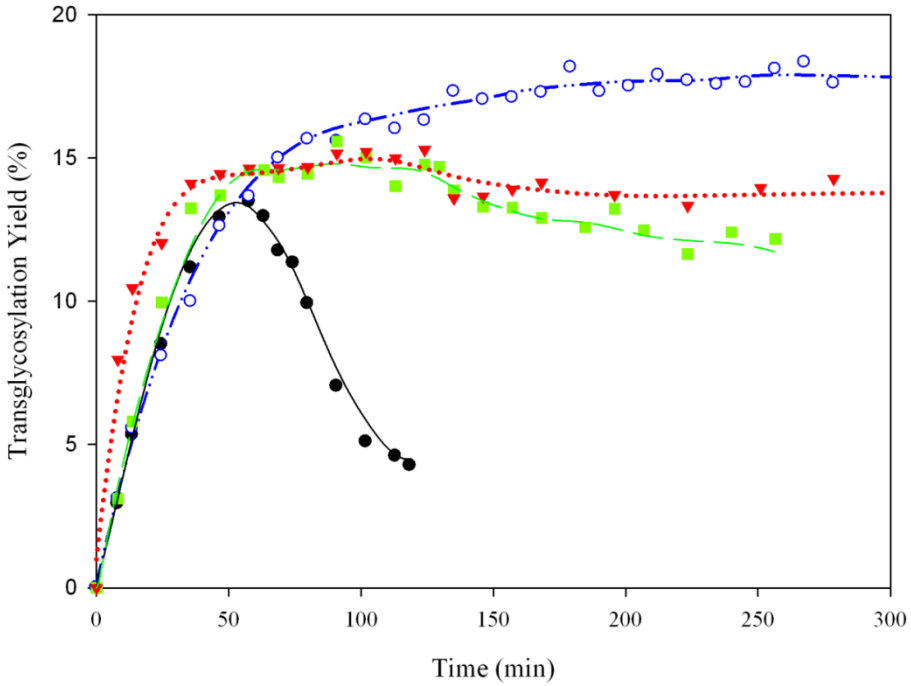


Fig. SI 2. Time-course NMR monitoring of Bn α-L-Ara*f*-(1,2)-α-D-Xyl*p* transglycosylation products obtained by reactions catalyzed by *Tx*Abf (●, black solid line), N344G (■, green dashed line), N344P (▼, red dotted line) and N344Y (○, blue dash-dot-dot line), with *p*NP-α-L-Ara*f* (5 mM) as donor and Bn-α-D-Xyl*p* (10 mM) as acceptor. Assays were conducted in 25 mM deuterated acetate buffer (pD 5.9) at 45 °C.

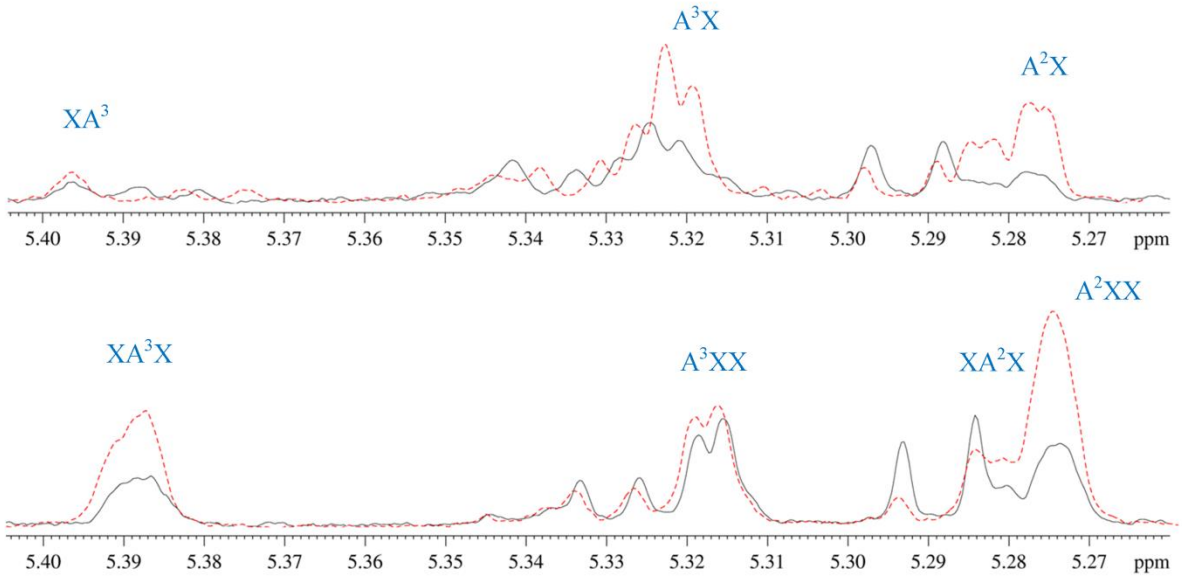


Fig. SI 3. Zoom on NMR spectra anomeric region showing the different transglycosylation products obtained via transglycosylation reaction between xylobiose (A) or xylotriose (B) and *p*NP-α-L-Ara*f* catalyzed by *Tx*Abf (solid black line) or N344Y (dashed red line).

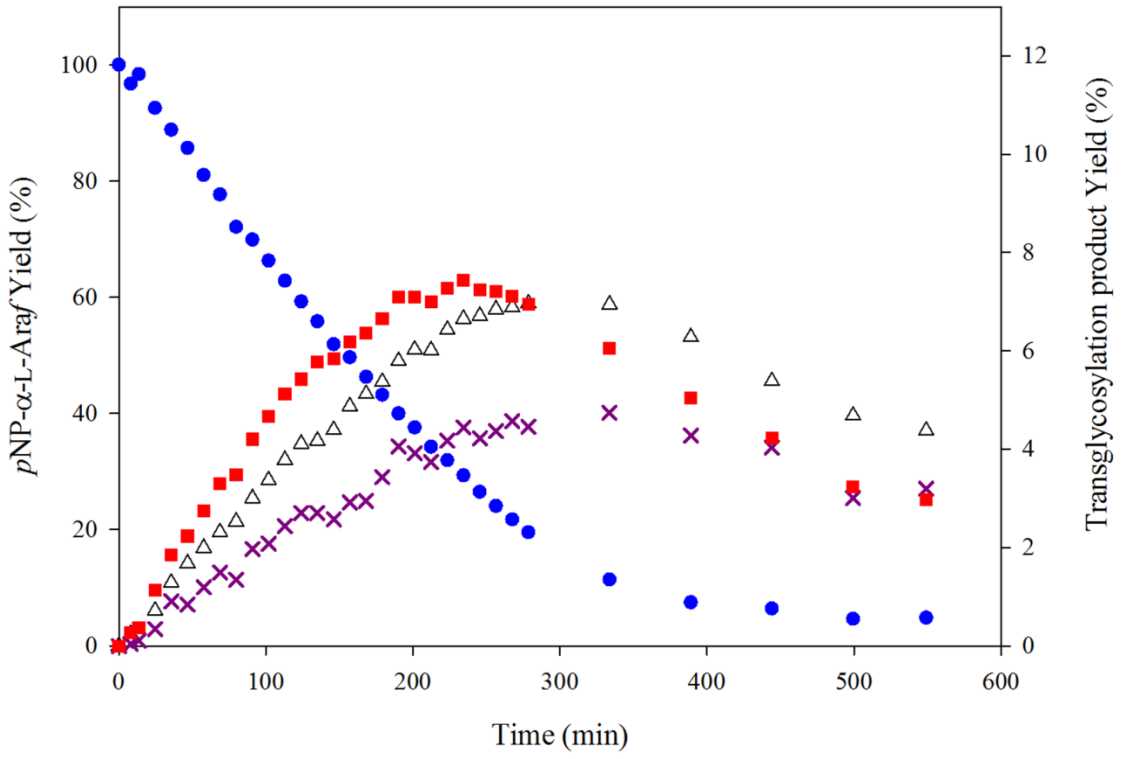


Fig. SI 4. Time-course NMR monitoring of the three transglycosylation products XA^{3}X (5.40 ppm, ■), A^{3}XX (5.32 ppm, △) and A^{2}XX (5.27 ppm, ×) synthesis catalyzed by *Tx*Abf (6 nM), with *p*NP-α-L-Ara*f* (5 mM) as donor (●) and xylotriose (10 mM) as acceptor. Assays were conducted in 25 mM deuterated acetate buffer (pD 4.65) at 45 °C.

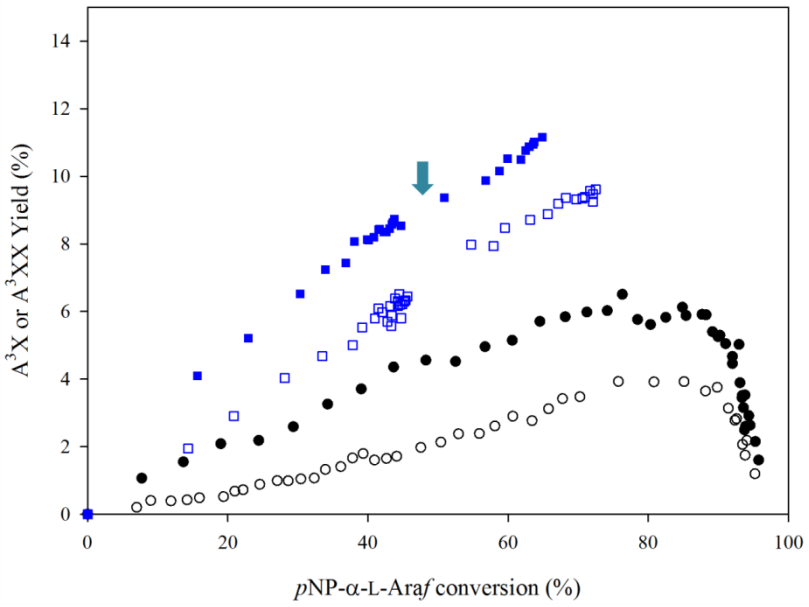


Fig. SI 5. A^3XX (5.32 ppm) transglycosylation product yield (% of 5 mM) as a function of *p*NP-α-L-Ara*f* conversion (% of 5 mM) extracted from time-course NMR monitoring of reactions with *p*NP-α-L-Ara*f* (5 mM) as donor and xylobiose or xylotriose (10 mM) as acceptor catalyzed by *Tx*Abf (8 nM) (● or ○, respectively) and N344Y (400 nM) (■ or □, respectively). The cluster observed around X = 45% for N344Y-catalyzed reactions corresponds to the plateau phase and the blue arrow to mutant addition (400 nM) once the steady state was reached. Assays were conducted in 25 mM deuterated acetate buffer (pD 5.9) at 45 °C.

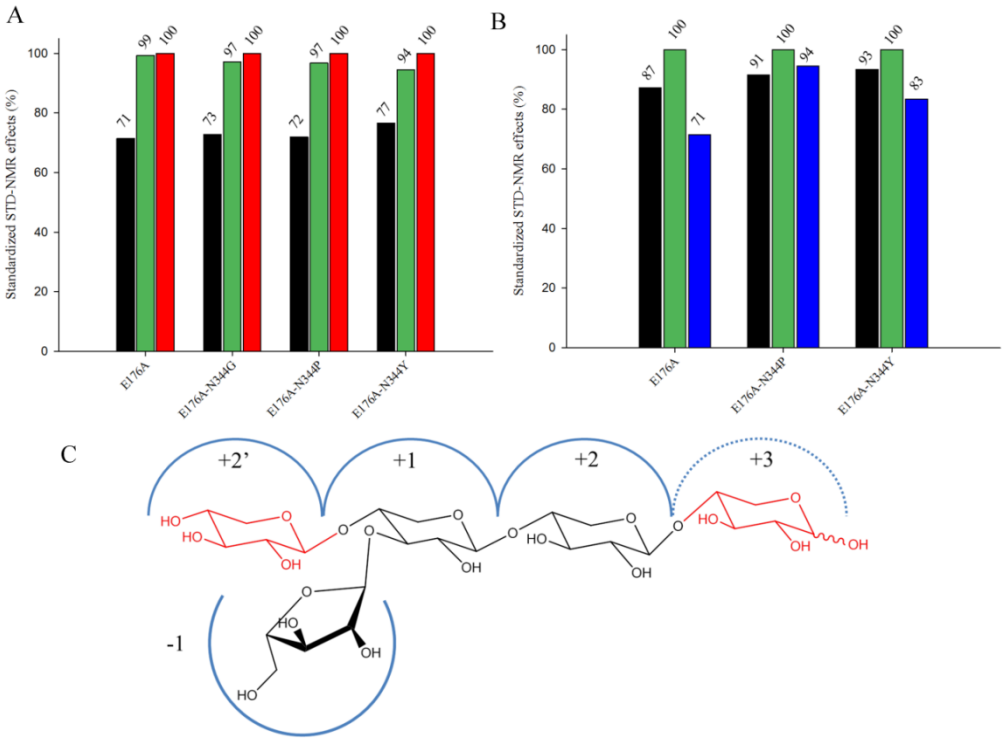


Fig. SI 6. Intra-sugar standardized STD NMR effects, for *Tx*Abf‡, N344G‡, N344P‡ and N344Y‡, of the L-Ara*f* moiety protons of A^3X (A): H-1 (black), H-2/H-4 (green), H-3 (red) and of XA^3XX; (B): H-1 (black), H-2 (green) and H-4 (blue). STD effects are expressed as relative percentages of the STD effect of H-3 (red) for A^3X (A) and H-2 (green) for XA^3XX (B). (C) Schematic representation of A^3X and XA^3XX in *Tx*Abf active site.

Table SI 1. Transglycosylation features (Y_{max}, R_T) calculated from time-course NMR kinetics of *Tx*Abf and N344G, P or Y mutants with *p*NP-α-L-Ara*f* (5 mM) as donor and xylobiose (10 mM) or xylotriose (10 mM) as acceptors, at 45 °C.

Enzyme	Features	Xylobiose (10 mM)			Xylotriose (10 mM)		
	Transglycosylation product	XA^3	A^3X	A^2X	XA^3X	A^3XX	A^2XX
	δ (ppm)	5.40	5.32	5.28	5.39	5.32	5.27
***Tx*Abf**	Y_{max} (%)[a]	1.3	6.5	2.1	2.9	3.9	5.8
	R_T[b]	nd[c]	8.6	2.6	nd	3.5	5.8
N344G	Y_{max} (%)	0.9	6.8	2.1	4.1	5.6	7.6
	R_T	2.0	14.2	4.9	7.2	10.0	14.7
	$Y_{max\ (mutant)}/Y_{max\ (wt)}$	0.74	1.04	1.01	1.45	1.42	1.29
	$R_{T\ (mutant)}/R_{T\ (wt)}$	nd	1.65	1.87	nd	2.91	2.53
N344P	Y_{max} (%)	0.9	7.4	2.1	4.9	6.7	7.0
	R_T	1.8	15.5	4.3	6.8	11.0	14.0
	$Y_{max\ (mutant)}/Y_{max\ (wt)}$	0.68	1.14	1.01	1.72	1.71	1.20
	$R_{T\ (mutant)}/R_{T\ (wt)}$	nd	1.81	1.68	nd	3.19	2.41
N344Y	Y_{max} (%)	1.7	8.7	4.5	7.2	6.5	10.5
	R_T	2.5	20.2	nd	15.1	13.9	26.7
	$Y_{max\ (mutant)}/Y_{max\ (wt)}$	1.31	1.34	2.17	2.50	1.66	1.79
	$R_{T\ (mutant)}/R_{T\ (wt)}$	nd	2.35	nd	nd	4.03	4.60
N344K	Y_{max} (%)	2.8	12.4	6.1	9.6	10.4	14.2
	R_T	2.5	22.7	24.8	19.9	13.7	27.5
	$Y_{max\ (mutant)}/Y_{max\ (wt)}$	2.18	1.90	2.95	3.37	2.65	2.43
	$R_{T\ (mutant)}/R_{T\ (wt)}$	nd	2.65	9.6	nd	3.98	4.74

[a] Y_{max} denotes the maximum transglycosylation product yield (as a percentage of 5 mM), which can be reached at different times depending on the regioisomer.

[b] Transfer rates (R_T, in µmol of transglycosylation product/100 µmol of donor substrate) were determined from plots of transglycosylation yields (Y) *versus* donor conversion (X) for X values inferior to 40%.

[c] R_T were not determined when R^2 values associated to the Y = f(X) plot were inferior to 0.95.

Table SI 2. Solvent Kinetic Isotope Effects for *Tx*Abf, N344P and Y mutants.

Enzyme	Relative Activity (%)[a]			Mean Ratio	Error	R-fold increase[b]
	A	B	C			
***Tx*Abf**	75	71.6	71	72.5	1.8	1
N344P	102	111.1	102	105.1	5.2	1.45
N344Y	115	114.8	121	116.7	3.4	1.61

[a] Relative Activity = $SA_{(without\ xylobiose)}/SA_{(with\ xylobiose)}$ were determined in triplicates
[b] R-fold increase = $Mean\ Ratio_{(mutant)}/Mean\ Ratio_{(wt)}$

Table SI 3. Determination of the relative activity of the *Tx*Abf and N344P and Y mutants toward *p*NP-α-L-Ara*f* (5 mM) in absence or presence of xylobiose (10 mM), at 45 °C.

Enzyme	$SA_{max,\ H}$ (IU/mg) mean value	$SA_{max,\ H}$ error	$SA_{max,\ D}$ (IU/mg) mean value	$SA_{max,\ D}$ error	k_H/k_D	k_H/k_D error	k_H/k_D Relative increase (%)
***Tx*Abf**	383.06	12.61	203.39	24.29	1.88	0.29	0
N344P	53.21	2.07	27.27	1.79	1.95	0.20	3.6
N344Y	4.27	0.28	1.66	0.11	2.58	0.34	36.8

II-3. Does N344Y constitute a good protein scaffold?

The thorough analysis conducted on N344 mutations led us to wonder whether these mutations, affecting the global catalytic efficiency, could be suitable starting points for further improvements of the transglycosylation activity through molecular evolution. For this, a random library was created with *Tx*Abf-N344Y as DNA template, and served to set up the overall protocol of library design and screening methodology. For this library (2.6 mutation/kB), 50,000 *E.coli* clones expressing *Tx*Abf mutants were screened according to the two-step methodology described by Koné et al. (2009). The selection was based on the calculation of a color ratio obtained from blue intensities developed by clones in presence or absence of acceptor (XOS) in the medium. Among these 50,000 clones, 200 were selected and assessed for their capacity to consume *p*NP-α-L-Ara*f* in presence or absence of XOS thanks to a purpose-made automated kinetics protocol using a liquid handling robot (TECAN), and leading to an *in vitro* activity ratio. Unfortunately, none of the clones displayed an activity ratio > 100%. Sequencing of the 15 "best" clones revealed that 9 of them (60%) were actually parental enzymes (*Tx*Abf-N344Y), while true mutants displayed mutations at the enzyme surface and are thought to not be relevant substitutions (data not shown). This screening failure might be explained by one of the following hypotheses in decreasing order of credibility:

(1) The screening strategy was not adapted, since the *in vivo* detection of acceptor-activated mutants is actually a tricky task. On one side, the actual software version is probably not efficient enough to reach this goal and on the other side XOS-mediated inhibition makes the detection of apparent activation even more complicated. This is why for the following random libraries the *in vivo* screening was only employed for the detection of hydrolytically-crippled mutants (i.e. only first step in presence of X-α-L-Ara*f* donor).

(2) N344Y represents an evolutionary dead-end or unfortunately none additional positive mutation was created during this random mutagenesis round.

With hindsight, it appears that it was not so relevant to perform a random mutagenesis on this protein background since the catalytic efficiency is initially well crippled, notably through a strong decrease in k_{cat}. Such impact means that the interaction network involved in TS stabilization is already dramatically affected and further alterations may lead to a too important mutation load creating non-functional proteins. In the light of findings acquired later on (described in part III and IV), recombination with acceptor subsite mutations could have been a better option, but were not available at that moment. Unlike N344 mutants, L352M constituted a more promising background protein for a 2nd round of random mutagenesis (Part V) because only the L-Ara*f* moiety recognition is essentially altered rather than the catalytic power, as showed by large K_M increase but only slight k_{cat} decrease (TS2 not strongly affected).

Part III

Computational approach for H/T modulation and synergistic effects with random mutations

(Article in preparation)

In the quest to tip the H/T balance in favour of transglycosylation, it has been demonstrated that certain mutations in the -1 subsite can drastically diminish the global catalytic efficiency of *Tx*Abf. One such mutation, N344Y, has been extensively described earlier in this manuscript (Part II), whereas a second one, L352M, is described in terms of basic kinetic parameters in the following study. Considering the properties of this mutant, it appears as a rather interesting protein background for the further construction of an enzyme displaying target properties. For this the sequence encoding the mutant L352M was used as parental DNA for recombination. The details of this work and the characterization of the resulting mutants are described in more detail in Part IV.

One clear piece of knowledge that can be gleaned from the literature (c.f. bibliographic review) is that the improvement of acceptor molecule recognition is essential to favour transglycosylation and/or to hamper the access of water to the active site. In part I, a semi-rational approach developed was used to probe three acceptor subsite residues (N216, W248 and W302), but this work was mainly focused on understanding the role of these amino acids in the anomeric control of regioselectivity. In the present study, described hereafter, a different approach was adopted in order to identify mutations that would increase acceptor recognition. To achieve this goal, an *in silico* strategy was deployed using an algorithm, called *BindScan* (presented at the 26th International Carbohydrate Symposium, Madrid, 2012), which has been developed by Dr. Xevi Biarnés. Since *BindScan* is new and was developed with other objectives in mind, the aim of our study was to investigate how *BindScan* would perform when applied to the specific study of the development of stronger natural acceptor recognition in *Tx*Abf to shift further the H/T balance.

Predicting Mutational "Hot-spots" for Hydrolysis/Transglycosylation Modulation in *Tx*Abf using *BindScan* Algorithm

Bastien Bissaro[1,2,3], Xevi Biarnés[4], Pierre Monsan[1,2,3,5], Régis Fauré[1,2,3], Michael J. O'Donohue[1,2,3*] and Antoni Planas[4*]

[1] *Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France*
[2] *INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France*
[3] *CNRS, UMR5504, F-31400 Toulouse, France*
[4] *Laboratory of Biochemistry Bioengineering Department, Institut Químic de Sarrià, Universitat Ramon Llull, Barcelona, Spain*
[5] *Toulouse White Biotechnology, UMS INRA/INSA 1337, UMS CNRS/INSA 3582, 3 Rue des Satellites 31400 Toulouse, France*

*** Correspondance:**

M. J. O'Donohue, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, 135 Avenue de Rangueil, 31077 Toulouse cedex 4, France; Fax +33 5 6155 9400; Tel: +33 5 6155 9428; E-mail: michael.odonohue@insa-toulouse.fr

A. Planas, Laboratory of Biochemistry Bioengineering Department, Institut Químic de Sarrià, Universitat Ramon Llull, Barcelona, Spain; Fax +34 93 205 6266; Tel: +34 93 267 2000; E-mail: antoni.planas@iqs.edu

Keywords: computational enzyme engineering, glycoside hydrolase, synergic effects, enzyme mechanism, furanoses/pentoses

Running title: "Hot spots" *in silico* prediction for H/T modulation

Abbreviations: Abfs, α-L-arabinofuranosidases; AXOS, arabinoxylo-oligosaccharides; GH, glycoside hydrolase; ITC, isothermal titration calorimetry; *Tx*Abf, α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*; T/H, transglycosylation/hydrolysis ratio; TS, transition state; X, donor substrate conversion rate; XOS, xylo-oligosaccharides; D-Xyl*p*, D-xylopyranosyl; Y, yield

1. Introduction to the "*BindScan*" method

The algorithm *BindScan* is fruit of developmental research that has been performed by Dr. Xevi Biarnés and Dr. Antoni Planas in the "Institut Quimic de Sarria" of the University Ramon Llull of Barcelona. The *in silico* analysis described in this work was thus performed by Dr. **X. Biarnés**, in the framework of a collaboration between Prof. A. Planas' team in Barcelona and Dr. M. O'Donohue's group (one component of the CIMEs team led by Prof. M. Remaud-Siméon) in Toulouse.

BindScan is a computational algorithm initially developed to provide a rational approach for focusing experimental directed evolution of enzymes aimed at introducing new substrate specificities. ***BindScan* virtually identifies positions along a given protein sequence that, upon mutation, are likely to affect the binding of a given ligand.** The algorithm can detect three different kinds of mutational outcomes: positive and negative ones that either increase or decrease the binding affinity and neutral ones that have no effect on binding affinity. To generate results using *BindScan,* it is necessary to provide an initial data set that describes the 3D structure of protein bound to a ligand, this being the result of structural analysis (e.g. crystallographic data), or of standard molecular modeling. Initially, *BindScan* was devised to detect amino acids along a protein sequence that upon mutation would be likely to allow the binding of a new ligand that is different from, but structurally-related to, the ligand present in the initial structural data set. In such cases, these positions are clearly identified in the algorithm as those showing a significant positive effect. However, when the target ligand is rather similar to the initial one, the interpretation of the *BindScan* results will strongly depend on the starting structure, in particular on the starting geometry provided for the ligand. For instance, some positions may have a negative effect upon binding with the ligand in a given orientation, but may become positive when the orientation of the ligand is changed. In such cases, *BindScan* may potentially be used to introduce new substrate regioselectivities.

In the present study, *BindScan* was employed to detect positions along the *Tx*Abf sequence that upon mutation might increase the acceptor recognition and thus favour the transglycosylation reaction over the dominant hydrolytic activity. Since *BindScan* is only able to detect ligand binding effects, the study was designed to focus **on the deglycosylation step, in which the xylotriose acceptor binds to the covalent β-L-arabinofuranosyl-enzyme intermediate** (i.e. *Tx*Abf-β-L-Ara*f*). The goal of the study was thus to scope for mutations of critical amino acids that appear to be well-positioned to influence xylotriose binding, with the expected outcome being the identification of mutations that would either affect access and binding of water to the active site (although the latter cannot be directly tested with *BindScan*) or increase binding of xylotriose acceptor. An unknown factor in this study was the actual relationship between acceptor affinity and transglycosylation rate. Indeed, current knowledge indicates that there is a causal relationship, but the exact way in which

this governs the H/T equilibrium is unclear.

Of note, the *BindScan* methodology and the implementation of this novel computational algorithm to identify enzyme positions sensible to the binding of new substrates have not been published yet.

2. Results

2.1. *In silico* predictions via *BindScan* algorithm

2.1.1. Strategy

The fact that no structural data describing the **glycosyl-enzyme intermediate** *Tx*Abf-β-L-Ara*f* is available meant that it was necessary to construct a model. The model was built starting from the X-ray structure of *Tx*Abf-E176Q in complex with the substrate L-arabinofuranosyled xylotetraose XA^3XX (2VRQ). In this structure, an important conformational change has been described in the β2α2 loop surrounding the active site upon ligand binding (Paës et al., 2008; Arab-Jaziri et al., 2012). First, the substrate was removed from this structure (2VRQ), and the E176Q mutant was reverted to glutamic acid. Importantly, the closed loop configuration was maintained during these operations, since this was assumed to be the catalytically operational geometry of the loop. Next, the L-Ara*f*-nucleophile adduct geometry was obtained from the crystallographic structure of homologous enzyme, *Gs*Abf in the form of β-L-Ara*f*-enzyme intermediate (PDB code 1PZ2) (Hovel et al., 2003). *Gs*Abf and *Tx*Abf structures were superimposed, using all of the known active site residues as reference points. The β-L-Ara*f* unit was then directly transferred from the *Gs*Abf into *Tx*Abf structure. It is noteworthy the positions of the key active site residues in *Tx*Abf maintain their positions irrespective of whether the β2α2 loop adopts an open or closed conformation (Fig. SI 1).

Once the structural model of *Tx*Abf-β-L-Ara*f* had been established, a **xylotriose molecule was incorporated** using *in silico* docking methods (Trott & Olson, 2010), keeping all protein side chains and the L-Ara*f* unit in fixed positions during docking. All internal rotamers of xylotriose were freely allowed during docking and the search space was defined as an orthorhombic box around the active site. After the docking simulations, the binding mode of xylotriose showing the strongest binding affinity to the *Tx*Abf-β-L-Ara*f* intermediate spans subsites +1, +2 and +3. This binding mode indicates a preference towards the transglycosylation at the non-reducing end of xylotriose. Other orientations of xylotriose were also obtained with weaker binding affinities, in particular those spanning subsites +2', +1, +2. From the pool of docked structures, four representative binding modes of xylotriose were selected amongst the ones with lower binding affinity: firstly, two structures in which the hydroxyl group at *O*-3 or *O*-2 position of the **central** D-Xyl*p* unit was orientated towards the anomeric center of the covalent *Tx*Abf-β-L-Ara*f* intermediate (Fig. SI 2A-B), leading to transglycosylation products XA^3X (run 1) and XA^2X (run 2), respectively, according to the AXOS nomenclature (Fauré et al., 2009).

Secondly, two equivalent structures in which the *O*-3 or *O*-2 positions of the **non-reducing** end D-Xyl*p* unit were actually orientated towards the β-L-Ara*f* anomeric center (Fig. SI 2C-D), meant to lead to the formation of A^3XX (run 3) and A^2XX (run 4), respectively. When selecting these structures, special care was taken to assure a proper orientation of the xylotriose substrate towards the catalytic machinery of the glycosyl-enzyme intermediate (average distance D-Xyl*p*-O_x···*C*-1-(β-L-Ara*f*): 3 Å; D-Xyl*p*-O_x···E176 (acid/base): 3 Å, with x = 2 or 3).

2.1.2. BindScan runs

The results obtained using *BindScan* runs are summarized in profile graphs (Figs. 1 and Fig. SI 3-5). These reveal two different types of amino acid positions: (1) "**hot-spots**", which are variants that show energy values clearly below the reference region, delimited by dotted lines; and (2) "**cold-spots**" which are variants whose energy values are clearly above the reference region (Table 1). Nevertheless, hot- and cold-spots are relatively rare, as one might expect, with most positions being neutral with respect to acceptor binding.

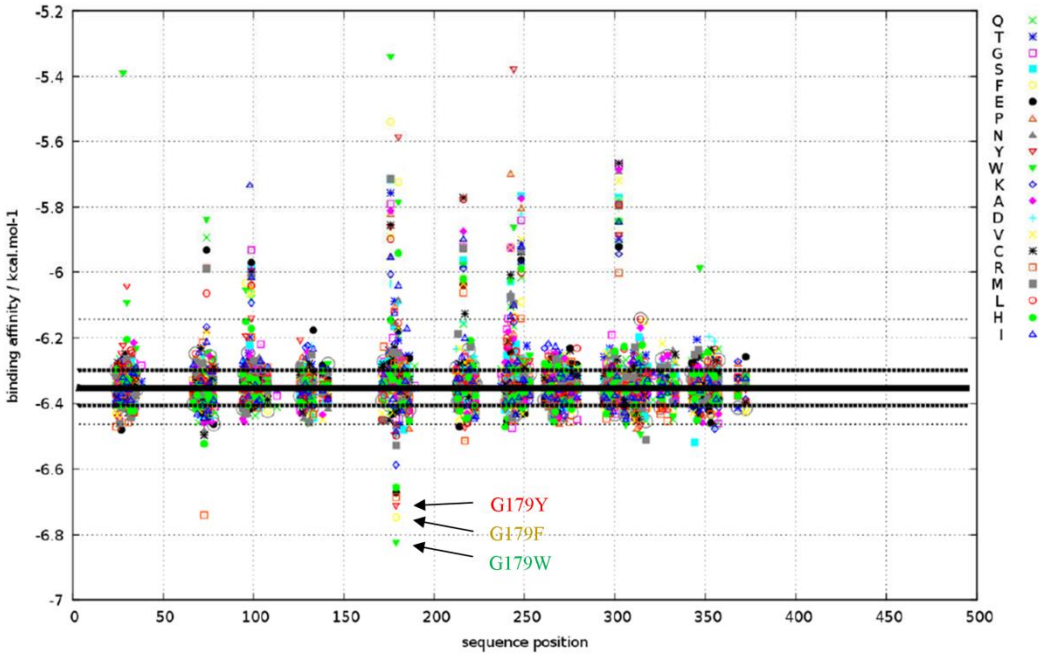


Fig. 1. *BindScan* run 1. (Binding energy-mutation) profile of the covalent intermediate *Tx*Abf-β-L-Ara*f* in complex with xylotriose substrate in the central *O*-3 orientation for XA^3X product formation. The x-axis represents each single amino acid position along the protein query sequence. Each position is virtually mutated, one by one, into each of the 20 natural amino acids. For each protein variant, the binding affinity for the substrate is evaluated by means of a computationally derived scoring function. This binding affinity is represented as an energy term in the y-axis. Each mutant is represented following using a coloring scheme that is depicted in the right legend. The reference binding affinity for the wild-type enzyme is determined as an average of all the energy values obtained for all the natural variants on each position along the protein sequence. This reference value is represented as a thick horizontal line in the graph. The standard deviation of these values and the minimum and maximum values are represented as horizontal dotted lines. The region within the thin dotted lines is the confidence interval, indicating that only positions deviating notably from this region are meaningful.

The results of *BindScan* run 1 (Fig. 1), which targeted the improvement of xylotriose binding with a view to forming the product XA^3X (Fig. SI 2A), revealed just one significant hot-spot (G179),with a ΔΔG of binding of –0.5 kcal·mol^{-1} for the G179W mutation relative to the wt enzyme, and several cold-spots (Table 1) with raises in ΔΔG of binding of up to +1 kcal·mol^{-1}. A similar result was obtained in *BindScan* run 2 (Fig. SI 3), in which the formation of XA^2X was targeted (Fig. SI 2B). Indeed, the principal hotspot is the same (G179), this amino acid being located in the +1 acceptor subsite (Fig. 2). Overall, the concordance of these results indicates that certain mutations at this position should increase the binding of the xylotriose substrate and allow transfer of the L-Ara*f* donor moiety onto the central D-Xyl*p* unit, irrespective of the orientation of the latter.

Table 1. Hot-spot and Cold-spot identified in the course of *BindScan* runs

		Hot-Spot	Cold-Spot									
Residues			C74	W99	E176	W178	C180	N216	Y242	V244	W248	W302
Location (subsite)			+2’	-1/+2’	Cat.	+1	+2’	+2	-1/+1	+2	+1	-1/+1
Conservation (%)			51	97	100	51		26	100	20	17	98
Common to the 4 runs		G179			×			×	×		×	
Specific to	Run 1 (XA^3X)	-	×	×			×					×
	Run 2 (XA^2X)	G312E	×	×		×	×					
	Run 3 (A^3XX)	C180				×				×		×
	Run 4 (A^2XX)	C74		×		×				×		×

* The conservation analysis was realized with a multiple sequence alignment, using a total of 488 bacterial GH51 Abfs sequences. Percentages are calculated by taking into account only sequences for which a residue is found at an equivalent location (i.e. no indel) and therefore does not necessarily involve all the 488 sequences for each residue.

The introduction large side chains in the G179 position fills an empty space left in the substrates binding cavity. At the same time, the G179W variant establishes new stacking interactions with the substrate, in particular with the ring located at +1 subsite. This can explain the increase of roughly 0.5 kcal.mol^{-1} in substrate binding upon G179W mutation observed in the *BindScan* profile (Fig. 1). On top of that, for the particular case of the G179W variant, it is expected that the hydrophobicity of the active site will increase, and this may eventually disfavor the association of water molecules in the active site that may lead to hydrolysis. Of note, a second hot-spot revealed in run 2 is amino acid G312 (Fig. SI 3). The best variant, G312E, would establish a new hydrogen-bond interaction with the substrate, in particular with the ring located at +2’ subsite (Fig. 2) and could explain the predicted

increase of roughly 0.2 kcal.mol^{-1} for substrate binding. The G312E mutation may potentially lead to enzymatic variants that are able to transfer the L-Ara*f* moiety onto xylotriose acceptor to obtain XA^{2}X tetrasaccharide.

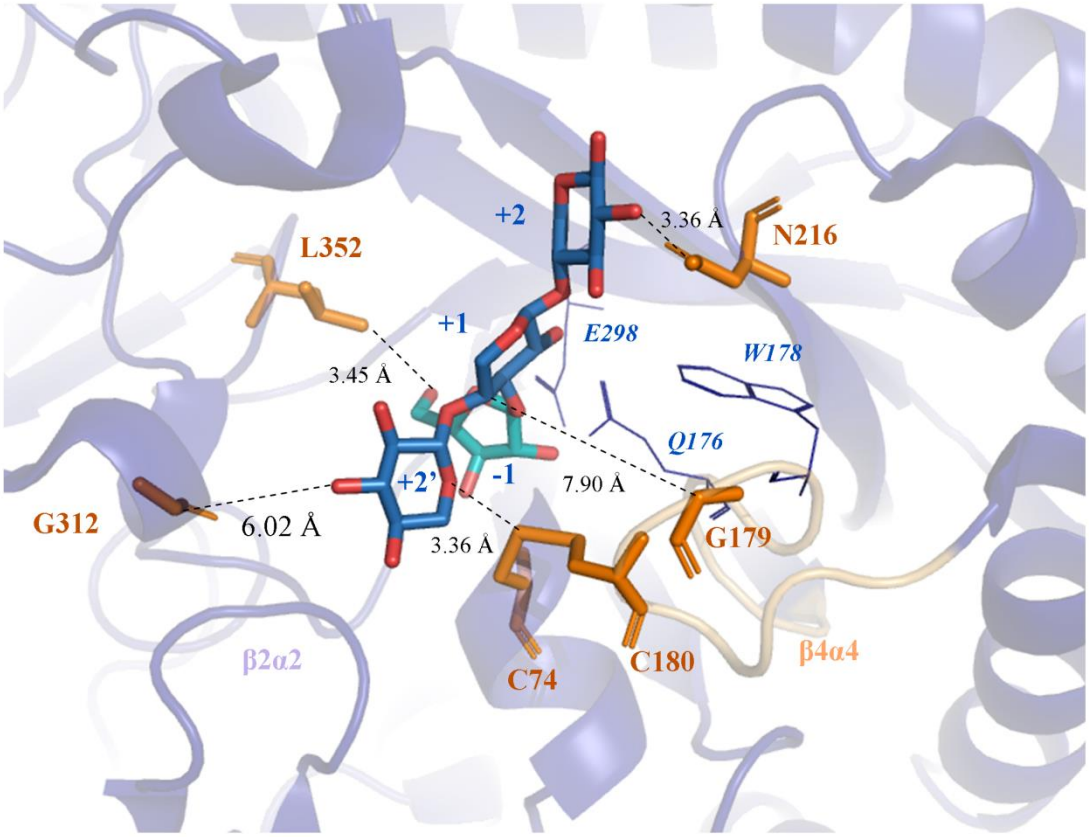


Fig. 2. Location of residues C74, G179, C180, N216, G312 and L352 (orange sticks) within *Tx*Abf active site co-crystallized with XA^{3}XX (PDB 2VRQ, (Paës et al., 2008). The nucleophile E298, mutated acid/base E176Q and +1 residue W178 are represented as lines (blue).

Regarding the positioning of xylotriose spanning from +1 to +3 subsites (i.e. branching on non-reducing end, Fig. SI 2B and C), the same G179 hot-spot was predicted (Fig. SI 4 and Fig. SI 5). This further implies that if position 179 is mutated, at best **only the efficiency of transglycosylation will be improved**, **but not the regioselectivity of the reaction**. Additionally, C74 and C180 were also detected as hot-spots during runs 4 and 3, respectively. The introduction of large side chains in the C74 position also fills the empty space left at +2' subsite for this new binding mode of xylotriose. At the same time, a hydrogen bond is formed between the non-reducing end xylotriose and the new E74 residue (the best representative of C74 hot-spot in terms of binding affinity). Likewise, the introduction of large side chains in the C180 position fills the empty space left at +2' subsite for "A3XX binding mode" of xylotriose. At the same time, a couple of hydrogen bonds are formed between the non-reducing end of xylotriose and the new N180 residue (the best representative of C180 hot-spot in terms of binding affinity).

Concerning residues identified as cold-spots, these are common between runs 3 and 4 (except W99) and it is worth mentioning that, unlike in run 1 and 2, C74 or C180 are no longer predicted as cold-spots since, in run 3 or 4, either of these two is found as a hot-spot, and these two cysteines are actually involved in a disulfide bridge.

Besides, among the 4 common cold-spots (E176, N216, Y242 and W248) it is noteworthy that N216 was detected as a stronger mutational point during run 4 (Fig. SI 5), for "A^2XX mode", with binding affinities increasing by approximately 2.3 kcal.mol^{-1}, as opposed to 0.4-0.6 kcal.mol^{-1} for other binding modes.

2.2. *In vitro* assessment of the impact of predicted mutations on H/T modulation

2.2.1. Hydrolytic behaviour

The three aromatic substitutions at G179 position (F, W and Y) as well as G312E were created and the purified recombinant enzymes were first assayed for their hydrolytic ability onto the donor substrate *p*NP-α-L-Ara*f* (Table 1). Besides, L352M, a -1 mutation identified in the course of a directed evolution process ((Arab-Jaziri et al., 2014)*, Supporting information A3*), considered as an interesting protein background for T/H improvement and recombination with these acceptor mutations, was characterized here in the same conditions (Fig. 2). Furthermore, N216W, a + 2 mutation previously identified through a semi-rational approach that leads to the inability to perform transfer with (1,3) regioselectivity (Part I). Therefore, it was thought relevant to further investigate this position on a biochemical point of view in the frame of this study.

Regarding the catalytic constant (k_{cat}) on *p*NP-α-L-Ara*f*, G179 substitutions lead to 54, 50 and 63% of *Tx*Abf k_{cat} (433 s^{-1}), for F, W and Y respectively. A slight substrate excess inhibition was also observed and K_M values are lower than *Tx*Abf one (divided by 2.3, 8.7 and 2.9-fold, for G179F, W and Y, respectively). Regarding the L352M derivatives, the introduction of G179F, W or Y mutations led to 49, 51 and 48% of L352M k_{cat} value (343 s^{-1}), which correspond to a similar impact than the one observed on *Tx*Abf. K_M values for these double mutants are also lower than for L352M (divided by 3.1, 6.7 and 4.8-fold, respectively), as observed for the same mutations introduced on the wild-type background, and which might be related to alteration of interactions with the *p*NP moiety located in +1 subsite. Regarding G312E mutation (+2' subsite), as a single mutation, it does not affect drastically the catalytic constant (75% relative to k_{cat} of *Tx*Abf) whereas when combined to L352M it leads to a relative k_{cat} of 11% compared to L352M. Finally, the +2 mutation N216W does not modify drastically the rate-limiting deglycosylation-associated constant k_{cat} but introduces a strong substrate excess inhibition, coupled to an apparent decrease in K_M value for *p*NP-α-L-Ara*f*. It is noteworthy that the incorporation of the L352M mutation (-1 subsite) removed the apparent substrate-excess inhibition.

Table 2. Steady-state kinetic parameters[a] for the hydrolysis of *p*NP-α-L-Ara*f* by *Tx*Abf and *BindScan* related mutants.

Enzyme	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}.mM^{-1}$)	Relative K_M (-fold)	Relative k_{cat}/K_M (%)	K_{is}[b] (mM)
*Tx*Abf	433 ± 10	0.27 ± 0.03	1671 ± 224	1	100	-
L352M	343 ± 18	16.2 ± 1.5	21.2 ± 3.0	60	1.3	-
G179F	232.6 ± 2.0	0.115 ± 0.006	2029.9 ± 123.7	0.42	121	288.9 ± 75.2
G179F-L352M	166.8 ± 5.9	5.15 ± 0.46	32.41 ± 4.03	19.07	1.9	-
G179W	214.2 ± 3.5	0.031 ± 0.004	6842.4 ± 985.4	0.12	409	56.4 ± 7.9
G179W-L352M	173.5 ± 3.0	2.43 ± 0.14	71.3 ± 5.3	9.01	4.3	-
G179Y	272.5 ± 4.9	0.094 ± 0.009	2899.4 ± 104.4	0.35	174	49.2 ± 6.6
G179Y-L352M	162.9 ± 1.7	3.38 ± 0.11	48.2 ± 2.0	12.52	2.9	-
G312E	323.0 ± 4.1	0.66 ± 0.03	492.4 ± 28.7	2.43	29	-
G312E-L352M	47.2 ± 0.9	5.48 ± 0.28	8.6 ± 0.6	20.29	0.5	-
N216W	310.3 ± 8.8	0.0578 ± 0.0080	5369.3 ± 895.9	0.21	321	12.89 ± 1.251
N216W-L352M	475.5 ± 5.0	1.50 ± 0.06	316.2 ± 16.6	5.56	19	-

[a] Reactions were carried out in triplicate in sodium acetate buffer (pH 5.8, 50 mM) at 45 °C.
[b] K_{is} values were obtained from fitting of experimental values to a model of excess-substrate inhibition.

2.2.2. Transglycosylation potency

What is the impact of donor concentration on transglycosylation yield?

Given the significantly altered saturation profile (large increase in K_M values) observed for L352M mutant, which also displays improved transfer abilities (Arab-Jaziri et al., 2014), it was considered that the impact of donor substrate concentration on transglycosylation efficiency should be analysed

before focusing on the characterization of *BindScan* mutants and derivatives. It was observed that increasing the donor substrate concentration (from 5 to 15 mM) affected modestly the transfer rate when keeping a constant xylotriose concentration (at 10 mM), coupled to a significant increase of autocondensation rate (ca. 3-fold) (Table SI 1). Importantly, with the increase in *p*NP-α-L-Ara*f* concentration, a greater accumulation of transglycosylation product was made possible (Fig. SI 6), for reactions catalysed by *Tx*Abf as well as by L352M mutant. This is probably due to a delayed secondary hydrolysis since the activated donor is present in more important quantities for a longer period of time. As a consequence, conditions using (15:30, mM/mM) ratio were considered as the best ones for transglycosylation monitoring while still using moderate quantities of xylotriose acceptor and staying under the solubility limit of *p*NP-α-L-Ara*f* (ca. 20 mM).

Impact of acceptor subsites mutations on transfer ability

As previously observed (Part I), N216W (+2 subsite, Fig. 2) leads to the creation of a (1→2) regiospecific mutant, deleting *Tx*Abf transglycosylation potency for *O*-3 branching, whatever binding mode is adopted (Fig. SI 7). Besides, A^2XX maximum yield is not significantly modified compared to *Tx*Abf (~ 10%), as its transfer rate (Table SI 2). Interestingly, when combined to L352M mutant, the regiospecificity phenotype of N216W and transfer rate of L352M are conserved. A synergistic effect is observed for N216W-L352M leading to a 2- and 4-fold increase in A^2XX yield (~ 40%) compared to L352M and *Tx*Abf (Fig. SI 7), respectively. This shows that L352M constitutes a rather promising protein background for recombination with acceptor subsites mutations.

According to *BindScan* predictions, the introduction of aromatic substitutions at G179 position (+1 subsite) should improve the affinity for xylotriose (ca. -6.8 $kcal.mol^{-1}$, net gain of ca. -0.5 $kcal.mol^{-1}$ for G179W for instance). Therefore, the transarabinofuranosylation capacity of G179F, W and Y mutants as well as that of L352M-recombined mutants was investigated. Since *BindScan* runs were performed individually for a given substrate orientation it is relevant to analyse the impact of mutations according to each product independently, i.e. XA^3X (run 1) A^3XX (run 3) and A^2XX (run 4). Albeit not included in *BindScan* predictions the analysis of autocondensation product was also carried out. Regarding the XA^2X product (run 2) the ^1H NMR signal resolution unfortunately precluded any analysis with respect to the impact of *BindScan* mutations on this particular acceptor orientation.

Concerning XA^3X, whichever mutant is considered, G179F, W or Y, no secondary hydrolysis is observed (as opposed to *Tx*Abf) (Fig. SI 8) whereas A^2XX, the main product, is more sensitive to this degradation once the donor substrate is totally consumed (Fig. SI 9). However, the apparent L-Ara*f* transfer rate (R_T) increase remains modest for XA^3X when taking into account the error range, the most significant enhancement being observed for G179F mutant (Table SI 2).

Regarding A^2XX, R_T values are increased to different extents: 1.8, 1.3 and 2-fold compared to *Tx*Abf, for G179F, W and Y respectively. Besides, maximum transglycosylation yields obtained for each reactions (N.B., corresponding donor conversion yield are not equivalent) reach 2.2, 1.9 and 2.4-fold increase for G179F, W and Y, respectively, compared to *Tx*Abf (10%) (Figs 3 and Fig. SI 9). When these mutations are introduced on L352M background transglycosylation yields are multiplied by 4.1, 4.0 and 2.1-fold, respectively, compared to *Tx*Abf, and by 2.3, 2.3 and 1.2-fold relatively to L352M. Therefore, the positive impact of G179 mutations is less marked when combined with L352M compared to the wild-type background.

About A^3XX, G179F, W or Y mutations reduce extremely L-Ara*f* transfer rates and consequently transglycosylation yields. Indeed, if looking at a common *p*NP-α-L-Ara*f* conversion rate (80%) A^3XX yields reach 15 and 10% for L352M and *Tx*Abf respectively, whereas these lie between 2 and 3% for G179W, F, Y and corresponding L352M mutants thereof.

For the autocondensation product (α-L-Ara*f*-α-L-Ara*f*-*p*NP, linkage not characterized), G179W provokes a 2-fold increase in transfer rate whereas F and Y substitutions display a *Tx*Abf-like behavior. This property is also conserved when combined to L352M.

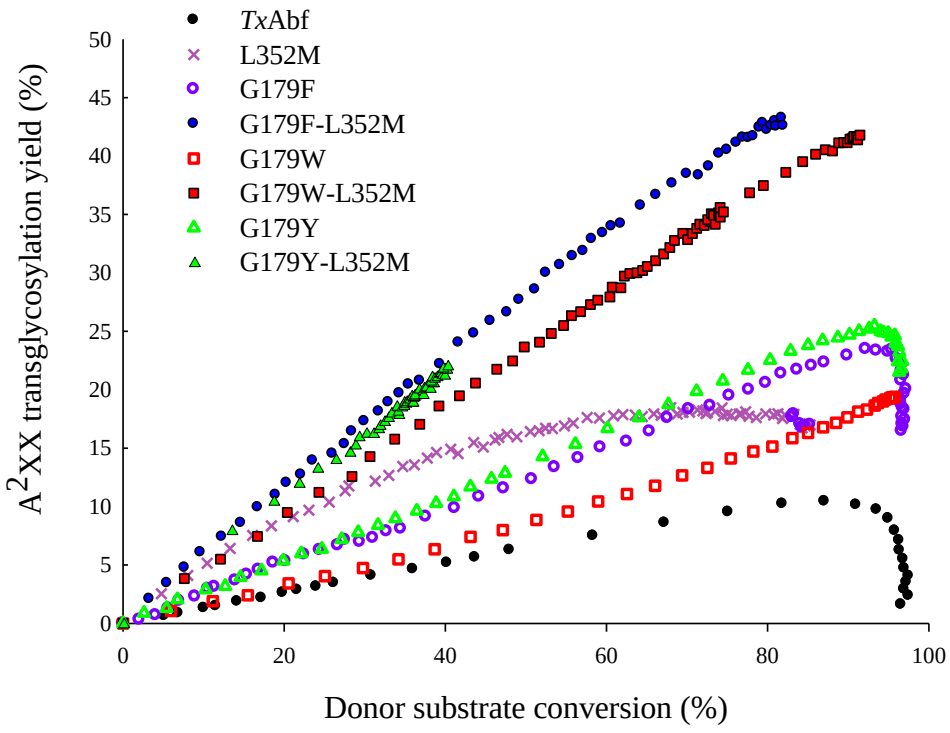


Fig. 3. Donor conversion-dependent plot of the evolution of tranglycosylation product A^2XX by *Tx*Abf and mutants thereof. Reactions were catalyzed by *Tx*Abf (18 nM), G179F (50 nM), G179W (50 nM), G179Y (50 nM), L352M (50 nM), G179F-L352M (50 nM), G179W-L352M (100 nM) and G179Y-L352M (150 nM) mutants. For the latter one a strong inhibition was observed, although more enzyme was employed, leading to maximum conversion yield of 40%. Each assay was carried out in deuterated acetate buffer (50 mM, pD 5.8) at 45 °C with 50 nM of enzyme (or 18 nM for *Tx*Abf and 150 nM for G179Y-L352M).

Besides, to know if a synergistic effect could be observed between +1 (G179W) and +2 (N216W) mutations an analysis of the triple mutant G179W-N216W-L352M was carried out. Although the regiospecificity phenotype brought by N216W was conserved this revealed that transglycosylation yields were not further improved and that it was even detrimental for R_T values (Table SI 2).

Finally, regarding the mutation G312E suggested by *BindScan* run 2 (XA^2X), no significant modification was observed on the production of A^2XX and XA^3X products (Fig. SI 10 and Fig. SI 11). A^3XX yields were even lower (by 2 to 3-fold). Regarding XA^2X, signal overlapping precluded a proper integration but superimposed spectra revealed nevertheless that no transglycosylation improvement was introduced trough G312E mutation, as a mono-mutant as well as recombined with L352M (data not shown).

2.2.3. Probing affinity modifications by isothermal titration calorimetry

To provide a finer understanding of behaviours issued from this *in silico*/biochemical hybrid study two questions have to be tackled. First, as suggested by biochemical values (K_M notably) on *p*NP-α-L-Ara*f,* do BindScan mutations really increase the affinity for ligand? Second, are the observations made on secondary hydrolysis postponing phenomenon the result of a greater difference of recognition between the donor substrate and the transglycosylation product? To answer both questions, and allow a direct comparison between *in silico* and *in vitro* experiments, isothermal titration calorimetry (ITC) analyses were performed. Regarding the protein targets, two archetypes were chosen, G179F as a representative of *BindScan* mutants and N216W for its regiospecifity phenotype. To inactivate these enzymes the acid/base catalytic residue was mutated (E176Q). Based on available crystallographic data the E176Q mutation occupy a similar steric space compared to the native E176 (Fig. SI 12) and should not introduce important parasite effects relative to ITC-derived constants. Whatever the extent of these is, all measures are compared to a third enzyme serving as reference, *Tx*Abf-E176Q. Regarding the ligands, the donor substrate *p*NP-α-L-Ara*f* and the main transglycosylation product A^2XX were employed. As a result, it was unfortunately observed that a residual hydrolytic activity was still detectable with the enzyme E176Q and *p*NP-α-L-Ara*f* as ligand, which is actually an activated substrate. Although several enzyme concentrations and (enzyme:ligand) ratio were attempted, this undesirable activity precluded the determination of reliable thermodynamic binding values. A similar residual activity was observed with E176Q-N216W but not with E176Q-G179F. Therefore, to avoid erroneous interpretations only a qualitative assessment was performed when possible (Fig. SI 13). Regarding the titration with *p*NP-α-L-Ara*f* of E176Q-G179F an exothermic signal was observed with an enthalpic contribution approximating -12. $kcal.mol^{-1}$. Concerning A^2XX tetrasaccharide ligand, representing the main transglycosylation product, unlike *p*NP-α-L-Ara*f* no residual hydrolytic activity was observed, and isothermal titration revealed

stronger interactions than those observed with *p*NP-α-L-Ara*f*, given the absence of signal for the latter. Significantly, endothermic reactions were observed with A^2XX and it is noteworthy that the interactions are mainly enthalpy-driven, irrespective of the mutant under investigation. Regarding the *BindScan* mutation impact, the titration of E176Q-G179F by A^2XX induced greater change in enthalpy energy (ΔH) relative to E176Q (approximately 2-fold), and K_d estimations tend to show better affinity for A^2XX when G179F is present at +1 subsite (Fig. SI 13). Furthermore, for this mutant the absolute value of ΔH is slightly higher with A^2XX than for *p*NP-α-L-Ara*f*. However K_d values comparison cannot be carried out. Finally, about E176Q-N216W, the ΔH values lies in the same range than that of E176Q.

3. Discussion

3.1. Do BindScan predictions match with already-known features of *Tx*Abf?

In the perspective of developing efficient predicting tools it is worthwhile to analyse the relevance of positions suggested by this novel computational algorithm in the light of the structure-function knowledge acquired on *Tx*Abf during the past decade. Regarding the **hot-spot G179** highlighted by *BindScan*, this position has never been submitted to site-directed mutagenesis or found through directed evolution process and thus represented as such an interesting new target. About **cold-spots**, interestingly most of them were already proved to be detrimental for transglycosylation potency, which is a positive point with respect to the reliability of the method. Indeed, W99 mutation leads to a transglycosylation yield decrease on *p*NP-β-D-Gal*f*/Bn-α-D-Xyl*p* substrate pair and was demonstrated to be of utmost importance for *Tx*Abf β2α2 loop dynamic and global catalysis (Arab-Jaziri et al., 2012). E176 and Y242 are two residues of the eight key amino acids defining clan A (Durand et al., 1997). The former is the acid/base catalytic residue necessary for the overall mechanism and the latter was shown to interact with the nucleophile E298 as well as with both the donor and acceptor substrates (Paës et al., 2008), and is essential for catalysis to occur properly (data not shown). Through site-saturation mutagenesis at W248 (+1 subsite) and W302 (-1/+1) positions we demonstrated that the product partition could be modified but that the global yield was severely reduced (Part I). Regarding C74 and C180, these are cysteines involved in a disulfide bond and C180A mutation provoked a 15-fold decrease in k_{cat} (Paës et al., 2008), which was suggested to be related to a modified orientation of the acid/base catalyst E176. Therefore, given the importance of these residues with regard to their involvement in *Tx*Abf structure integrity, the potential hot-spot nature of C74 or C180 may not result in the expected phenotype upon mutation. Finally, cold spots V244 and N216 are both constitutive of subsite +2, located on either side of the groove, suggesting that the topology of this subsite might be determinant for the efficient branching

of L-Ara*f* onto non-reducing end of xylotriose (run 3 and 4). For the former no structure-function data is available but for the latter it was shown that N216W substitution leads to a (1,2)-regiospecific transfer on the non-reducing end of xylotriose (i.e. A^2XX end product). This is in agreement with the identification of N216 as a cold-spot albeit rising further questioning regarding the regioselectivity prediction since predicted by the four runs, as discussed later down (cf. section 3.5).

Additionally, it was noticed that during the *in silico* docking calculations of xylotriose to prepare the starting structures for the *BindScan* runs, the binding mode of xylotriose acceptor showing the strongest binding affinity towards the *Tx*Abf-β-L-Ara*f* intermediate was that spanning from subsites +1 to +3 (Fig. SI 2 C-D). In other words, this binding mode indicates a preference towards the production of substituted xylotriose at its non-reducing end (approximately -7.7 and -6.6 kcal.mol^{-1} for A^3XX or A^2XX, as opposed to -6.3 and -6.4 kcal.mol^{-1} for XA^3X and XA^2X binding modes, respectively). Experimentally, A^2XX is usually obtained as the major transglycosylation product, whereas A^3XX and XA^3X are produced in similar yields. This apparent discrepancy underlines the complexity of establishing a straightforward relationship between acceptor affinity and transfer rates. Maybe, other parameters such as the hydroxyl reactivity intervene in efficiency of the nucleophilic attack of the acceptor onto the enzyme-bounded anomeric carbon.

3.2. Are donor recognition and "hydrophobicity" altered by acceptor subsite mutations?

Since transglycosylation and hydrolysis compete for the outcome of deglycosylation step of the two-step displacement mechanism characteristic of retaining GHs (Koshland Jr., 1953; Sinnott, 1990), it was judged essential to analyse first the impact of *BindScan* hot-spot mutants on *Tx*Abf hydrolytic behaviour (i.e. using only the donor substrate *p*NP-α-L-Ara*f*), to then better understand their role in H/T modulation. Regarding the global efficiency, the decrease in k_{cat}, reflecting the rate-limiting deglycosylation step (Part IV) carried out essentially by water in "hydrolytic mode", is consistent with the introduction of bulky aromatic residues (F, W or Y) at the active site entrance. Besides, stacking interactions are likely to be enhanced in +1 subsite between aromatic residues (in place of G179) and the *p*NP moiety of *p*NP-α-L-Ara*f*, as suggested by the decrease in K_M values and the slight excess-substrate inhibition. Furthermore, the increase in autocondensation reaction, principally for G179W, may also be related to an increase in affinity for a second donor substrate molecule acting as an acceptor during the deglycosylation step (spanning +1 to +2 or +1 to +2' subsites). Calorimetric experiments aiming at providing a more accurate view of this affinity unfortunalely failed to give any reliable K_d values.

Aiming at designing more efficient transarabinofuranosylases *BindScan* mutations were recombined with L352M, a -1 mutation previously identified in the course of random

mutagenesis/screening steps allowing to enhance *Tx*Abf transglycosylation rate (Arab-Jaziri et al., 2014) (SI A.3). This mutation provokes a dramatic increase in K_M value on *p*NP-α-L-Ara*f*, with low impact on water-mediated deglycosylation (modest decrease in k_{cat}), reflecting an influence on the L-Ara*f* moiety recognition (-1 subsite). Astoundingly, the phenotype induced by *BindScan* mutations G179F, W and Y was actually shown to be rather independent of the enzyme background, since similar relative impacts are observed when introduced in wild-type *Tx*Abf or L352M enzymes. This demonstrates that the impact of G179 mutations is confined within the +1 region with none or limited consequences on the intrinsic catalytic machinery. Similar conclusions can be withdrawn from the hydrolytic behaviour of the +2 mutant, N216W.

3.3. How to explain that secondary hydrolysis is delayed?

Consistently with the previous remark on protein background-independent feature of the G179-induced hydrolytic phenotype, a total conservation of both enzymatic behaviors brought by G179 and L352 mutations is observed in transglycosylation mode. This results in synergistic effects when both of them are combined. The insertion of G179F, W or Y mutations permits to **delay the apparent secondary hydrolysis to the very end of the reaction**, yielding a significantly higher quantity of transglycosylation product. To provide an explanation to such a phenomenon, comparative calorimetry experiments were performed as a complement to hydrolysis catalytic constants determination and transglycosylation behaviour monitoring. From the latter, the more likely hypothesis is that a greater difference of affinity between the donor *p*NP-α-L-Ara*f* and the xylotriose acceptor is engineered through G179 mutations in +1 subsite. This may then explain that during the course of the reaction the enzyme first targets the donor *p*NP-α-L-Ara*f* as a preferential substrate before hydrolysing the transglycosylation product (A^2XX). As a consequence, the *p*NP-α-L-Ara*f* conversion threshold, above which secondary hydrolysis is significantly triggered, is probably raised. In spite of attempts to assess this hypothesis via ITC assays, the complexity of these experiments did not allow to definitely validate it. Notwithstanding, these calorimetric experiments allowed to show that greater interactions (in terms of enthalpy) are developed with A^2XX when G179F is introduced in +1 subsite, which is consistent with its predicted role on an incoming xylotriose acceptor. Furthermore, it is worthwhile mentioning that with E176Q-G179F opposite enthalpic effects are observed, i.e. exothermic or endothermic with *p*NP-α-L-Ara*f* or A^2XX ligand, respectively. First, the enthalpic predominant contribution over entropy denotes that polar interactions (hydrogen bonds) are probably a main driver (Olsson et al., 2008; Olsson et al., 2011), which is understandable given the poly-hydroxylated nature of the tetrasaccharide A^2XX. Second the exothermic characteristic of the binding between E176Q-G179F and *p*NP-α-L-Ara*f* might appear as surprising since in a previous study we described as endothermic (albeit also enthalpy-driven) the behaviour of E176A with this

same substrate (Arab-Jaziri et al., 2012). Maybe this radical change comes from the interaction between the *p*NP moiety of *p*NP-α-L-Ara*f* and the phenyl ring of G179F, although we are currently not able to provide further explanations for this. Overall, although frustrating since no quantitative results could be obtained, these calorimetric experiments allowed nevertheless to confirm indirectly (i.e. on product A^2XX and not on the incoming acceptor xylotriose) the predicted role of G179F as a hot-spot.

3.4. Could G179 'aromatic mutants' phenotype be expanded to other GH51 Abfs?

In the perspective of withdrawing wider lessons from this case of study it is worthwhile to analyse mutagenesis/activity data with respect to protein evolution. Among the diversity of GH51 Abfs, G179 (*Tx*Abf numbering) is a totally conserved position (Fig. SI 14 C). However, structure and sequence-based analyses reveals that two different main configurations in the region surrounding G179 can be observed within the GH51 diversity. In *Tx*Abf-like cases, G179 is found within the motif [$W^{178}\underline{G}^{179}C^{180}G^{181}$] whereas in *Gs*Abf-like case it is the motif [$D^{177}\underline{G}^{178}P^{179}W^{180}$] where *Gs*Abf-$G^{178}$ corresponds to *Tx*Abf-G^{179}. Both motifs are approximately found at 50% each within GH51 Abfs, and the absence of *Tx*Abf-W^{178} equivalent is, for most of sequences, correlated with the presence of *Gs*Abf-W^{180} equivalent. In those cases, an aspartate (e.g. *Gs*Abf-D^{177}), or a tyrosine to much lesser extent, are found in place of *Tx*Abf-W^{178}equivalent. On a structural standpoint, *Tx*Abf is the only representative of its kind out of the six available GH51 Abfs crystallographic structures (Fig. SI 14A-B). Indeed, *Tx*Abf-W^{178} is perpendicular to acceptor subsites (between +1 and +2) whereas in *Gs*Abf-like cases W180 is parallel to the catalytic groove (subsite +1) (Fig. SI 14A). In the latters, the opposite side of the cleft is not constrained (short loop β6α6 loop) unlike *Tx*Abf displaying a longer β6α6 loop, harbouring W248 which is constitutive of +1 subsite (parallel to the groove), and represents a sort of symmetric of *Gs*Abf-W^{180} equivalent. In *Tx*Abf-like cases a non-bulky amino acid (G181) is found in place of *Gs*Abf-W^{180} equivalent. Accordingly, a phylogeny analysis revealed that GH51 Abfs can be subdivided into three different groups, showing that *Tx*Abf belongs to group C whereas the five other crystallized GH51 Abfs belong to clade B (Lagaert et al., 2014). In the light of these structural and sequence data it appears that G179F substitution might have similar impact if introduced in *Tx*Abf-like enzymes, whereas for *Gs*Abf-like ones such substitution appear much more difficult given the active site topology (i.e. presence of W^{180}, hiding G^{178} equivalent and precluding any functional aromatic substitution).

3.5. Strengths and weaknesses of *BindScan* for the prediction of regioselectivity-related hot-spots

BindScan mission was initially quite challenging with respect to the nature of the target ligand (i.e. acceptor substrate), which is close to the natural one. Although absolute binding affinities measured *in silico* for the acceptor spanning +1 to +3 subsites were in agreement with the "natural" transfer regioselectivity of *Tx*Abf (i.e. A^2XX as main product), none hot-spot that could introduce a regiospecific phenotype was identified via *BindScan*. Indeed, the same significant hot-spot (G179) was identified whether a *O*-2 or a *O*-3 orientation was chosen for xylotriose attack on the covalent intermediate *Tx*Abf-β-L-Ara*f*, irrespective of the binding mode (central or reducing end D-Xyl*p* unit). Of note, this was corroborated by *in vitro* experiments since *BindScan* mutations did not introduce any new regioselectivity.

Regarding the sensitivity threshold of *BindScan*, the example of G312E (+2' subsite) can shed light on the limits that should be taken into account to state whether a substitution would be relevant or not. G312E substitution (run 2) was meant to slightly increase xylotriose binding for XA^2X configuration (by only -0.2 kcal·mol^{-1}) but provided a similar behaviour than the wild-type. This could mean that either no real improvement in binding was introduced or that transglycosylation reactivity and acceptor affinity cannot be linked so straightforwardly. In the former case it teaches that a predicted gain of -0.2 kcal.mol^{-1} may not be considered as highly significant and lies below the sensitivity threshold of *BindScan* method for natural acceptor recognition. Also, the case of N216W mutant raises a few questions. Indeed, the fact that transfer rates are not modified combined to the observation that calorimetry experiments did not show drastic energetic changes with A^2XX would lead to the conclusion that the (1,2)-regiospecificity introduced by N216W mutation essentially comes from steric hindrances allowing only the "A^2XX mode". Besides, *BindScan* runs detected N216 as a cold-spot for each configuration, the A^2XX being even the most unfavourable in terms of binding energy (ΔΔG up to 2.4 kcal·mol^{-1} with respect to a maximum of 0.4 kcal·mol^{-1} for the other three orientations). It is noteworthy to recall that *BindScan* imposes the position (for a given regiospecificity) of the ligand according to preliminary docking experiment performed on the wild-type enzyme, before looking for mutations to improve its binding. The predicted cold-spot nature of the N216X substitutions, particularly destabilizing for A^2XX binding mode, suggest that the wild-type-associated xylotriose positioning might not be adapted for the observed *O*-2 regiospecific transfer in the case of the N216W mutant. Therefore, these results suggest that the xylotriose acceptor probably adopts a significantly different position within the active site of *Tx*Abf and N216W. Molecular dynamics could be considered to further study this particular regiospecific phenotype.

4. Conclusions

Finding the molecular determinants of H/T modulation in GHs and identifying/predicting the role of amino acid changes is a field of research that is currently being revolutionized through the employment of computational approaches as a complement to more traditional *in vitro* methods. Because retaining GHs proceed through a two-step mechanism increasing the affinity of the incoming acceptor, which serves as deglycosylating reagent during the second step, may appear as a relevant strategy. In the present study, the challenge consisted in predicting hot-spots for increasing the affinity between the covalent intermediate *Tx*Abf-β-L-Ara*f* and xylotriose, which is a natural acceptor. The *BindScan* analyses revealed a highly conserved position, G179, the aromatic substitutions of which introduce staking interactions and lead to doubled transfer rates as well as delayed secondary hydrolysis. The underlying reasons for this were proposed to be (1) an absolute increase of affinity of the D-xylose-containing ligand, and (2) a relative increase of the difference of affinity between donor substrate and transglycosylation product. The former is in agreement with *BindScan* predictions and consistent with transfer rates increases while the latter provides an explanation to the postponed secondary hydrolysis phenomenon. Such an approach constitutes a first of its kind in the corpus of approaches available in the literature that were developed to modulate the H/T balance in favour of transglycosylation catalyzed by GHs. Besides, among the putative strategies for T/H ratio increase, identifying mutations altering the TS energy level can be considered as one of the key approaches (Pierdominici-Sottile et al., 2011; Pierdominici-Sottile et al., 2013). In our case, this condition was fulfilled by introducing a -1 mutation (L352M), previously found in the course of random libraries screening. Satisfyingly, synergistic effects were then obtained after recombination with the rationally predicted (G179F, W or Y) or semi-rationally found (N216W) substitutions located in acceptor subsites, offering transglycosylation yields > 40%. Overall, this study proposes a global strategy that could be developed for H/T modulation and provides a rather convincing illustration of new computational tools, such as *BindScan*, that could be employed for other GHs in this perspective.

5. Experimental section

5.1. Substrates and chemicals

Unless otherwise stated, routine experimental work was performed using chemicals purchased mainly from Sigma-Aldrich (Illkirch, France) and molecular biology reagents purchased from New England Biolabs (Evry, France). The substrate, *p*NP-α-L-Ara*f*, was purchased from CarboSynth (Compton, U.K.). Xylotriose and XOS, used as acceptors, were purchased from Wako Chemicals GmbH (Neuss, Germany). Xylobiose was prepared in-house as previously described (Vrsanská et al.,

2008), using XOS as the starting material. ESI-HRMS: *m/z* calcd for $C_{10}H_{18}O_9Na$: $[M+Na]^+$ 305.0843; found: 305.0845 (1 ppm)(Hoffman et al., 1991).

A^2XX tetrasaccharide was produced in house as follows. In a 100 mL round bottom flask, the donor substrate *p*NP-α-L-Ara*f* (123 mg, 0.45 mmol, eq. 15 mM) and acceptor xylotriose (373 mg, 0.9 mmol, eq. 30 mM) were dissolved in sodium acetate buffer (25 mM, pH 5.8, 30 mL final volume). After incubation at 45 °C, the reaction was initiated by the addition of R69H-N216W-L352M mutant (1 µM). After 14 h of reaction (45 °C, under magnetic stirring) the reaction was stopped by inactivating the enzyme with a cycle of liquid nitrogen freezing/95 °C, 10 min/liquid nitrogen freezing. The reaction mixture was then freeze-dried. Acetylation of the lyophilized residue (with an acetic anhydride/pyridine mixture (1:1 v/v, 50 mL) in the presence of catalytic amount of dimethylaminopyridine (DMAP)) was carried out at room temperature and quenched after 16 h by adding MeOH (35 mL) at 0 °C. After concentration *in vacuo*, the residue was dissolved in CH_2Cl_2 and washed with aqueous saturated solution of $KHSO_4$, $NaHCO_3$ (twice) and brine. The aqueous solutions were back extracted with CH_2Cl_2 (twice). The organic layers were dried (Na_2SO_4), concentrated and purified on a silica column using the Reveleris flash chromatography System (Grace, Epernon, France), with a gradient of ethyl acetate in petroleum ether, with elution occurring at 60%. The peracetylated tetrasaccharide per-*O*-Ac-[A^2XX] (147 mg, 0.153 mmol) was isolated in 34% yields. For de-*O*-acetylation, the compound per-*O*-Ac-[A^2XX] (114 mg, 0.118 mmol) was treated in anhydrous MeOH/CH_2Cl_2 (1:1, v/v, 5 mL) with sodium methoxide (0.5 eq., 1 M in methanol), under nitrogen atmosphere, 1 h at 0 °C and 4 h at room temperature. The solution was neutralized with Amberlite IR-120 (H^+), filtered and then concentrated under reduced pressure. The residue dissolved in deionized water was freeze-dried to give A^2XX in quantitative yields (ca. 95 %).

2,3,5-Tri-*O*-acetyl-α-L-arabinofuranosyl-(1,2)-3,4-di-*O*-acetyl-β-D-xylopyranosyl-(1,4)-1,2,3-tri-*O*-acetyl-D-xylopyranose (Per-*O*-Acetyl α-L-Ara*f*-(1,2)-β-D-Xyl*p*-(1,4)- β-D-Xyl*p*-(1,4)-D-Xyl*p*)

^{1}H NMR ($CDCl_3$, 500 MHz, 298K): δ = 6.21 (d, $J_{1,2}$ = 3.6 Hz, H-1 of $Xyl^{I\alpha}$), 5.67 (d, $J_{1,2}$ = 7.2 Hz, H-1 of $Xyl^{I\beta}$), 5.38 (dd, $J_{3,4}$ = 9.5 Hz, H-3 of $Xyl^{I\alpha}$), 5.18 (t, $J_{3,4}$ = 8.5 Hz, H-3 of $Xyl^{I\beta}$), 5.14 (m, H-3 of $Xyl^{III(I\beta)}$), 5.12, 5.08 (br. s, H-1 of Ara*f* $^{(I\alpha \text{ or } I\beta)}$), 5.10 (m, H-3 of $Xyl^{III(I\alpha)}$), 5.07 (m, H-3 of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 4.97 (m, H-2 of $Xyl^{I\alpha}$), 4.96 (m, H-2 of $Xyl^{I\beta}$), 4.94 (m, H-3 of Ara*f*), 4.91 (m, H-2 of Ara*f*), 4.86 (m, H-4 of $Xyl^{III(I\beta)}$), 4.85 (m, H-4 of $Xyl^{III(I\alpha)}$), 4.77 (m, H-2 of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 4.55-4.51 (m, H-1 of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 4.46 (d, $J_{1,2}$ = 6.3 Hz, H-1 of $Xyl^{III(I\beta)}$), 4.41 (dd, H-5a of Ara*f*), 4.37 (m, H-1 of $Xyl^{III(I\alpha)}$), 4.36 (m, H-4 of Ara*f*), 4.20 (dd, H-5b of Ara*f*), 4.09 (m, H-5a of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 4.02 (m, H-5a of $Xyl^{I\beta}$), 4.02 (m, H-5a of $Xyl^{III(I\beta)}$), 3.96 (dd, $J_{5a,5b}$ = 5.5 Hz, H-5a of $Xyl^{I\alpha}$), 3.89 (m, H-5a of $Xyl^{III(I\alpha)}$), 3.86 (m, H-4 of $Xyl^{I\beta}$), 3.84 (m, H-4 of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 3.84 (m, H-4 of $Xyl^{I\alpha}$), 3.66 (m, H-5b of $Xyl^{I\alpha}$), 3.57 (m, H-2 of $Xyl^{III(I\beta)}$), 3.48 (m, H-2 of $Xyl^{III(I\alpha)}$), 3.47 (m, H-5b of $Xyl^{I\beta}$), 3.40 (m, H-5b of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 3.34 (dd, $J_{5a,5b}$ = 8.2 Hz, H-5b of $Xyl^{III(I\beta)}$), 3.26 (m, H-5b

of $Xyl^{III(I\alpha)}$).

^{13}C NMR ($CDCl_3$, 125 MHz, 298K): 106.8, 106.5 (C-1 of Ara*f*$^{(I\alpha \text{ or } I\beta)}$), 102.3 (C-1 of $Xyl^{III(I\alpha)}$), 101.3 (C-1 of $Xyl^{III(I\beta)}$), 100.7, 99.3 (C-1 of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 92.2 (C-1 of $Xyl^{I\beta}$), 89.3 (C-1 of $Xyl^{I\alpha}$), 81.2 (C-2 of Ara*f*), 80.8 (C-4 of Ara*f*), 77.1 (C-2 of $Xyl^{III(I\alpha)}$), 76.9 (C-3 of Ara*f*), 75.9 (C-2 of $Xyl^{III(I\beta)}$), 75.4 (C-4 of $Xyl^{I\alpha}$), 75.1 (C-4 of $Xyl^{I\beta}$), 75.1, 74.2 (C-4 of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 71.9 (C-3 of $Xyl^{I\beta}$), 71.8 (C-3 of $Xyl^{III(I\beta)}$), 71.8 (C-3 of $Xyl^{III(I\alpha)}$), 70.9, 70.2 (C-2 of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 70.2(C-3 of $Xyl^{II(I\alpha \text{ and } I\beta)}$), 70.0 (C-3 of $Xyl^{I\alpha}$), 69.7 (C-2 of $Xyl^{I\alpha}$), 69.5 (C-2 of $Xyl^{I\beta}$), 69.3 (C-4 of $Xyl^{III(I\alpha)}$), 68.2 (C-4 of $Xyl^{III(I\beta)}$), 63.3 (C-5 of Ara*f*), 63.2 (C-5 of $Xyl^{I\beta}$), 62.2 (C-5 of $Xyl^{III(I\beta)}$), 61.4 (C-5 of $Xyl^{I\alpha}$), 61.3 (C-5 of $Xyl^{II(I\alpha \text{ and } I\beta)}$).

5.2. *In silico* analyses

The details of the BindScan method are to be published elsewhere. In short, during a BindScan simualtion, each position is virtually mutated, one by one, into each of the 20 natural amino acids. For each protein variant, the binding affinity for the substrate is evaluated by means of a computationally derived scoring function.

5.3. Mutagenesis, protein expression, and purification

5.3.1. Site-directed mutagenesis

The plasmid construction pET24a-*Tx*Abf (original pET vectors from Novagen, Fontenay-sous-Bois, France) was used as template DNA for *in vitro* mutagenesis using the QuikChange II Site-Directed Mutagenesis kit (Agilent, Courtaboeuf, France), following the manufacturer's instructions. For the introduction of mutations G179F, W or Y, G312E and L352M the following primers (5'-3') were employed (underlined codon and mutated base in bold):

G179F	GGCAACGAGAACTGG**TT**CTGCGGCGGCAACAT
G179W	GCAACGAGAACTGG**TGG**TGCGGCGGCAACAT
G179Y	CGGCAACGAGAACTGG**TAT**TGCGGCGGCAACATGC
G312E	CGGGCACGAATCCGG**AG**TTCCTGTATCAGCAG
L352M	CAGCTCGTCAACGTG**A**TGCAATCCGTCATCC

For ITC experiments, *Tx*Abf and N216W enzymes were inactivated by mutation of the acid/base catalytic residue (E176Q) by using the following primers (5'-3'):

E176Q$^{wt\ template}$	GGCGTCGGCAAC**C**AGAACTGGGGCT

and the following one for G179F mutant encoding DNA matrix:

E176Q$^{G179F\ template}$	GGGCGTCGGCAAC**C**AGAACTGGTTCTG

The successful introduction of mutations was confirmed by DNA sequencing (GATC Biotech, Mulhouse, France).

5.3.2. Protein Expression and purification

Wild-type and mutated recombinant *Tx*Abf protein expression (1 L) and purification (IMAC) were performed as previously described (Bissaro et al., 2014).

5.4. Kinetic Studies

Michaelis-Menten kinetic parameters K_M, k_{cat} and k_{cat}/K_M of *Tx*Abf and mutants thereof were determined using a discontinuous enzyme assay, based on *p*NP release, as previously described (Bissaro et al., 2014). Reactions were performed in triplicate at 45 °C in sodium acetate buffer (50 mM, pH 5.8) using *p*NP-α-L-Ara*f* (0.25-18 mM) and 0.1 $mg.mL^{-1}$ BSA. Negative controls containing all of the reactants except the enzyme were used to correct for spontaneous hydrolysis of the substrate. One unit (IU) of enzyme specific activity corresponds to the amount of enzyme releasing one µmol of *p*NP per minute. Appropriate enzyme quantities (in the nM range) were used in combination with suitable substrate concentrations, such that less than 10% of the total substrate was hydrolyzed over the course of the measurement.

5.5. Monitoring transglycosylation

To monitor reactions by collecting ^{1}H NMR spectra on a Bruker Avance 500 spectrometer, enzyme solutions were first prepared by 10-fold dilution in D_2O (99.90%) and then concentration on an Amicon® Ultra filter (regenerated cellulose 10 kDa, Millipore), performing the operation twice. Afterwards, reactions were performed at 45 °C in NMR tubes containing 600 µL of buffered D_2O, containing *p*NP-α-L-Ara*f* as donor and xylotriose as acceptor, at a ratio of 1:2 (15 and 30 mM, respectively). To obtain a pD value of 5.8, deuterated acetate (Euriso-Top, France) buffer was employed, while pD 7.1 was achieved using a deuterated sodium phosphate buffer, which was prepared by dissolving sodium phosphate in D_2O, followed by lyophilisation. This two-step protocol was repeated three times to obtain sufficient deuteration. Values of pD were measured by determining pH using a glass pH electrode and then applying the equation $pD = pH_{electrode} + 0.41$ (Glasoe & Long, 1960). To initiate reactions, an aliquot of enzyme solution (never more than 5% of total reaction volume) was added to the reaction, the quantity of enzyme (18 nM for *Tx*Abf and N216W, 50 nM for G179F, Y or W, G312E, L352M and combined mutants, unless otherwise stated) being adjusted to suit an 8-12 h reaction time frame. ^{1}H NMR spectra were acquired semi-continuously, accumulating a series of 5.52 min scans (i.e., 32 scans) with 6 s delays between scans. Each ^{1}H NMR spectrum was acquired using an excitation flip angle of 30° at a radiofrequency field of 29.7 kHz, and the reference residual water signal, calibrated at δ = 4.55 ppm at 45 °C (Gottlieb et al.,

1997), was pre-saturated during the repetition delay (with a radiofrequency field of 21 Hz). The following acquisition parameters were used: relaxation delay (6 s), dummy scans (4). The actual duration of the reaction was dependent on the enzyme used and before enzyme addition, an ^{1}H NMR spectrum of the reaction mixture was acquired, which served as the zero point of the reaction, from which integrals were then corrected according to the small dilution factor induced by the enzyme addition.

Data processing was performed as previously described (Bissaro et al., 2014). Briefly, the time-dependent evolution of donor (*p*NP-α-L-Ara*f*) and acceptor (xylotriose) concentrations were quantified by integrating the relative anomeric proton signals. Molar balances, based on acceptor and donor consumption, were used to convert the transglycosylation product signal integral into concentration, and to evaluate concentrations of hydrolysis products, respectively. From this, the donor substrate conversion rate (X in %) was calculated as the mean of three distinct signals; those of the *ortho* and *meta* aromatic protons (*p*NP) and that of the anomeric proton (L-Ara*f*), with conversion accounting for both reaction outcomes (i.e., hydrolysis and transglycosylation), and identifying self-condensation. For all time-course NMR kinetics, the absolute error mean value on X ranged from 1 to 3%. The transglycosylation yield, $Y_{(product)}$, was determined by relative integration of product anomeric proton signals, these being at 5.39, 5.32 and 5.27 ppm for XA^3X, A^3XX and A^2XX respectively (Gruppen et al., 1992; Vietor et al., 1994; Ferré et al., 2000; Pastell et al., 2008; Pitkänen et al., 2011; Correia et al., 2011). Finally, $Y_{(product)}$ (%) was plotted against X (%) to provide the transfer rate (R_T) of the donor substrate to a given product (µmol of product/µmol of consumed substrate), a value which is independent of the reaction duration and thus suitable to provide a standardized comparison of the behavior of different enzymes. Finally, initial rates were derived from the linear regions in which the variation of X was from 0-25% up to 0-40%, with root mean square deviations between 0.96 and 0.99.

5.6. Isothermal titration calorimetry

ITC experiments were performed at 10°Cwith an ITC200 isothermal titration calorimeter (Microcal). Analysis of the data was performed according to the one-binding-site model using the MicroCal Origin software provided by the manufacturer. To ensure minimal buffer mismatch, samples of protein and ligand were prepared in the same relevant buffer (sodium phosphate buffer 50 mM, pH 7). Experiments were conducted with protein concentration of 111- 579, 139-554 and 146-437 µM in the microcalorimeter cell for E176Q, E176Q-G179F and E176Q-N216W, respectively. A total of 20 injections of 2 µl of ligand at 2.5 mM (*p*NP-α-L-Ara*f* or A^2XX) were made at intervals of 180 s while stirring at 1000 r.p.m Binding-related ITC data were corrected considering the heat of dilution for the ligand. As usual the heat of dilution of the macromolecule measured is negligible.

The data were integrated to generate curves in which the areas under the injection peaks were plotted against the ratio of injected sample to cell content.

Competing interests

The authors declare they have no competing interests.

Acknowledgements

PhD fellowship of B. Bissaro was supported by the Institut National de la Recherche Agronomique (INRA) [CJS]. MetaToul (Metabolomics & Fluxomics Facitilies, Toulouse, France, www.metatoul.fr) and its staff members are gratefully acknowledged for technical support and access to NMR facility. MetaToul is part of the national infrastructure MetaboHUB (The French National infrastructure for metabolomics and fluxomics, www.metabohub.fr). MetaToul is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, the SICOVAL, the Infrastructures en Biologie Sante et Agronomie (IBiSa, France), the Centre National de la Recherche Scientifique (CNRS) and the Institut National de la Recherche Agronomique (INRA). NMR experiments were also performed on the PICT – Genotoul platform of Toulouse and funded by CNRS, Université de Toulouse-UPS, Ibisa, European structural funds and the Midi-Pyrénées region.

References

Arab-Jaziri, F., Bissaro, B., Barbe, S., Saurel, O., Débat, H., Dumon, C., Gervais, V., Milon, A., André, I., Fauré, R. & O'Donohue, M.J. (2012) Functional roles of H98 and W99 and β2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *FEBS J.*, 279 (19), pp.3598–3611.

Arab-Jaziri, F., Bissaro, B., Tellier, C., Dion, M., Fauré, R. & O'Donohue, M.J. (2014) Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α-L-arabinofuranosidase using a directed evolution approach. *Submitted Results*.

Bissaro, B., Saurel, O., Arab-Jaziri, F., Saulnier, L., Milon, A., Tenkanen, M., Monsan, P., O'Donohue, M.J. & Fauré, R. (2014) Mutation of a pH-modulating residue in a GH51 α-L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products. *Biochim. Biophys. Acta Gen. Subj.*, 1840 (1), pp.626–636.

Correia, M.A.S., Mazumder, K., Brás, J.L.A., Firbank, S.J., Zhu, Y., Lewis, R.J., York, W.S., Fontes, C.M.G.A. & Gilbert, H.J. (2011) Structure and function of an arabinoxylan-specific xylanase. *J. Biol. Chem.*, 286 (25), pp.22510–22520.

Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B. & Mornon, J.P. (1997) Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology*, 7 (2), pp.277–284.

Fauré, R., Courtin, C.M., Delcour, J.A., Dumon, C., Faulds, C.B., Fincher, G.B., Fort, S., Fry, S.C., Halila, S., Kabel, M.A., Pouvreau, L., Quemener, B., Rivet, A., Saulnier, L., Schols, H.A., Driguez, H. & O'Donohue, M.J. (2009) A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls. *Aust. J. Chem.*, 62 (6), pp.533–537.

Ferré, H., Broberg, A., Duus, J.O. & Thomsen, K.K. (2000) A novel type of arabinoxylan arabinofuranohydrolase isolated from germinated barley analysis of substrate preference and specificity by nano-probe NMR. *Eur. J. Biochem.*, 267 (22), pp.6633–6641.

Glasoe, P.K. & Long, F.A. (1960) Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.*, 64, pp.188–190.

Gottlieb, H.E., Kotlyar, V. & Nudelman, A. (1997) NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem.*, 3263 (3), pp.7512–7515.

Gruppen, H., Hoffmann, R.A., Kormelink, F.J., Voragen, A.G., Kamerling, J.P. & Vliegenthart, J.F. (1992) Characterisation by ^{1}H NMR spectroscopy of enzymically derived oligosaccharides from alkali-extractable wheat-flour arabinoxylan.

Carbohydr. Res., 233, pp.45–64.

Hoffman, R.A., Leeflang, B.R., de Barse, M.M.J., Kamerling, J.P. & Vliegenthart, J.F.G. (1991) Characterisation by ^{1}H-NMR spectroscopy of oligosaccharides, derived from arabinoxylans of white endosperm of wheat, that contain the elements →4)[α-L-Ara*f*-(1→3)]-β-D-Xyl*p*-(1→or→4)[α-L-Ara*f*-(1→2)][α-L-Ara*f*-(1→3)]-β-D-Xyl*p*-(1→. *Carbohydr. Res.*, 221, pp.63–81.

Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. & Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 α-L-arabinofuranosidase. *EMBO J.*, 22 (19), pp.4922–4932.

Im, D.-H., Kimura, K., Hayasaka, F., Tanaka, T., Noguchi, M., Kobayashi, A., Shoda, S., Miyazaki, K., Wakagi, T. & Fushinobu, S. (2012) Crystal structures of glycoside hydrolase family 51 α-L-arabinofuranosidase from *Thermotoga maritima*. *Biosci., Biotechnol. Biochem.*, 76 (2), pp.423–428.

Koshland Jr., D.E. (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Cambridge Philos. Soc.*, 28 (4), pp.416–436.

Lagaert, S., Pollet, A., Courtin, C.M. & Volckaert, G. (2014) β-xylosidases and α-L-arabinofuranosidases: accessory enzymes for arabinoxylan degradation. *Biotechnol. Adv.*, 32 (2), pp.316–332.

Lagaert, S., Schoepe, J., Delcour, J.A., Lavigne, R., Strelkov, S.V., Courtin, C.M., Mikkelsen, N.E., Sandgren, M. & Volckaert, G. (2011) Elucidation of the substrate specificity and protein structure of Abfb, a family 51 alpha-L-arabinofuranosidase from *Bifidobacterium longum*. *To be Published*, PDB 2Y2W.

Olsson, T.S.G., Ladbury, J.E., Pitt, W.R. & Williams, M. a (2011) Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci.*, 20 (9), pp.1607–1618.

Olsson, T.S.G., Williams, M. a, Pitt, W.R. & Ladbury, J.E. (2008) The thermodynamics of protein-ligand interaction and solvation: insights for ligand design. *J. Mol. Biol.*, 384 (4), pp.1002–1017.

Paës, G., Skov, L.K., O'Donohue, M.J., Rémond, C., Kastrup, J.S., Gajhede, M. & Mirza, O. (2008) The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry*, 47 (28), pp.7441–7451.

Pastell, H., Tuomainen, P., Virkki, L. & Tenkanen, M. (2008) Step-wise enzymatic preparation and structural characterization of singly and doubly substituted arabinoxylo-oligosaccharides with non-reducing end terminal branches. *Carbohydr. Res.*, 343 (18), pp.3049–3057.

Pierdominici-Sottile, G., Horenstein, N.A. & Roitberg, A.E. (2011) A free energy study of the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. From the Michaelis complex to the covalent intermediate. *Biochemistry*, 50 (46), pp.10150–10158.

Pierdominici-Sottile, G., Palma, J. & Roitberg, A.E. (2013) Free energy computations identify the mutations required to confer trans-sialidase activity into *Trypanosoma rangeli* sialidase. *Proteins*, 82 (3), pp.424–435.

Pitkänen, L., Tuomainen, P., Virkki, L. & Tenkanen, M. (2011) Molecular characterization and solution properties of enzymatically tailored arabinoxylans. *Int. J. Biol. Macromol.*, 49 (5), pp.963–969.

Rémond, C., Plantier-Royon, R., Aubry, N. & O'Donohue, M.J. (2005) An original chemoenzymatic route for the synthesis of β-D-galactofuranosides using an α-L-arabinofuranosidase. *Carbohydr. Res.*, 340 (4), pp.637–644.

Sinnott, M.L. (1990) Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.*, 90 (7), pp.1171–1202.

Souza, T., Santos, C.R., Souza, A.R., Oldiges, D.P., Ruller, R., Prade, R. a., Squina, F.M. & Murakami, M.T. (2011) Structure of a novel thermostable GH51 α-L-arabinofuranosidase from *Thermotoga petrophila* RKU-1. *Protein Sci.*, 20 (9), pp.1632–1637.

Taylor, E.J., Smith, N., Turkenburg, J., D'Souza, S., Gilbert, H., Davies, G. (2006) Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from Clostridium thermocellum. *Biochem. J.*, 395, pp.31–37.

Trott, O. & Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, 31 (2), pp.455–461.

Vietor, R.J., Hoffmann, R.A., Angelino, S.A.G.., Voragen, A.G.J., Kamerling, J.P. & Vliegenthart, J.F.G. (1994) Structures of small oligomers liberated from barley arabinoxylans by endoxylanase from *Aspergillus awamori*. *Carbohydr. Res.*, 6215 (93), pp.245–255.

Vrsanská, M., Nerinckx, W., Claeyssens, M. & Biely, P. (2008) An alternative approach for the synthesis of fluorogenic substrates of endo-β-(1→4)-xylanases and some applications. *Carbohydr. Res.*, 343 (3), pp.541–548.

Supporting Information of Part III: "Predicting Mutational "Hot-spots" for Hydrolysis/Transglycosylation Modulation in *Tx*Abf using *BindScan* Algorithm".

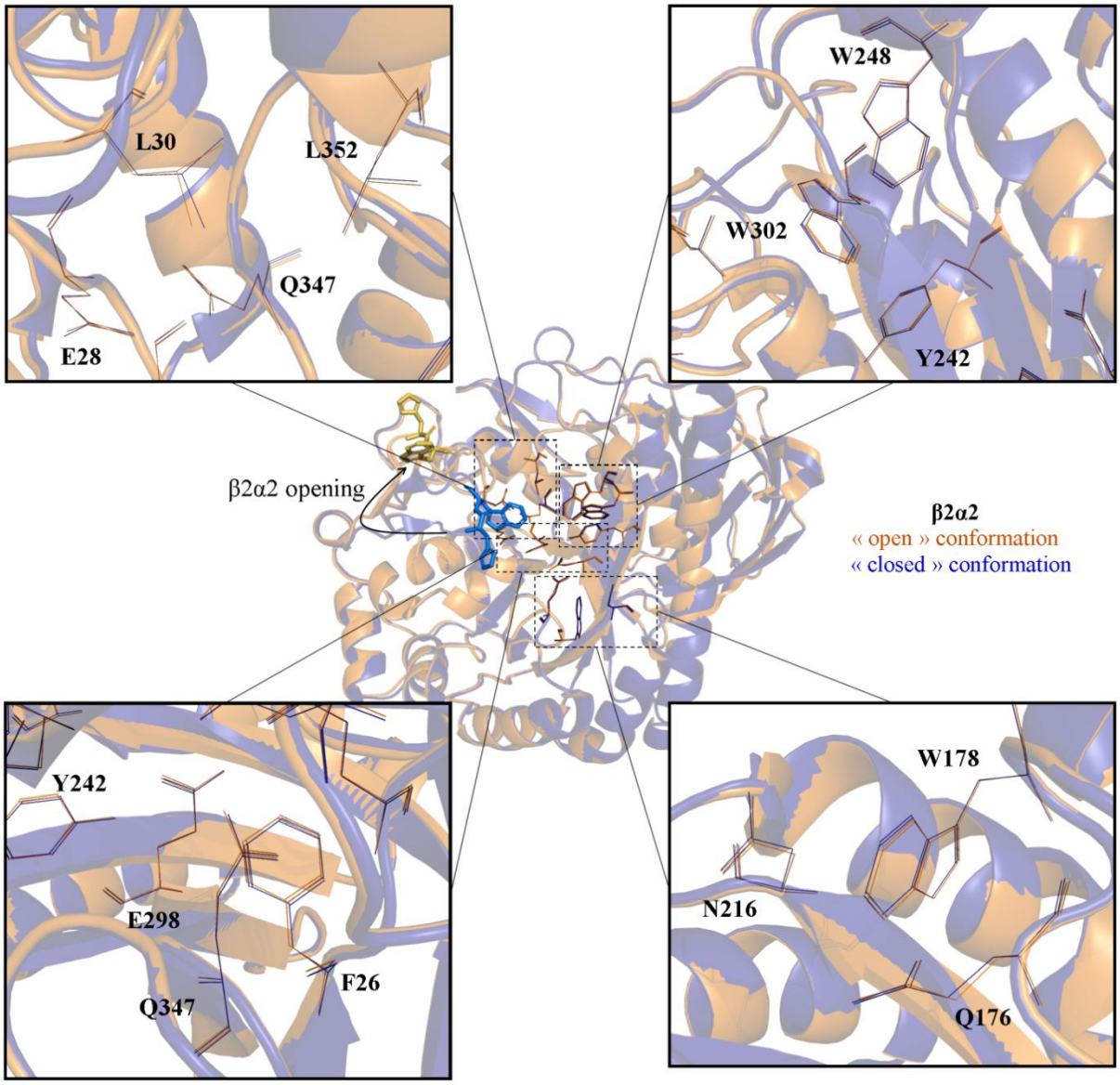


Fig. SI 1. Comparison of side chains orientation of key residues within *Tx*Abf active site between structures with opened (orange, PDB 2VRK) or closed (blue, PDB 2VRQ) conformation of β2α2 mobile loop. A catalytically operational active site is formed upon closure of β2α2 loop (Paës et al., 2008; Arab-Jaziri et al., 2012), at one end of which are located H98 and W99 (represented as sticks) constituting subsites +2' and -1, respectively. Key residues constitutive of -1 (F26, E28, L30, Y242, L314, Q347 and L352), +1 (W178, W248 and W302) or +2 (N216) subsites as well catalytic residues (E176Q and E298 for mutated acid/base and nucleophile respectively) were compared.

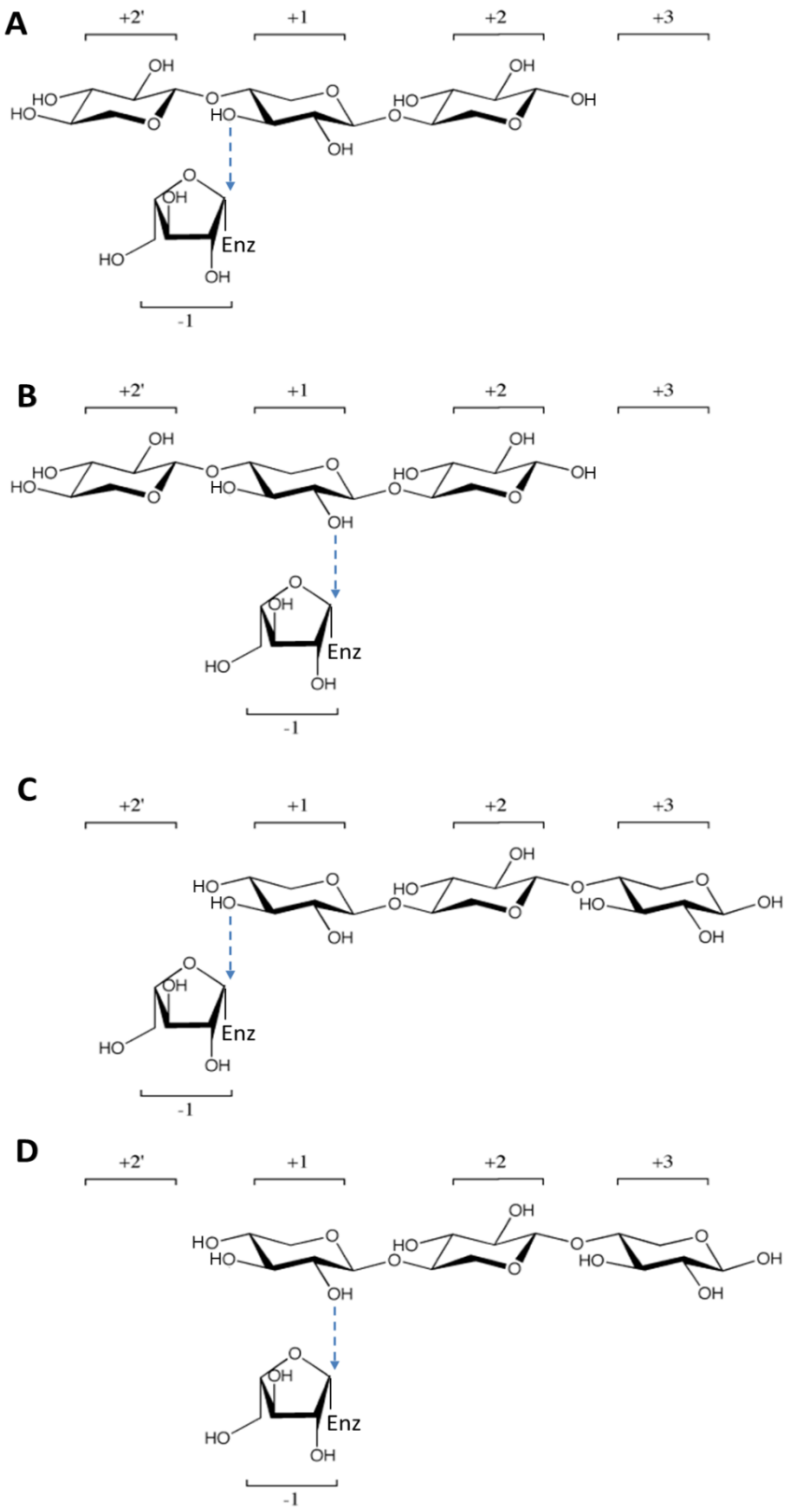


Fig. SI 2. Xylotriose orientation modes to favour (A) XA^{3}X, (B) XA^{2}X, (C) A^{3}XX or (D) A^{2}XX transglycosylation products formation.

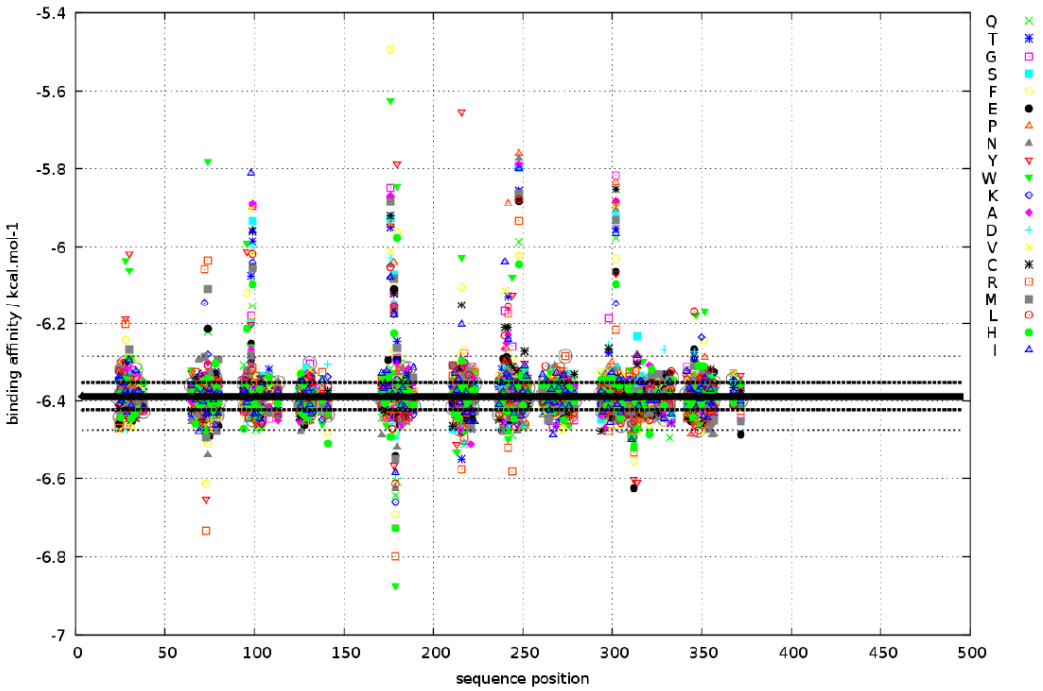


Fig. SI 3. *BindScan* run 2. (Binding energy-mutation) profile of the covalent intermediate *Tx*Abf-β-L-Ara*f* in complex with xylotriose substrate in the central *O*-2 orientation for $XA^{2}X$ product formation.

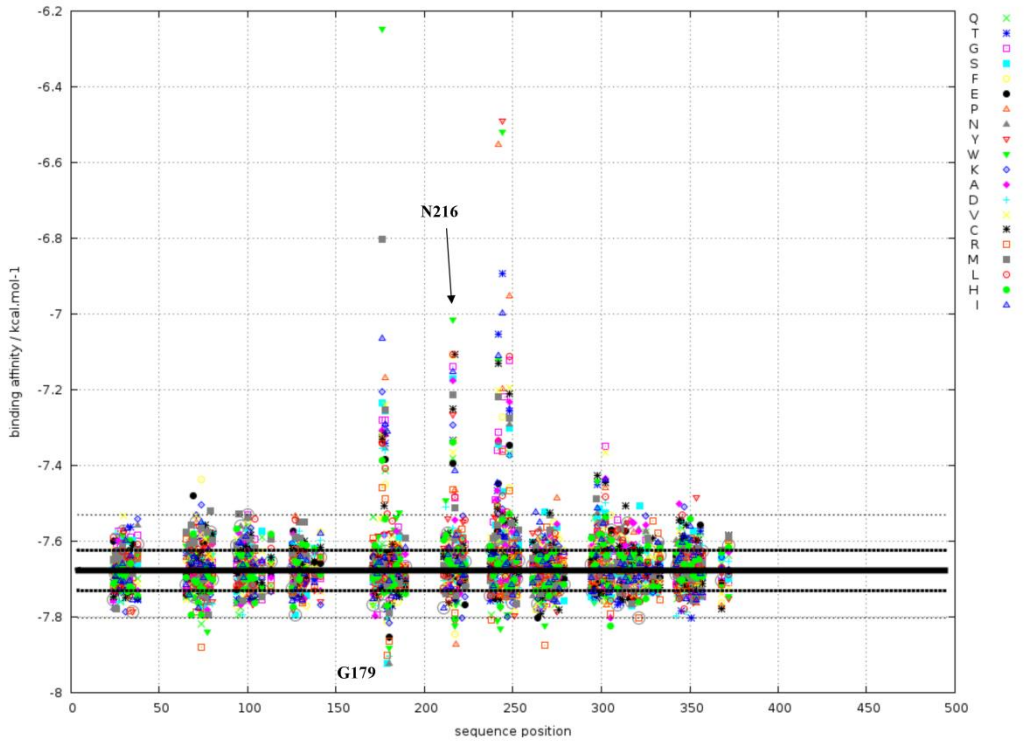


Fig. SI 4. *BindScan* run 3. (Binding energy-mutation) profile of the covalent intermediate *Tx*Abf-β-L-Ara*f* in complex with xylotriose substrate in the non-reducing *O*-3 orientation for $A^{3}XX$ product formation.

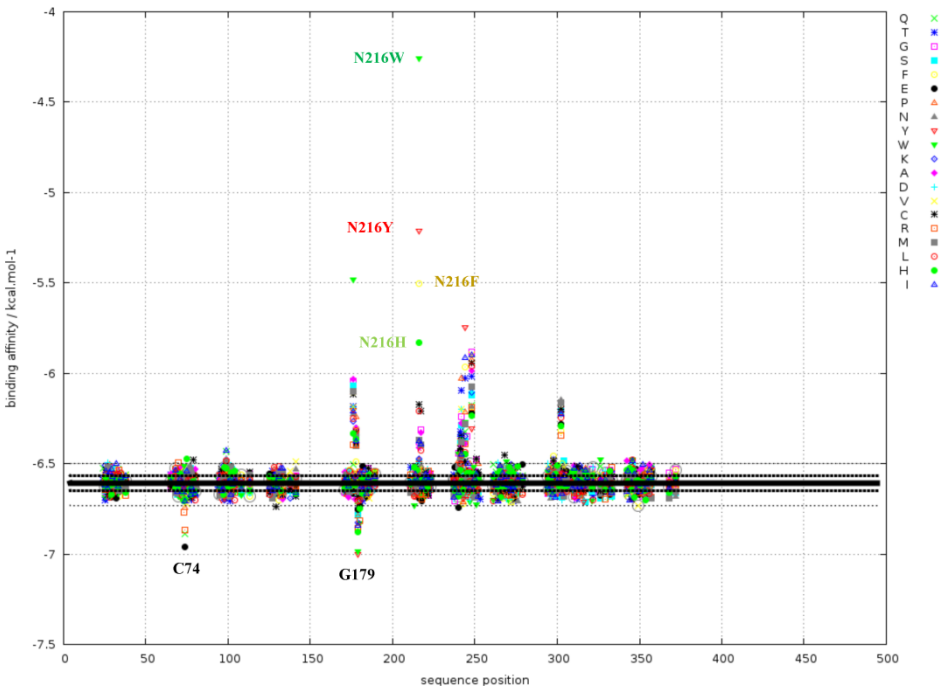


Fig. SI 5. *BindScan* run 4. (Binding energy-mutation) profile of the covalent intermediate *Tx*Abf-β-L-Ara*f* in complex with xylotriose substrate in the non-reducing *O*-2 orientation for A^2XX product formation.

Table SI 1. Impact of (donor:acceptor) ratio on L-Ara*f* transfer rates[a]

Donor:Acceptor Ratio (mM/mM)	Transglycosylation product	R_T (µmol P for 100 µmol S)			
		XA^3X	A^3XX	A^2XX	Auto[b]
	δ (ppm)	5.40	5.32	5.28	5.96
5:10	wt	3.2	3.5	5.8	4.7
	L352M	15	15	32	6
	Increase Factor	4.7	4.2	5.5	1.3
15:10	wt	3.3	6.2	6.8	17.7
	L352M	10.6	11.0	25.7	16.8
	Increase Factor	3.2	1.8	3.7	1.0
15:30	wt	7.7	13.5	14.3	11.3
	L352M	16.8	15.9	43.9	8.6
	Increase Factor	2.2	1.2	3.1	0.8

[a] Reactions were performed at 45 °C, in deuterated sodium acetate buffer (25 mM, pD 5.8), with *p*NP-α-L-Ara*f* as donor and xylotriose as acceptor.
[b] no standard is available to assign this chemical shift to a given linkage, putatively corresponding to (1,3) regioisomer.

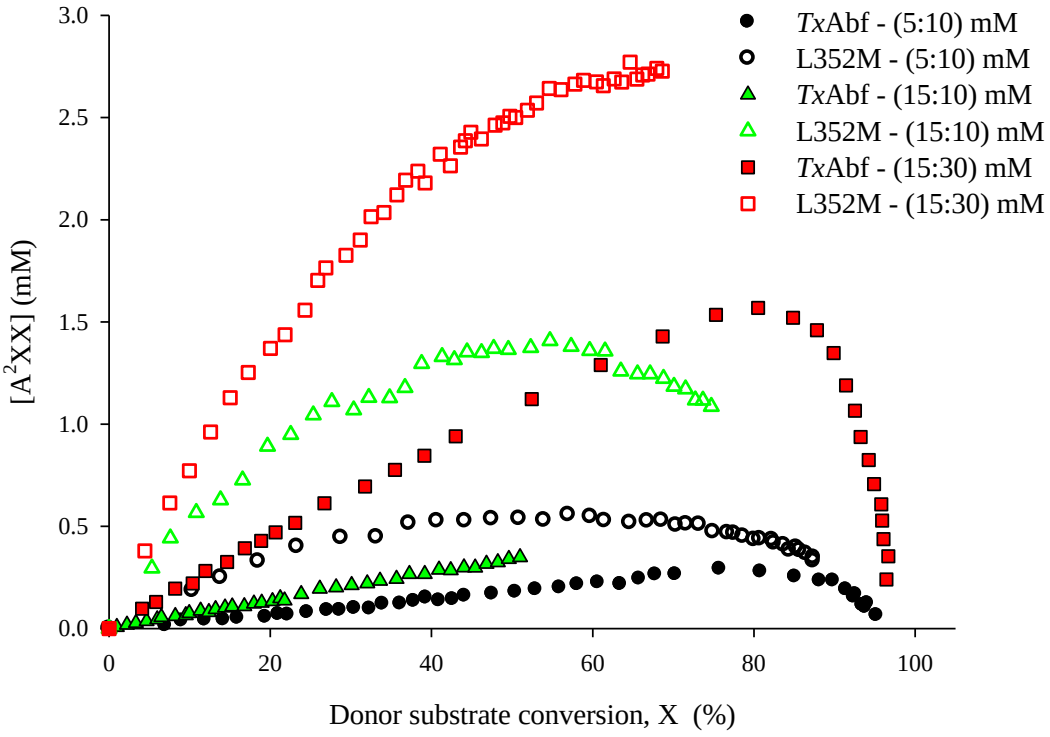


Fig. SI 6. Impact of donor:acceptor concentration and ratio on the quantity of enzymatically synthesized A^2XX. Data are represented as a time-course monitoring of A^2XX (mM) produced by transglycosylation in presence of *p*NP-α-L-Ara*f* (donor) and xylotriose (acceptor) catalyzed by *Tx*Abf (closed symbols) or L352M (open symbols) using different (donor:acceptor) molar ratios.

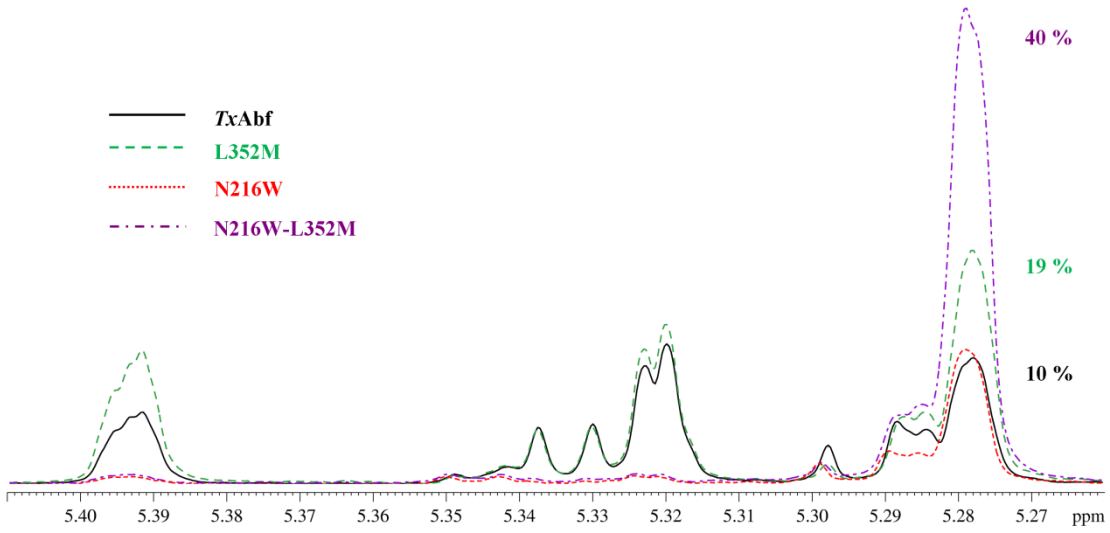


Fig. SI 7. Transglycosylation potency comparison of *Tx*Abf, L352M, N216W and combined N216W-L352M mutants. Time-course NMR monitoring experiments were carried out at 45 °C with a 15:30 (mM/mM) ratio of *p*NP-α-L-Ara*f*:xylotriose. Transglycosylation products, through their H-1 of L-Ara*f* signals, can be observed at 5.39 ppm (XA^3X), 5.32 ppm (A^3XX) and 5.28 ppm (A^2XX). The superimposed ^1H NMR spectra correspond to maximum A^2XX transglycosylation yield obtained for each enzyme.

Table SI 2. Transfer rate analysis of *BindScan* mutants and recombinant ones[a]

Enzyme		Transglycosylation product (δ in ppm)						
		XA^3X	A^3XX	A^2XX	Auto 1[b]	Auto 2[b]		
Variant	[Enzyme] (nM)	5.4	5.32	5.28	5.96	5.89	R^2 values	X error (%)
*Tx*Abf	18	7.3	12.1	13	11.9	nd	> 0.99	1.6
L352M	50	17.6	17.3	41.8	8.4	nd	> 0.98	1.5
G179F	50	12	nd	24	11.6	nd	> 0.97	1.4
G179F-L352M	50	16.8	nd	55.9	7	nd	> 0.98	1.3
G179W	50	9.3	2.7	16.7	21.9	7.7	> 0.97	1.4
G179W-L352M	100[c]	21.8	2.9	46.7	15.5	nd	> 0.98	1.7
G179W-N216W-L352M	50	nd	nd	17.1	nd	36.4	> 0.98	1.2
G179Y	50	10.1	nd	26.3	12.3	nd	> 0.97	1.4
G179Y-L352M	150[c]	14.8	nd	51.9	7.8	nd	> 0.95	1.8
G312E	50	10.5	5.7	17.2	8.1	nd	> 0.98	1.0
G312E-L352M	50	18.0	4.6	38.1	6.3	nd	> 0.98	1.0
N216W	19	np	np	8.6	nd	36.6	> 0.96	2.0
N216W-L352M	50	np	np	38.8	6.7	21.0	> 0.99	1.0

[a] Reactions were carried out in deuterated sodium acetate buffer (pD 5.8, 25 mM), at 45 °C.

[b] Auto 1 and auto 2 correspond to two regioisomers of self-condensation products (Rémond et al., 2005)

[c] These reactions were first done with 50 nM of enzyme but an early plateau phase was reached. For G179W-L352M, the maximum donor conversion (X_{max}) reached 53% with 50 nM whereas X_{max} = 77% with 100 nM. For G179Y-L352M, X_{max} reached 23 and 40% with 50 and 150 nM respectively.

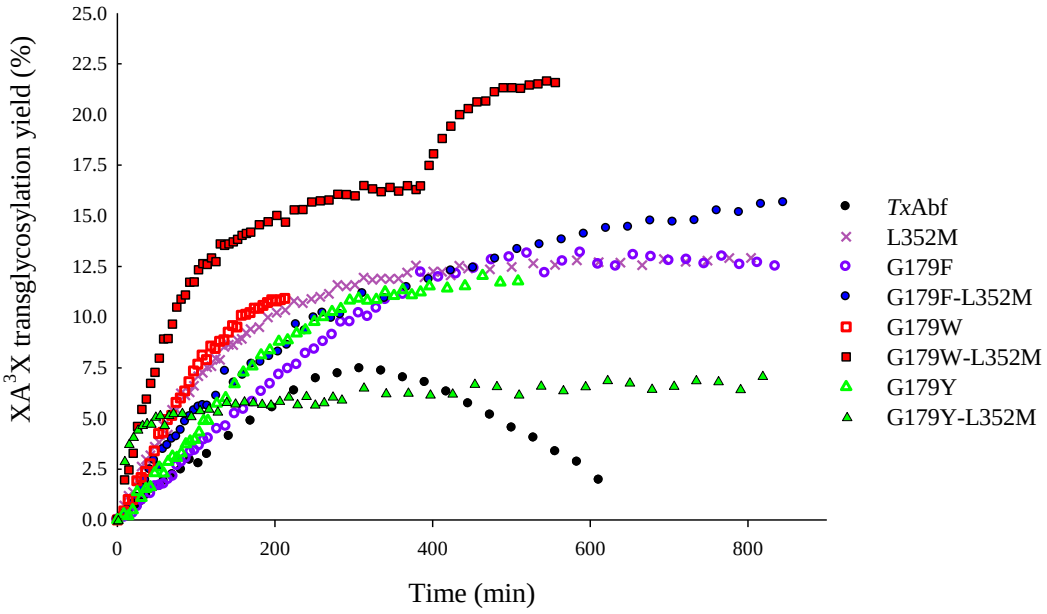


Fig. SI 8. Impact of G179 mutants and derivatives on XA^{3}X transglycosylation products formation. Data are represented as a time-course monitoring of XA^{3}X in reactions catalyzed by *Tx*Abf (18 nM), G179F (50 nM), G179W (50 nM), G179Y (50 nM), L352M (50 nM), G179F-L352M (50 nM), G179W-L352M (100 nM) and G179Y-L352M (150 nM) mutants, with *p*NP-α-L-Ara*f* as donor (15 mM) and xylotriose as acceptor (30 mM), at 45 °C in deuterated sodium acetate buffer (pH 5.8, 25 mM).

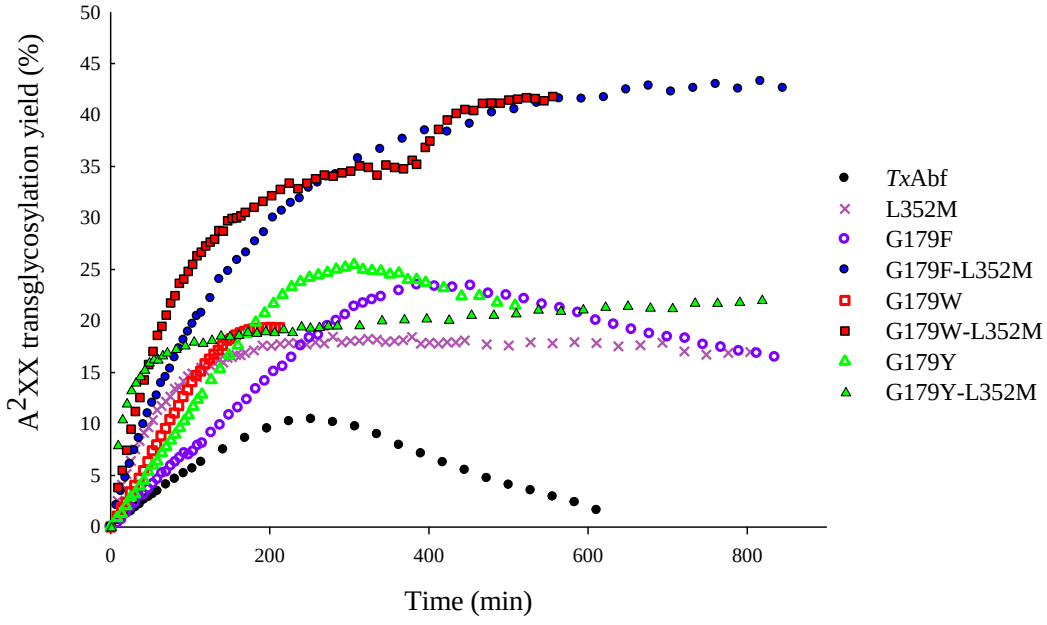


Fig. SI 9. Impact of G179 mutants and derivatives on A^{2}XX transglycosylation products formation. Data are represented as a time-course monitoring of A^{2}XX in reactions catalyzed by *Tx*Abf (18 nM), G179F (50 nM), G179W (50 nM), G179Y (50 nM), L352M (50 nM), G179F-L352M (50 nM), G179W-L352M (100 nM) and G179Y-L352M (150 nM) mutants, with *p*NP-α-L-Ara*f* as donor (15 mM) and xylotriose as acceptor (30 mM), at 45 °C in deuterated sodium acetate buffer (pH 5.8, 25 mM).

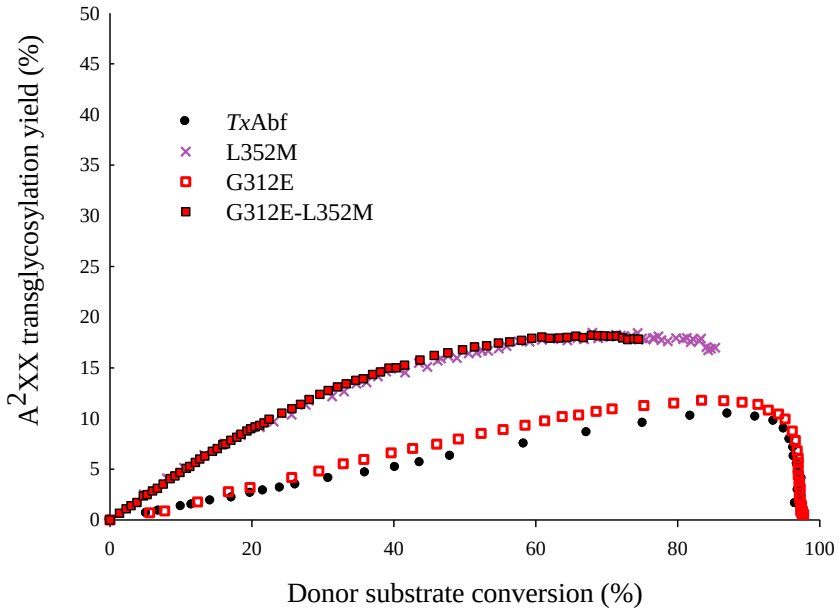


Fig. SI 10. G312E impact on A^{2}XX transglycosylation product formation. Reactions were catalyzed by *Tx*Abf (18 nM), G312E (50 nM), L352M (50 nM) and G312E-L352M (50 nM) mutants, with *p*NP-α-L-Ara*f* as donor (15 mM) and xylotriose as acceptor (30 mM), at 45 °C in deuterated sodium acetate buffer (pD 5.8, 25 mM). Data are represented as a *p*NP-α-L-Ara*f* conversion-dependent plot.

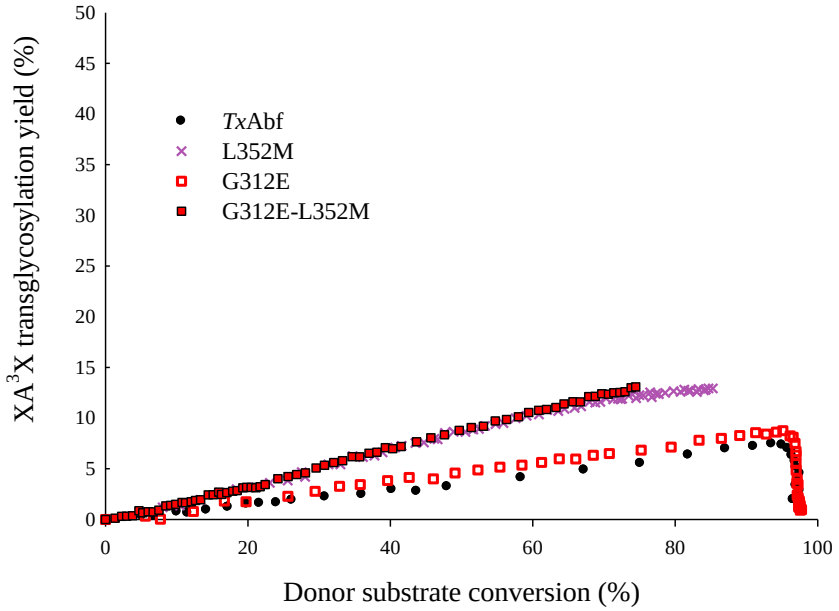


Fig. SI 11. G312E impact on XA^{3}X transglycosylation product formation. Reactions were catalyzed by *Tx*Abf (18 nM), G312E (50 nM), L352M (50 nM) and G312E-L352M (50 nM) mutants, with *p*NP-α-L-Ara*f* as donor (15 mM) and xylotriose as acceptor (30 mM), at 45 °C in deuterated sodium acetate buffer (pD 5.8, 25 mM). Data are represented as a *p*NP-α-L-Ara*f* conversion-dependent plot.

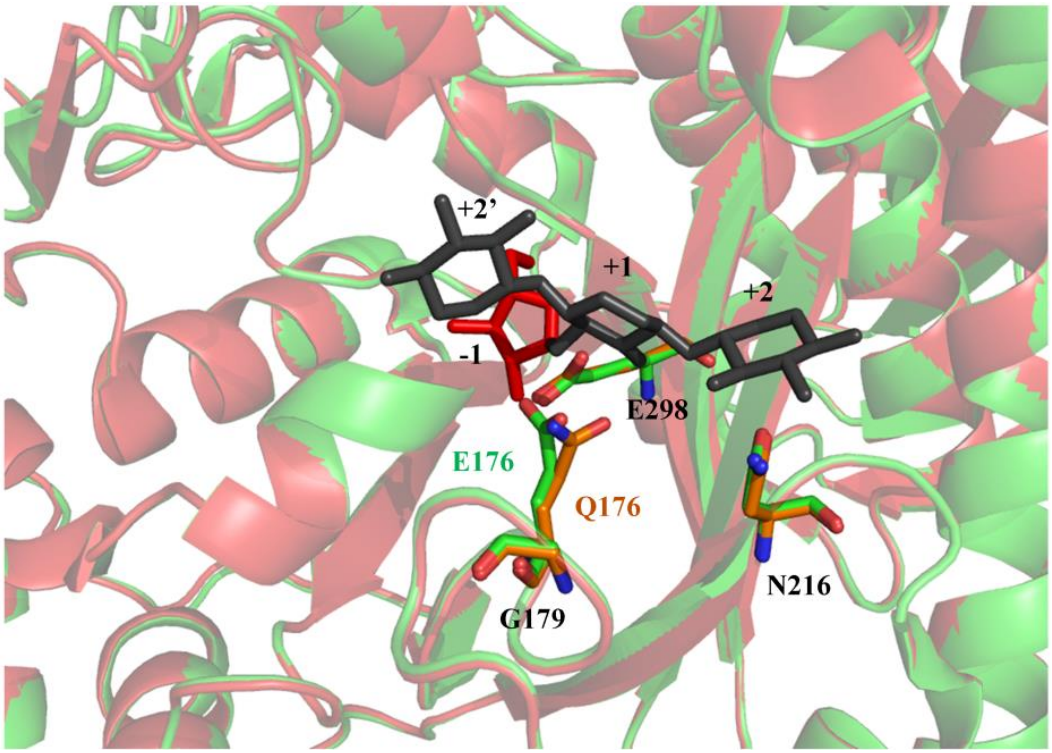


Fig. SI 12. Comparison of E176 and Q176 side chains positioning. E176 is the natural catalytic acid/base residue (green stick), which was mutated into Q176 (orange stick) for co-crystallization purposes (Paës et al., 2008). Within the 3D-structure containing E176 (PDB 2VRK chain A, green) the β2α2 loop adopts an open conformation. Therefore, for the present superimposition a 3D structure containing Q176 in which β2α2 loop adopts a similar positioning was chosen (PDB 2VRQ chain C, orange).

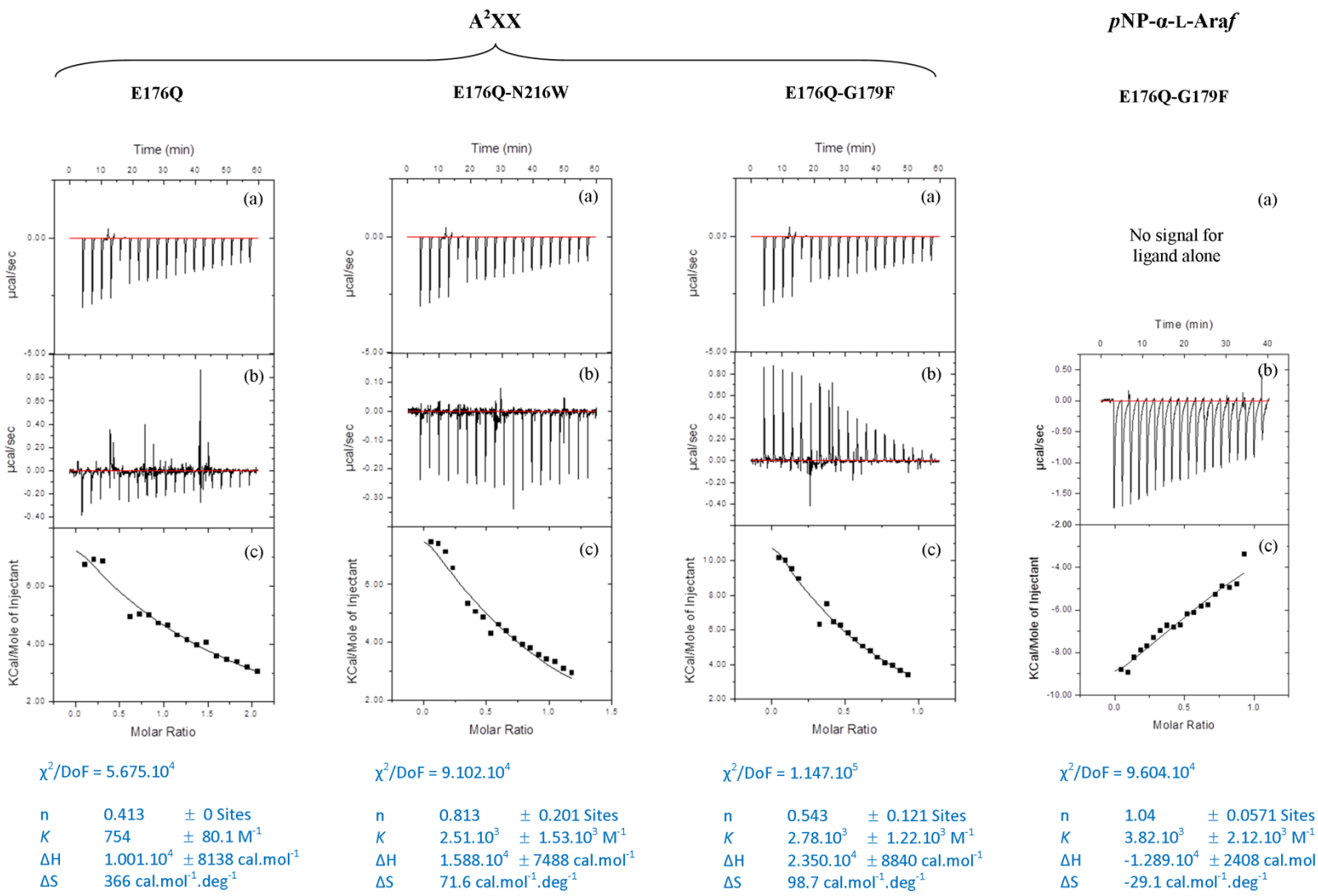


Fig. SI 13. Binding interaction analyses by ITC. E176Q, E176Q-G179F and E176Q-N216W mutants were titrated with *p*NP-α-L-Ara*f* or A²XX. Only those represented gave exploitable results: a) control titration to correct for ligand signal (i.e. no protein), b) titration with protein and ligand, c) integrated data after correction.

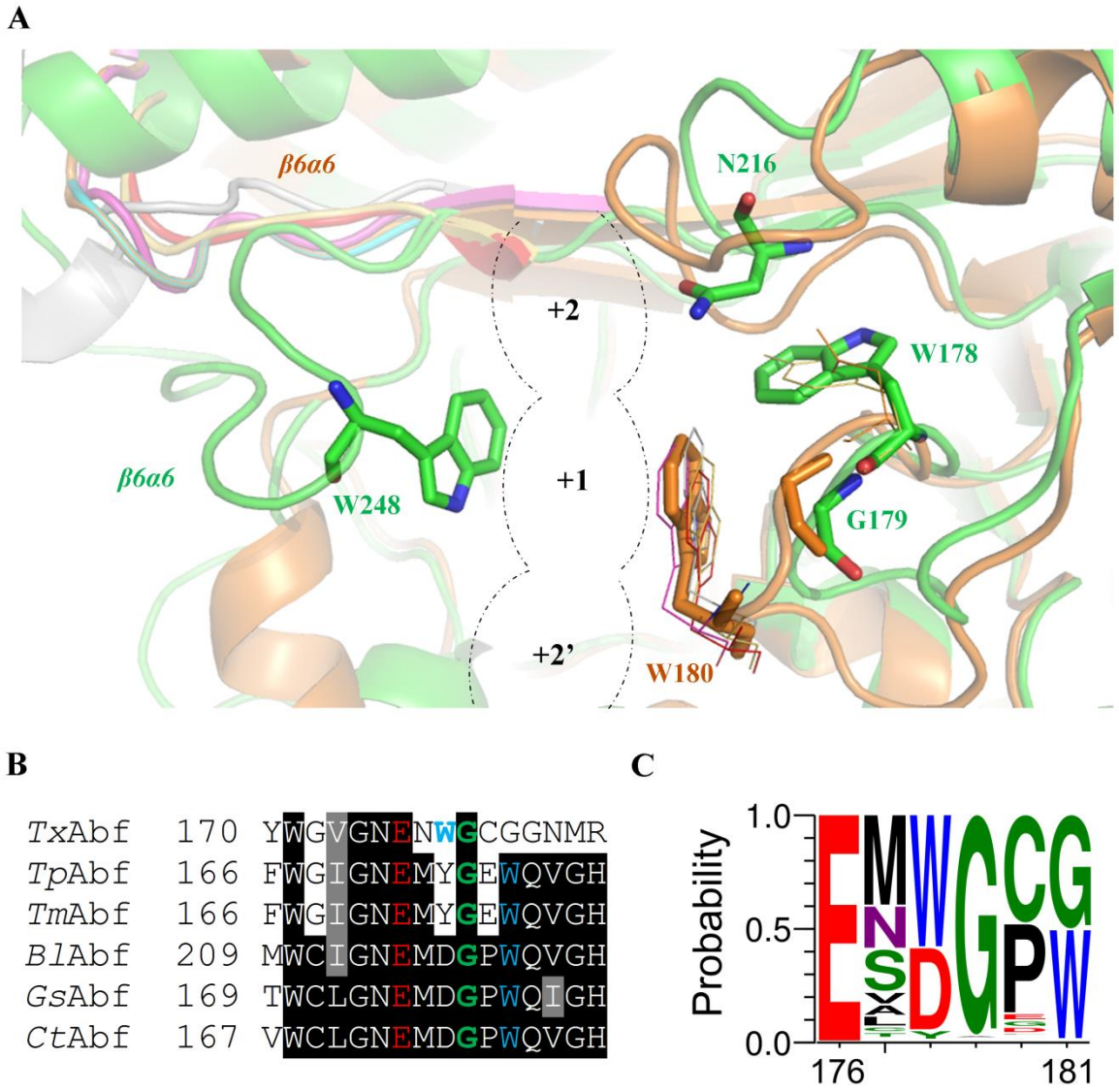


Fig. SI 14. (A) Location of G179 in *Tx*Abf active site (PDB 2VRQ, in green) superimposed with *Gs*Abf (PDB 1QW8, in orange) showing the presence of tryptophans thought to be involved in stacking interactions with the +1 D-Xyl*p* unit. The tryptophan involved in "parallel stacking" is located on the right hand side for *Gs*Abf (W180, orange stick) whereas it is on the left hand side for *Tx*Abf (W248, green stick), located on β6α6 loop, which is much shorter for *Gs*Abf-like enzymes. Were superimposed structures of Abfs from *Geobacillus stearothermophilus* (*Gs*Abf, orange, PDB 1QW8, (Hovel et al., 2003)), *Bifidobacterium longum* (*Bl*Abf, grey, PDB 2Y2W, (Lagaert et al., 2011)), *Clostridium thermocellum* (*Ct*Abf, pink, PDB 2C8N, (Taylor et al., 2006)), *Thermotoga maritima* (*Tm*Abf, yellow, PDB 3UG5, (Im et al., 2012)), *Thermotoga petrophila* (*Tp*Abf, red, PDB 3S2C, (Souza et al., 2011)). The introduction of G179F, W or Y mutations in *Tx*Abf probably creates a much narrower groove for acceptor positioning. **(B) Sequence alignment of the six GH51 Abfs for which a structure is available. (C) WebLogo constructed with an alignment of 486 sequences belonging to GH51 Abfs family.** This shows that G179 is a highly conserved position. On a structural standpoint, one can see that in *Gs*Abf-like cases, the presence of W180 (*Gs*Abf numbering) may preclude the introduction of any aromatic substitution at equivalent G179 position (*Tx*Abf numbering).

Part IV

Molecular Design of the First 'non-Leloir' Transarabinofuranosylases

(Article ready to be submitted)

This part of my thesis work is the logical follow up of the previously described results, in which I attempted to take into account all the knowledge acquired on *Tx*Abf to make it evolve in a smarter way. For this, based on the deep characterization realized for N344Y and L352M single mutants we wondered if these mutants could represent good platforms for further improvement through random mutagenesis. Therefore, two libraries (Alpha 2 and Alpha 3) were created and screened, by notably using an “automated kinetics assay” that was developed for this purpose. From these additional rounds of directed evolution, whereas the Alpha 2 library (N344Y) did not provide interesting transglycosylating enzymes, the Alpha 3 library (L352M) screening led to the identification of a new mutation, R69H, allowing a several-fold acceptor-mediated activation. A deconvolution allowed to characterize the role of each mutation (R69H and L352M) to understand their involvement in the T/H modulation. In the perspective of creating a transfuranosidase R69H and L352M were then recombined with acceptor region mutations, G179F (*BindScan*) or N216W (regioselectivity study), known to both increase the overall transglycosylation yield and provide (1$\rightarrow$2) regiospecicity for the latter one.

This case constitutes the first example of a trans(arabino)furanosidase creation obtained through molecular evolution by combining mutations found by random, semi-rational and *in silico* approaches.

MOLECULAR DESIGN OF A 'NON-LELOIR' FURANOSE-TRANSFERRING ENZYME FROM AN α-L-ARABINOFURANOSIDASE

Bastien Bissaro[1,2,3], Julien Durand[1,2,3], Xevi Biarnés[4], Antoni Planas[4], Pierre Monsan[1,2,3,5], Régis Fauré[1,2,3] and Michael J. O'Donohue[1,2,3,*]

[1] *Université de Toulouse; INSA,UPS,INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France*
[2] *INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France*
[3] *CNRS, UMR5504, F-31400 Toulouse, France*
[4] *Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Via Augusta, 08017 Barcelona, Spain*
[5] *Toulouse White Biotechnology, UMS INRA/INSA 1337, UMS CNRS/INSA 3582, 3 Rue des Satellites 31400 Toulouse, France*

Keywords: Transglycosylation, molecular evolution, glycoside hydrolase, enzyme mechanism, pentose-based compounds.

ABSTRACT: Mastering the synthesis of glyco-conjugates and branched oligosaccharides is of utmost interest since those are involved in a myriad of life-sustaining or threatening biological events. Enzymatic approaches offer invaluable features regarding the control of stereo-and regio-selectivity of the glycosylation reaction. In this respect, the vast biodiversity of glycoside hydrolases (GHs) constitute a promising reservoir of starting material for their evolution toward glycosynthetic enzymes, which, unlike glycosyltransferases, are easy to produce and employ simple substrates. However, GHs usually display an unbalanced hydrolysis/transglycosylation partition in favor of hydrolysis and the modulation of this ratio is nowadays hampered by a lack of knowledge on the actual molecular determinants involved in it. Such a challenge was tackled through the molecular evolution of the hydrolytic α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*Tx*Abf), which constitutes today a paradigm for furanose-acting enzymes. In this study is presented how the first 'non-Leloir' transarabinofuranosylase was obtained by the combination of random mutagenesis/screening iterative steps and knowledge previously acquired through semi-rational as well as *in silico* approaches. The random strategy led to the identification of a residue, R69, which is actually a highly conserved position among GHs from clan A, which permits the introduction of a quasi-strict transglycosylating phenotype in TxAbf. Indeed, transfer rates close to 100% and overall transarabinofuranosylation yields reaching 80% (vs 10% for the wild-type) are obtained when recombined with rationally-found mutations. Beyond this achievement, in the light of already available, albeit scattered, literature knowledge, the thorough mechanistic study realized on these mutants allowed to draw a more general picture of hydrolytic GH conversion on the way to transglycosylation.

1. INTRODUCTION

Considering the importance of carbohydrates in biological processes, both in molecular recognition and the storage of energy, the improvement and widening of our ability to synthetize glycosidic motifs and glycoconjugates is of utmost importance. In this respect, chemo-enzymatic approaches are proving to be highly useful alternatives to purely chemical methods, although these have so far focused mainly on hexopyranoses, leaving furanoses aside, despite the fact that the synthesis of furanosides has often provided to be quite tricky.[1,2]

Carbohydrate active enzymes (CAZymes), such as glycosyltransferases and so-called retaining glycosides hydrolases (GHs) are the workhorses of chemo-enzymatic approaches that have been developed for the synthesis of complex carbohydrates assembly.[3–5] In the case of glycosyltransferases,the vast majority is designated as sugar nucleotide-dependent (Leloir) enzymes.

Regarding GHs, these have proven to be especially attractive for chemo-enzymatic strategies, because they are highly abundant, display a wide range of substrate specificities and often employ readily accessible, simple glycosides as donor compounds. The synthetic capability of GHs resides in their double-displacement mechanism, the outcome of which can be either hydrolysis or transglycosylation, the latter being dependent on the relative availability and the recognition of acceptors other than water in the reaction medium (Figure 1).[6] Nevertheless, most GHs principally display hydrolytic activity in aqueous medium, and only a minority of GHs is better described as non-Leloir transglycosylases (i.e., using 'glycosidic aglycon'). These latter GHs catalyze the formation of glycosidic bonds, even at high water activity (in the reaction medium), and display a low (sometimes undetectable) propensity to perform hydrolysis. Intriguingly, non-Leloir transglycosylases are scattered among several CAZy GH families[7] (e.g., XET from GH16, amylosucrases from GH13, trans-sialidases from GH33), although no non-Leloir transfuranosylases have been described to date. Significantly, such transglycosylases are very frequently highly homologous to their

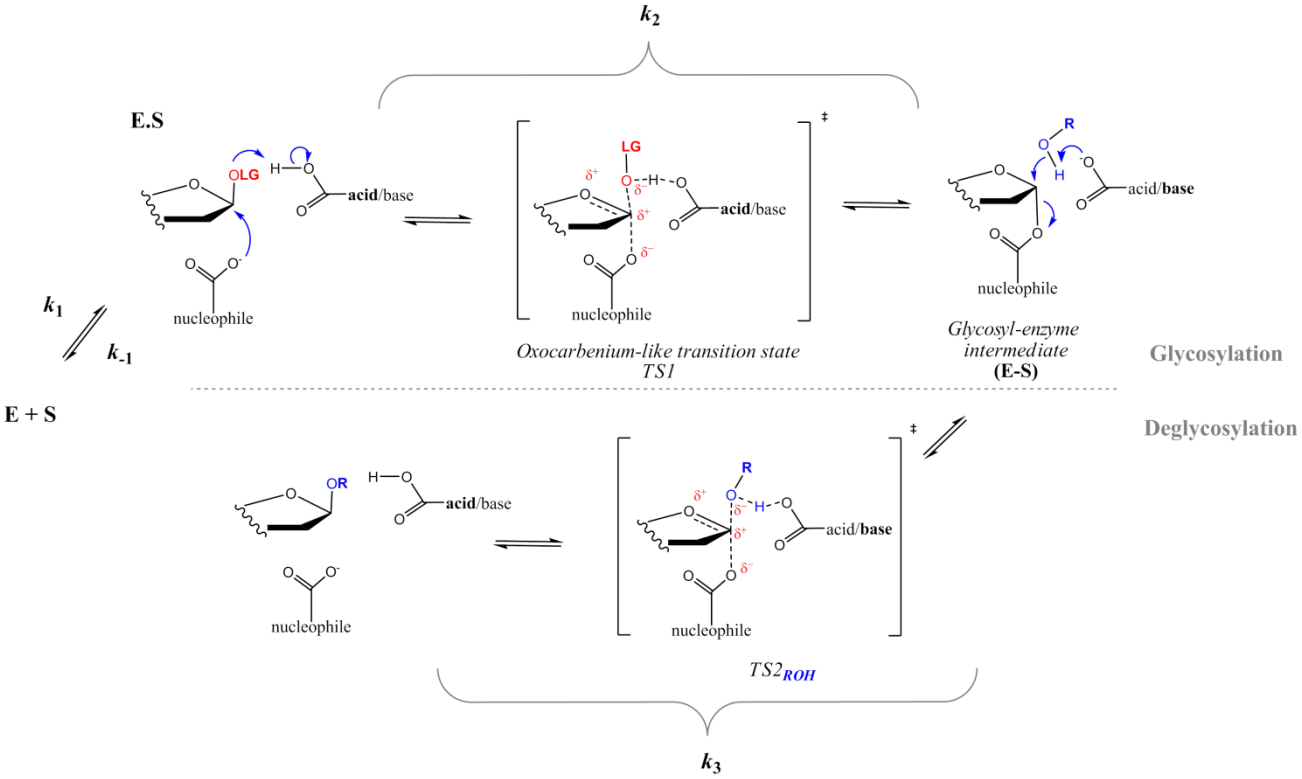


Fig. 1. Two-step displacement mechanism of retaining GHs. Along the reaction mechanism, a Michaelis complex (E.S) is formed between the enzyme and the donor substrate allowing a nucleophilic attack of the catalytic nucleophile residue onto the anomeric carbon, combined with a protonation of the interglycosidic oxygen by the catalytic acid/base residue, leading to a release of the leaving group (LG) and formation of the covalent glycosyl-enzyme intermediate (E-S). This step is so-called the glycosylation step. The donor LG can be either an activated group (e.g., *p*NP) or a sugar (e.g., fructose for glucansucrases). Regarding the second step, denoted as deglycosylation step, the covalent glycosyl-enzyme intermediate can be either attacked by a water molecule (hydrolysis, R = H) or an external hydroxylated acceptor (transglycosylation, R = sugar, alkyl chain, etc.), that are activated by the deprotonated form of the acid/base catalytic residue. In the case of secondary hydrolysis LG becomes R (donor substrate = transglycosylation product) and R = H (hydrolysis). In this study the donor substrate is *p*NP-α-L-Ara*f* and the acceptor is xylotriose.

hydrolytic family counterparts (e.g. amylosucrases/amylases from GH13). The implication of this observation is that the outcome of catalysis (i.e., hydrolysis or transglycosylation) is not determined by any single active site topology or protein 3-D structure, but is probably associated with more subtle molecular features which have yet to be revealed.

The quest to understand the underlying molecular determinants of the hydrolysis/transglycosylation (H/T) partition in retaining GHs has occupied researchers for at least 20 years and has given rise to several hypotheses that are mostly only applicable to single GH families.[8–12] Consequently, no overarching theories are currently available to provide more generic knowledge that would allow the rational design of non-Leloir transglycosylases using protein engineering. One way to tackle the conundrum of non-Leloir transglycosylases is to question the way in which molecular evolution has driven the appearance of retaining GHs that can perform glycosynthesis in environments in which water is ubiquitous. To do this, one can compare the amino sequences of GHs within single CAZy families. However, it is likely that a small number of amino changes relevant to the phenotype under investigation will be hidden by a larger number of amino changes that are either associated with other phenotypic differences, or are simply the result of neutral sequence drift. Another way to tackle the question is to analyze GHs that have been artificially evolved *in vitro*, selecting for the ability to perform transglycosylation in preference to hydrolysis. In the present study, the latter strategy has been developed.

For over more than one decade, our group has focused considerable effort on the study of GH51 α-L-arabinofuranosidase from *Thermobacillus xylanyliticus* (*Tx*Abf), which is a retaining GH that hydrolyzes the bonds linking α-L-arabinofuranosyl moieties to D-xylopyranosyl units in arabinoxylans and arabinoxylo-oligosaccharides (AXOS) of plant origin.[13] During the course of our studies we have revealed that, in the presence of appropriate sugar or alcohol acceptors, the H/T partition can be driven towards transglycosylation, allowing the synthesis of a variety of furanose-based compounds, including arabino- and galactofuranosides.[14,15] Moreover, in recent work, we used *in vitro* molecular evolution in a first attempt to artificially tip the H/T balance further towards transglycosylation and thus create a novel glycosynthetic tool for the synthesis of difficult to access complex furanoside-based structures.[16,17]

In the present study, we describe the pursuit of our work on *Tx*Abf, and in particular the artificial evolutionary pathway that has finally led us to our goal, creating what can best be described as the very first non-Leloir transarabinofuranosylases. More importantly, we also describe the thorough characterization of these mutant enzymes and provide mechanistic insight into the shift from hydrolysis toward transglycosylation.

2. EXPERIMENTAL SECTION

2.1. Substrates and chemicals

Unless otherwise stated, routine experimental work was performed using chemicals purchased from Sigma-Aldrich (Illkirch, France) and molecular biology reagents purchased from New England Biolabs (Evry, France). The substrate, *p*NP-α-L-Ara*f*, was purchased from CarboSynth (Compton, UK) and xylotriose and XOS were purchased from Wako Chemicals GmbH (Neuss, Germany). Regarding aryl arabinofuranosides used in this study, these were synthesized *in-house* using a modified version of a previously described synthetic route.[18] Since the efficiency of aromatic O-glycosidations strongly depend on the reactivities of different donor-acceptor combinations,[19] activated precursors were synthesized using either trichloroacetimidate (for phenol, Ph) or bromide (for 3,4-dinitrophenol, 3,4-dNP, 2-nitro-4-chlorophenol, NCP, and 3-nitrophenol, 3-NP).

Working under nitrogen, the aryl group was partially dissolved in anhydrous DCM in a 2 neck-flask equipped with a pressure-equalizing dropping funnel. With the aid of a syringe, $BF_3.Et_2O$ (1 eq) was added to the stirred yellow suspension at room temperature. Then, the trichloroacetimidate[18] (1.5 eq) was dissolved in anhydrous DCM with activated molecular sieves (4 Å), and then slowly transferred to the reaction mixture over 20 min using the dropping funnel. The funnel was rinsed twice with anhydrous DCM, which was further added to the reaction. After stirring at room temperature for 4 h, the mixture was neutralized with triethylamine. The residue was then first washed with a saturated solution of $NaHCO_3$, then with a small amount of water, and finally with brine. The aqueous phases were extracted twice with DCM and the combined organic extracts were dried ($MgSO_4$), filtered and concentrated. The residue was then re-crystalized from ethanol affording phenyl-2,3,5-tri-O-benzoyl-α-L-arabinofuranoside as a white powder after filtration and drying.

For the three other aryl groups, working under nitrogen, the aryl group, Ag_2CO_3 (1.15 eq) and 2,6-lutidine (1 eq) were partially dissolved in anhydrous acetonitrile in a 2 neck-flask equipped with a pressure-equalizing dropping funnel, at 75 °C. Then, the bromide[20] (0.95 eq) was dissolved in anhydrous acetonitrile with activated molecular sieves (4 Å), and then slowly transferred to the reaction mixture over 20 min using the dropping funnel. The funnel was rinsed twice with anhydrous acetonitrile, which was added to the reaction. After stirring overnight at 75 °C, the mixture was filtered on Celite®545 (Sigma-Aldrich) and concentrated. The residue was dissolved in DCM and first washed with a saturated solution of KH_2SO_4, then twice with $NaHCO_3$, and finally with brine. The aqueous phases were extracted twice with DCM and the combined organic extracts were dried ($MgSO_4$), filtered and concentrated. The yellow residue was purified on a silica column using the Reveleris flash chromatography System (Grace, Epernon, France), with a gradient of DCM in petroleum ether, with elution at 50%.

Debenzoylations was performed in a mixture of anhydrous MeOH and DCM (2:1, v/v) with MeONa (1M in methanol, 0.5 eq). For NCP-α-L-Ara*f* and 3,4-dNP-α-L-Ara*f*, a post-deprotection purification on flash chromatography column was needed, using a gradient of MeOH in DCM, with elution occurring between 5 and 10% in MeOH.

2.2. Mutagenesis, protein expression, and purification

2.2.1. Site-directed mutagenesis

The plasmid pET24a-*Tx*Abf (the original pET vector was from Novagen, Fontenay-sous-Bois, France) was used as template DNA for *in vitro* mutagenesis using the QuikChange Site-Directed Mutagenesis kit (Agilent, Courtaboeuf, France), following the manufacturer's instructions. For recombination of mutations R69H, L352M and G179F, the following primers (5'-3') were employed (underlined codon and mutated base in bold):

R69H	TCCGGTCCTC**CAC**TGGCCGGGCG
G179F	GGCAACGAGAACTGG**TTC**TGCGGCGGCAACAT
L352M	CAGCTCGTCAACGTG**ATG**CAATCCGTCATCC

The successful introduction of mutations was confirmed by DNA sequencing (GATC Biotech, Mulhouse, France).

2.2.2. Large scale enzyme expression and purification

To produce larger batches of wild-type and mutated recombinant *Tx*Abf, protein expression and purification were performed as previously described.[21]

2.2.3. Creation and screening of random mutant libraries

Design of random libraries

A randomly mutated gene library was constructed using pET24-*Tx*Abf-L352M as DNA template. For mutagenesis, the GeneMorph II Random Mutagenesis kit (Agilent) was employed following the suppliers recommendations. To generate optimal genetic diversity, three separate reactions were performed (50 µL each) in which the quantities of pET24-*Tx*Abf-L352M template DNA differed (50, 100 and 150 ng), but the amounts of dNTP mix (800 µM), primers mix (250 $ng.\mu L^{-1}$ each) and Mutazyme II DNA polymerase (2.5 U) were identical. The reactions were performed using a PCR thermocycler and the following protocol: one initial denaturing step at 95 °C for 2 min, followed by 25 cycles at 95 °C for 1 min, 55 °C for 1 min, 72 °C for 2 min and a final elongation step at 72 °C for 10 min. For the PCR reactions, the sequences of the forward and reverse

primers were 5'-GGAATTCCATATGAACGTGGCAAGCCGGGTAGTCG-3' and 5'-CGATGGCCCACTACGTGAACCATC-3' respectively. The forward primer contains a *Nde*I restriction site and the reverse primer anneals to a sequence within the pET24 vector. Accordingly, the size of the amplicon was 1933 bp, representing the target *Tx*Abf-encoding sequence (1490 bp), associated with a part of the pET24 vector (433 bp). After purification using a DNA clean and concentrator kit (Zymo Research, USA), the amplicon was digested using *Nde*I (20 IU) and *Hind*III (20 IU) at 37 °C for 4 h in a buffered (1 X Neb2 buffer) reaction (50 µL). Afterwards, the reaction was stopped by heating at 65 °C for 20 min and an aliquot (20 IU) of *Dpn*I was added. Following further incubation at 37 °C for 1 h, the DNA fragments were purified and eluted in 10 µL of ddH_2O (Zymo Research kit). The resulting purified, mutated *Tx*Abf-encoding sequence was then ligated with *Hind*III/*NdeI*-digested pET-24 vector. To achieve this, five separate ligation reactions (15 µL each) were prepared in PCR tubes, each containing 55 ng of mutated *Tx*Abf-encoding DNA and 66 ng of linearized pET24 DNA (i.e., a 3:1 molar ratio), 1.5 µL of T4 DNA ligase buffer (10X) and 1 µL of T4 DNA ligase (400 IU). The reactions were incubated for 12 h at 16 °C and then 2 µL aliquots were used to transform aliquots (50 µL) of chemically competent *E. coli* Top10 cells. Overall, 24 transformation reactions were performed, after which the bacteria were spread onto LB-agar Petri dishes containing kanamycin (40 $\mu g.mL^{-1}$). Colonies were allowed to develop for 14 h at 30 °C before being suspended in 500 µL of liquid LB medium containing kanamycin, which were added to the surface of each plate surface. After recovery, the various cell suspensions were pooled and the total volume was centrifuged (5000 x g, 10 min) providing a single cell pellet, which was used to prepare purified plasmid DNA. This DNA sample constituted the L352M-based random library, which was stored at -20 °C until further use.

Screening

For library screening, 10 ng of the plasmid library DNA stock were used to transform chemically competent *E. coli* BL21 DE3 star cells, which were then grown at 37 °C for 12 h on a nitrocellulose membrane (0.45 µm, 20 x 20 cm) placed on selective LB agar medium contained within a Q-tray, thus allowing the development of small colonies (Ø < 0.5 mm). Likewise, 8 Q-Trays were prepared, representing approximately 40,000 clones. Once colonies were formed, the nitrocellulose membranes were transferred onto selective, buffered (sodium phosphate buffer 100 mM, pH 7.0, agar 15 $g.L^{-1}$) minimal medium, containing IPTG (0.4 mM) and X-α-L-Ara*f* (0.1 mM final concentration, initially solubilized in DMSO), and incubated for 2 h at 30 °C. In parallel, a control membrane was prepared on which only colonies expressing the parental L352M variant were present. After 2 h, all membranes were examined and colonies (~ 400) displaying pale blue color were selected and transferred into four 96-well microtiter plates containing 200 µL of LB-kanamycin (40 $\mu g.mL^{-1}$). Plates were incubated overnight at 30 °C with shaking (700 rpm), after which the microtiter plates were replicated by removing an aliquot and combining it with a glycerol solution (25%, v/v final concentration). The replicate was stored at -80 °C and the remainders of the original cell cultures were used to prepare cell-free lysates.

Preparation of cell-free lysates of library clones

Using the previous microtiter cultures of the selected library clones to inoculate fresh LB-kanamycin medium (500 µL) in 96-deepwell (2 mL/well) microtiter plates, precultures were prepared and grown overnight (30 °C, 700 rpm). Afterwards, 100 µL of each pre-culture was used to inoculate 900 µL aliquots of fresh LB-kanamycin contained in wells of 96-well deepwell microtiter plates. These cultures were incubated at 37 °C and 700 rpm during 2 h and then IPTG (0.5 mM final) was added and incubation was pursued (18 h incubation at 18 °C and 700 rpm). After, cells from each culture were recovered using centrifugation (2250 x g, 20 min) and suspended in 300 µL of Tris-HCl buffer (20 mM, pH 8.0) containing lysozyme (0.5 $mg.mL^{-1}$), and incubated 1 h at 30 °C, followed by freezing overnight at -80 °C. After thawing (30 °C, 30 min), the cell suspensions were heat-treated (75 °C, 30 min) and clarified lysates were prepared by centrifugation (2250 x g, 50 min, 4 °C). Finally, lysates were transferred into fresh 96-well microtiter plates and stored at 4 °C.

Secondary screening of selected clones

The 400 cleared lysates representing individual library clones were used to perform kinetic analyses, using an automated protocol, which provided the means to simultaneously perform 96 analyses. Likewise, it was possible to compare the activities of individual lysates in the presence and absence of acceptor, and thus derive an activity ratio (R), which was equal to $Vi_{(donor+acceptor)}/Vi_{(donor)}$. Using a liquid handling system (Genesis RSP 200, Tecan), reactions were prepared in wells of 96-well polypropylene PCR microtiter plates (Corning®, Sigma Aldrich). The basic reaction mixture used to determine $Vi_{(donor)}$ values was composed of 135 µL substrate solution (15 mM *p*NP-α-L-Ara*f* in 50 mM sodium phosphate, pH 7.0) and 15 µL (10% of final volume) of lysate. To determine the value of $Vi_{(donor+acceptor)}$, the kinetic was performed using the same basic reaction mixture containing additional xylotriose as acceptor (10 mM). To perform reactions, the substrate solution (with or without acceptor) was first incubated at 45 °C for 15 min in a Peltier heating device, before initiating the reaction through the addition of lysates. Once the reactions were initiated, the progression of reactions was monitored by removing 20 µL aliquots every 6 min (4 points in total, each kinetic lasting 24 min). These were immediately combined with 180 µL of 1M Na_2CO_3 and absorbance (405 nm) was determined. Using a simple, purpose-designed Excel macro the activity (IU/mL) of the different clarified lysates was calculated in the presence and absence of acceptor and activity ratios were derived for each one. The relative mean error of this method was 4% for any single lysate. For each microtiter plate, standard deviations (σ) were calculated and 1.96 σ was

represented, which is the approximate value for the 97.5 percentile point of a normal distribution.

HPAEC-PAD screening

Reactions used for the determination of kinetics in the presence of acceptor (see above) were incubated for a total of 7 h at 45 °C, before being stopped using a freeze/heat cycle. This was achieved by plunging reactions into liquid nitrogen, heating to 95 °C for 10 min and then freezing anew in liquid nitrogen. For HPAEC-PAD analysis, each sample (in 96-well microtiter plate) was diluted 10-fold in ultrapure water (MilliQ) and 20 µL was removed for injection on a Carbopac PA1 (4 x 250 mm) column, equipped with a Carbopac PA1 pre-column (4 x 50 mm) and mounted on an ICS-3000 system (Thermo scientific Dionex). Elution was performed at 1 $mL.min^{-1}$ and 10 °C using the following eluants (A): NaOH (150 mM) and (B): NaOH (300 mM) + sodium acetate (500 mM). Each analysis lasted 30 min and was composed of the following steps: (i) 100% of (A), 5 min (ii) a gradient of 0 to 11% of (B) over 15 min and (iii) a second gradient up to 40% of (B), 5 min. Finally, 100% (A) was applied for 5 min. A gold probe was used for detection and data were recorded and analyzed using Chromeleon software (Dionex).

2.2.4. N216X site saturation library design and screening

A site saturation mutagenesis experiment of the type NNS, where S is G or C, was performed at position N216. To achieve this, the *Tx*Abf-encoding sequence was first mutated at position 216, introducing a stop codon (TAG) using QuickChange II Site-Directed Mutagenesis kit (Stratagen, Amsterdam, The Netherlands) and an appropriate primer pair (forward primer is shown below). This mutated DNA was then used as a template for NNS mutagenesis, which was performed using the same mutagenesis kit and appropriate primers (forward primer is shown below).

N216Stop TGCGTGCGGCGCG**TAG**ACGGCCGACTACCA
N216X TGCGTGCGGCGCG**NNS**ACGGCCGACTACCA

Taking into account the theoretical size of the NNS library, it was considered that the selection and screening of 96 clones would cover 95% of the diversity.[22] Therefore, 96 randomly picked *E. coli* BL21(DE3) colonies were used for enzyme production in microtiter plate format as previously described. After, the 96 clarified lysates were submitted to automated kinetic analyses in the absence and presence of a mixture of xylo-oligosaccharides as acceptor as described above. One mutant revealed an increased ratio (R = 1) relative to *Tx*Abf (R=0.4), and was identified as N216W after sequencing.

2.3. Determination of kinetic parameters

2.3.1. Hydrolytic behavior

The kinetic parameters describing hydrolysis catalyzed by wild-type or mutant *Tx*Abf were determined using a continuous assay in which the release of different phenolates was monitored using either an Eon spectrophotometer (BioTek), or in the case of PhOH, a Cary 100 Bio UV-Vis spectrophotometer (Agilent). Reactions were performed at 45 °C (an arbitrary choice of temperature that constitutes a compromise between enzyme activity and stability) in wells of microtiter plates containing 50 mM sodium phosphate buffer, pH 7.0, substrate, 0.1 $mg.ml^{-1}$ BSA and enzyme, in a final reaction volume of 200 µL. Taking into account the fact that the substrates were initially dissolved in DMSO, the final DMSO concentration in reactions was always less than 2% (v/v). To perform reactions, the buffered substrate mixture (180 µL) contained in microtiter plate wells was equilibrated at 45 °C for 5 min, and then 20 µL of an appropriately diluted enzyme solution was added and the release of LG was monitored at 400 (for 4-NP, 3-NP and 3,4-dNP) or 425 nm (for NCP). For reactions involving Ph-α-L-Ara*f*, these were performed in a similar way, but in a final volume of 500 µL in quartz cuvettes, and the release of Ph-OH was monitored at 280 nm. Control reactions containing all the reactants except the enzyme were used to correct for slight spontaneous hydrolysis of the substrates, which was observed in the case of 3,4-dNP-α-L-Ara*f* and NCP-α-L-Ara*f*. Furthermore, standard curves of substrates were prepared to correct for absorbance of the phenyl group in its bound form (i.e., Ph-α-L-Ara*f*). Moreover, a standard solution of 4-NP was used to compare the response coefficient variation between the Eon and Cary spectrophotometers. This proved to be less than 6%. Initial reaction rates were determined from the linear regions of time-dependent plots, which corresponded to 5-15% consumption of the substrate. Substrates concentrations were typically varied from 0.1 to 18 mM. One unit (IU) of enzyme specific activity corresponds to the amount of enzyme releasing one µmol of LG per minute. Steady-state kinetic parameters were determined using non-linear regression functions embedded in the SigmaPlot 11.0 software (Systat software Inc, Ritme, Paris, France). Three models were employed, including the classical Michaelis-Menten model, inhibition by excess substrate model, and a modified Michaelis-Menten that accounts for donor-induced activation through self-condensation. Steady-state kinetic parameters were then analyzed using the Brønsted correlation according to equations (1) and (2).

$$\log(k_{cat}/K_M) = \beta_{LG1}.pK_a^{LG} + c1 \quad \text{(Eq. 1)}$$

$$\log(k_{cat}) = \beta_{LG2}.pK_a^{LG} + c2 \quad \text{(Eq. 2)}$$

To measure solvent kinetic isotope effects (sKIE), reactions were performed at 45 °C, pH 7.0 as described earlier, using NCP-α-L-Ara*f* in D_2O. In reactions involving either R69H or R69H-L352M, substrate concentration was 2 mM, while those involving *Tx*Abf and L352M, had a substrate concentration of 5 mM. The D_2O content of the reaction was varied (35, 50, 65, 80 and 100%) adding appropriate amounts of H_2O to reach 180 µL. The enzymes were pre-diluted in D_2O and 20 µL of enzyme solution was used in each reaction, providing a final volume of 200 µL.

To determine the effect of pH on enzymes, an analysis of k_{cat}/K_M was performed, using the previously described protocols to determine in duplicate the values of k_{cat} and K_M. In these reactions, the

concentration of *p*NP-α-L-Ara*f* was varied between 0.15 and 18 mM and pH was varied in the range 4.5 to 8.7 using 100 mM citrate-sodium phosphate buffer (pH 4.5-7.0) and 100 mM sodium phosphate buffer (7.0-8.7), with a common point being measured in both buffers at pH 7.0 for each enzyme and each substrate concentration. Control experiments in the absence of enzyme were performed at each pH value in order to monitor any spontaneous hydrolysis of the substrate during the duration of the reaction (30 min). Finally the ratio k_{cat}/K_M was plotted as a function of pH and fitted to a bell-shaped activity profile according to equation 3.[23] This analysis provided the apparent pK_a values corresponding to the nucleophile (pK_{a1}) and the acid/base (pK_{a2}) catalytic amino acids, which were determined by nonlinear least-squares fitting using Microsoft Excel 2010. When a 'tail' was observed on the basic side of the pH-dependent profile equation 4 was employed.[24]

$$\frac{k_{cat}}{K_M} = \left(\frac{k_{cat}}{K_M}\right)_{max} \cdot \left(\frac{1}{1 + 10^{pK1-pH} + 10^{pH-pK2}}\right) \quad \text{(Eq. 3)}$$

$$\frac{k_{cat}}{K_M} = \frac{\left(\frac{k_{cat}}{K_M}\right)_1 + \left(\frac{k_{cat}}{K_M}\right)_2 \cdot 10^{(pH-pK_2)}}{1 + 10^{pK1-pH} + 10^{pH-pK2} \cdot (1 + 10^{pH-pK3})} \quad \text{(Eq. 4)}$$

2.3.2. Acceptor impact on global activity

To measure the influence of acceptor substrate concentration on *p*NP release, reactions were performed in which the donor (*p*NP-α-L-Ara*f*) concentration was held constant (5 mM), while the acceptor (xylotriose) concentration was varied. When acceptor-mediated inhibition was observed, a mixed model described by equation 5 was employed to fit experimental data, whereas equation 6 was used when acceptor-mediated activation occurred.

$$\frac{V_i}{V_{max}} = \frac{\left(1/\alpha'\right).S}{\left(\alpha/\alpha'\right).K_M + S} \quad , \quad \text{with } \alpha' = 1 + \frac{[I]}{K_i'} \text{ and } \alpha = 1 + \frac{[I]}{K_i}$$

(Eq. 5)

$$= \frac{V_{max}.[Donor].[Acceptor]}{K_M^{donor}.[Acceptor] + [Donor].[Acceptor] + K_M^{acceptor}.[Donor]} \quad \text{(Eq. 6)}$$

2.4. Monitoring transglycosylation

2.4.1. Time course NMR monitoring

To monitor reactions by collecting ^{1}H NMR spectra on a Bruker Avance 500 spectrometer, enzyme solutions were first prepared by 10-fold dilution in D_2O (99.90%) and then concentration on an Amicon® Ultra filter (regenerated cellulose 10 kDa, Millipore), performing the operation twice. Afterwards, reactions were performed at 45 °C in NMR tubes containing 600 µL of buffered D_2O, containing *p*NP-α-L-Ara*f* as donor and xylotriose as acceptor, at a ratio of 1:2 (15 and 30 mM, respectively). To obtain a pD value of 5.8, deuterated acetate (Euriso-Top, France) buffer was employed, while pD 7.1 was achieved using a deuterated sodium phosphate buffer, which was prepared by dissolving sodium phosphate in D_2O, followed by lyophilisation. This two-step protocol was repeated three times to obtain sufficient deuteration. Values of pD were measured by determining pH using a glass pH electrode and then applying the equation pD = pHelectrode + 0.41.[25] To initiate reactions, an aliquot of enzyme solution (never more than 5% of total reaction volume) was added to the reaction, the quantity of enzyme (18 nM for *Tx*Abf and N216W, 50 nM for G179F and L352M, and 2000 nM for all R69H-containing mutants) being adjusted to suit the 8-12 h reaction timeframe. ^{1}H NMR spectra were accumulated semi-continuously, accumulating a series of 5.52 min scans (i.e., 32 scans) with 6 s delays between scans. Each NMR spectrum was acquired using an excitation flip angle of 30° at a radiofrequency field of 29.7 kHz, and the reference residual water signal, calibrated at δ = 4.55 ppm at 45 °C,[26] was pre-saturated during the repetition delay (with a radiofrequency field of 21 Hz). The following acquisition parameters were used: relaxation delay (6 s), dummy scans (4). The actual duration of the reaction was dependent on the enzyme used and before enzyme addition, an NMR spectrum of the reaction mixture was acquired, which served as the zero point of the reaction, from which integrals were then corrected according to the small dilution factor induced by the enzyme addition.

2.4.2 NMR kinetics analysis

Data processing was performed as previously described.[21] Briefly, the time-dependent evolution of donor (*p*NP-α-L-Ara*f*) and acceptor (xylotriose) concentrations were quantified by integrating the relative anomeric proton signals. Molar balances, based on acceptor and donor consumption, were used to convert the transglycosylation product signal integral into concentration, and to evaluate concentrations of hydrolysis products, respectively. From this, the donor substrate conversion rate (X in %) was calculated as the mean of three distinct signals; those of the *ortho* and *meta* aromatic protons (*p*NP) and that of the anomeric proton (L-Ara*f*), with X accounting for both reaction outcomes (i.e., hydrolysis and transglycosylation), and identifying self-condensation. For all time-course NMR kinetics, the absolute error mean value on X ranged from 1 to 3%. The transglycosylation yield, $Y_{(product)}$, was determined by relative integration of product anomeric proton signals, these being at 5.39, 5.32 and 5.27 ppm for XA^{3}X, A^{3}XX and A^{2}XX respectively.[27–32] Finally, $Y_{(product)}$ (%) was plotted against X (%) to provide the transfer rate (R_T) of the donor substrate to a given product (µmol of product/µmol of consumed substrate), a value which is independent of the reaction duration and thus suitable to provide a standardized comparison of the behavior of different enzymes. Finally, initial rates were derived from the linear regions in which the variation of X was from 0-25% up to 0-40%, with root mean square deviations between 0.96 and 0.99.

3. RESULTS

3.1. Detection of increased transfuranosylation amongst randomly generated mutants

An earlier random mutagenesis experiment performed on *Tx*Abf identified L352M among several mutants that

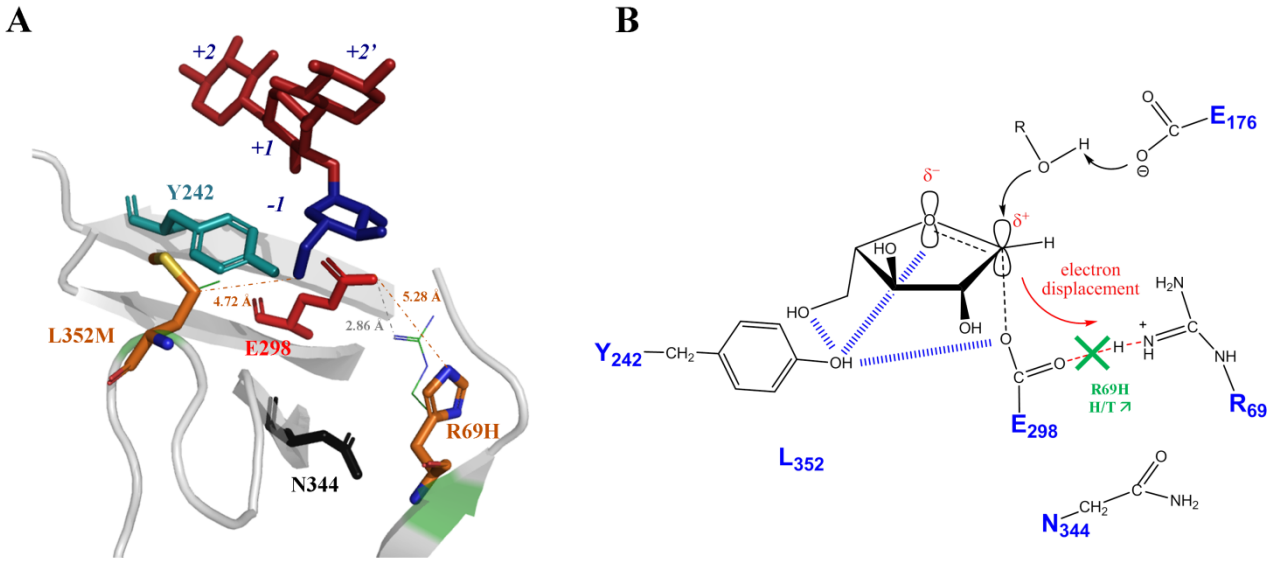


Fig. 2. (A) Location of R69H and L352M within *Tx*Abf -1 subsite and (B) putative impact of R69H mutation on electronic displacement system during transition state establishment at the deglycosylation step since R69 is at H-bond distance (2.86 Å) from the nucleophile catalyst E298, E176 playing the role of acid/base.

appeared to improve transglycosylation.[16,17] However, since the transglycosylation yield of L352M (approximately 18 %) was still rather modest, this mutant was used as a starting point for a new round of engineering. In the absence of clear-cut knowledge that would allow rational engineering we chose to pursue the random approach and thus created a new randomly-mutated library, using *Tx*Abf-L352M as the parental DNA. Sequence analysis of 24 randomly selected clones from this library revealed a mean mutation rate of 3.2 amino acid mutations/kb (Fig. SI 1) and screening of 40,000 library clones on solid medium containing the chromogenic substrate X-Ara*f* (Fig. SI 2) allowed the selection of 368 pale blue colonies. Referring to the results of previous work,[16,33] it was considered that the latter probably represented a subset of hydrolytically-impaired mutants, and thus a group of candidates for the identification of increased transglycosylation activity (Figure 1). Therefore, further *in vitro* screening was performed on these library clones in order to determine their relative activities on *p*NP-α-L-Ara*f*, in the presence or absence of the acceptor xylotriose (Fig. SI 3). Overall, secondary screening revealed a ratio mean value (μ) for the 4 micro-plates of 0.86, implying that a most clones displayed slight acceptor-mediated inhibition. Nevertheless, 4 clones appeared outside of the 95% confidence interval (μ ± 1.96 σ): 1 negative, clone 123 (Fig. SI 4), and 3 positives, clone 165, 197 and 241, forming a separate population that displayed more than 200% activity in the presence of xylotriose.

Further investigation of these clones, using HPAEC-PAD to monitor the reactions confirmed that the activity ratios of clones 165, 197 and 241 were significantly above the mean value (Fig. SI 5). DNA sequencing of clones 165, 197 and 241 revealed that each of these contained the mutation R69H, although the overall genotype was different in each case (Table SI 1), thus providing a strong indication that R69H substitution was responsible for the measured phenotype. This conclusion was further supported by the fact that examination of the molecular model of *Tx*Abf (PDB ID: 2VRQ) revealed that R69 is located close (2.86 Å) to the catalytic nucleophile (Figure 2A).

3.2. Deconvolution and recombination: synergistic effects between donor and acceptor mutations

3.2.1. Impact of R69H on the H/T ratio

To further investigate the consequences of the R69H mutation, this point mutation was introduced into the genetic backgrounds of both *Tx*Abf and the mutant L352M, thus creating mutant enzymes R69H and R69H-L352M respectively, and eliminating the extra substitution, V225M, present in clone 197. This was justified by the fact that V225 is located far from the active site (approximately 18 Å), is not very conserved (~ 31 %) in family GH51 and is substituted at equivalent positions in other GH51 members by a whole variety of amino acids (excluding cysteine, phenylalanine and tryptophan). Gratifyingly, subsequent analysis of the purified R69H-L352M enzyme confirmed the previous results obtained using the lysate of clone 197 and revealed that activity was increased up to 260% in the presence of xylotriose (Figure 3), unlike wild-type *Tx*Abf, which is inhibited (e.g., 30% of residual activity in the presence of 30 mM xylotriose). Factually, relative activity ratios are multiplied up to 7.2-fold (for R69HL352M) compared to that of *Tx*Abf, suggesting that transglycosylation might be significantly reinforced. Furthermore, the analysis of purified R69H demonstrated that this mutation alone is adequate to radically alter this ratio.

Table 1. Steady-state kinetic parameters[a] on *p*NP-α-L-Ara*f* for R69H and mutants thereof.

Enzyme	pH_{opt}	SA_{max} ($IU.mg^{-1}$)	k_{cat} (s^{-1})	K_M (mM)	Ns[b] ($s^{-1}.mM^{-1}$)
*Tx*Abf	5.8	145	139	0.25	-
L352M	6.5	243	233	8.98	-
R69H	6.5	2.383	2.287	0.09	0.049
R69H-L352M	6.5	1.012	0.972	1.23	0.031
R69H-G179F-L352M	6.5	0.556	0.535	0.69	0.021
R69H-N216W-L352M	6.5	0.603	0.580	0.48	0.034

[a] All assays were carried out at 45 °C in sodium phosphate buffer (pH 7, 50 mM) in triplicates. Relative errors were inferior to 10 %

[b] The Ns value corresponds to the non-specific constant attributed to autocondensation reaction and was obtained according the following michaelis-menten modified model: $SA = SA_{max}.[S](K_M + [S]) + Ns.[S]$.

3.2.2. Creation of acceptor subsite mutations G179F (+1) and N216W (+2)

To pursue the goal of creating a finely-tuned transfuranosylase, two other positions were targeted for mutation. The first of these, G179 located in subsite +1, was pinpointed using an *in silico* approach, which was used to identify point mutations that can potentially increase subsite +1 hydrophobicity and raise affinity for the acceptor (*Unpublished results BindScan*). Accordingly, the mutant G179F displays improved transglycosylation yield and transfer rate (Table SI 2). Furthermore, in agreement with *BindScan* initial purpose, a stronger xylotriose-mediated inhibition was observed suggesting a gain in the recognition of the acceptor (Fig. SI 6).

The second one, N216 located in subsite +2, though semi-conserved in family GH51 stood out in *Tx*Abf as a good target for site-saturation mutagenesis in the course of a previous study. Screening of the subsequent mutant library revealed one clone containing the mutation N216W, a substitution that is not found in any family GH51 members (Table SI 3) and which displayed here a quasi-total alleviation of the apparent xylotriose-mediated inhibition in the presence of xylotriose compared to *Tx*Abf (Fig. SI 6). Further analysis of a purified sample of this mutant revealed that it catalyzes the regiospecific synthesis of (1→2)-linked products (Fig. SI 7 and Table SI 2). Overall, encouraged by the results obtained with mutants G179F and N216W, a combinatorial approach was adopted, constructing triple mutants R69H-G179F-L352M and R69H-N216W-L352M (Figure 4).

3.2.3. Kinetic analyses of mutants combining R69H and L352M

Steady-state kinetic parameters were determined for the hydrolysis of *p*NP-α-L-Ara*f* catalyzed by the different combinatorial mutant enzymes bearing the substitutions R69H and L352M (Table 1). Regarding the single mutants (i.e., those containing only R69H or L352M), both displayed significantly modified behavior. In the case of L352M, a 37-fold increase of the K_M value led to a 22-fold reduction in the k_{cat}/K_M value, indicative of a modification in the glycosylation step. The behavior of the R69H mutant was radically different to that of *Tx*Abf, since the reaction was described by a biphasic profile (Fig. SI 8).

At *p*NP-α-L-Ara*f* concentrations below 5 mM, the reaction was described by the classical Michaelis-Menten model. However, at higher substrate concentrations a linear relationship was observed and, overall, the rate of the reaction was severely reduced. Regarding the double mutant R69H-L352M and the triple mutants R69H-G179F-L352M and R69H-N216W-L352M, their behavior was highly similar to that of the single mutant enzyme R69H. All displayed severely reduced specific activities and biphasic reaction profiles.

3.2.4. Measurement of the transglycosylation activity of the combinatorial mutants

Contrary to *Tx*Abf, mutants R69H-L352M, R69H-G179F-L352M and R69H-N216W-L352M revealed activation in presence of xylotriose (Figure 3), fitting to equation 6 and providing apparent $K_M^{xylotriose}$ values of 3.5, 5.9 and 45.5 mM, respectively. This results for instance in a 17.1-fold higher relative activity for R69H-N216W-L352M compared to TxAbf (Table SI 4). Moreover, when xylotriose concentration was 50 mM, the relative activity of R69H-N216W-L352M attained 700% without reaching a plateau-phase. To pursue the characterization of the transfer ability of the different mutants, the synthesis of arabinoxylo-oligosaccharides (AXOS, see Fauré et al.[34] for nomenclature) was monitored by ^{1}H NMR using *p*NP-α-L-Ara*f* (15 mM) as donor and xylotriose (30 mM) as acceptor and the results were displayed in two formats. The conversion-dependent plot (Figure 5A) constitutes the best way to compare the performance of mutants in the initial phase of the reaction, while plotting synthesis as a function of time (Figure 5B) provides a convenient method to evaluate the extent of secondary hydrolysis. Accordingly, results confirmed that the parental mutant L352M procured a 2-fold increase in transglycosylation yield compared to *Tx*Abf (Table SI 2) and displayed a significantly reduced level of secondary hydrolysis (Figure 5B). In the case of the other single mutant, R69H, the yield of A^2XX reached 31% (Table 2), but strong secondary hydrolysis at the end of the reaction period led to the degradation of the product. Nevertheless, overall R69H displayed a 3.6-fold increase in the apparent synthesis/hydrolysis rate (Rho) of A^2XX (Table SI 5). Moreover, this single mutation led to a 4.3-fold increase in transfer rate (R_T) for A^2XX and to a global R_T of 99% ($XA^3X + A^3XX + A^2XX + XA^2X$) in the

initial phase of the reaction (Table SI 6) demonstrating that in this phase, R69H confers strong transglycosylase character to *Tx*Abf. In comparison, the transglycosylation yields procured by the double mutant R69H-L352M, were slightly lower (27, 22 and 20%, for A^2XX, A^3XX and XA^3X). However, the contribution of the mutation L352M was reflected in the higher R_T value (for A^2XX) and slower secondary hydrolysis. Overall, the combination of R69H and L352M produced a Rho value of 55, which is a 27-fold increase compared to *Tx*Abf.

Analysis of the triple mutant, R69H-G179F-L352M, revealed that the presence of the mutation G179F led to a postponed secondary hydrolysis, which had a beneficial effect on the maximum yield of the principal product A^2XX (49%) and on the yields of A^3XX and XA^3X (17 and 35 % respectively). In the case of R69H-N216W-L352M, the (1→2) regiospecificity and the inhibition alleviation conferred by the N216W substitution, led to a dramatic increase in maximum yield (up to 79%) and the R_T reached 90% for A^2XX, which is equivalent to a 6.9-fold increase compared to *Tx*Abf (Table SI 5). However, inevitably the generation of a high concentration of the transglycosylation product in the reaction mixture also led to strong secondary hydrolysis once the donor substrate had been consumed. Compared to R69H-L352M, this led to a lower Rho value, which was nevertheless 7-fold higher than that measured for *Tx*Abf (Table SI 6). In a previous study[21] it was revealed how a pH-dependent inhibition of *Tx*Abf could be used to block the degradation of transglycosylation products. Therefore, in the present study this strategy was successfully employed in combination with the mutant R69H-N216W-L352M. Accordingly, by shifting the pH of the reaction from 7.0 to 5.8, it was possible to obtain A^2XX in yields > 70%, without the inconvenience caused by secondary hydrolysis (Figure 5 and Table 2). It is noteworthy, that the change in pH did not affect the transfer rate, indicating that the H/T partition is not pH-dependent for the mutant R69H-N216W-L352M. Moreover, a control experiment performed in H_2O/D_2O (90:10, v/v) mixture revealed that the substitution of H_2O by D_2O did not affect the H/T partition since equivalent transfer rates and yields were obtained, at pH 7.0 and 5.8.

3.3. Probing the mechanisms of mutant enzymes

3.3.1. Hammett-Brønsted analysis: Transition state structure and rate-limiting step determination

To investigate how mutations in subsite -1 (i.e., L352M and R69H), affect the catalytic mechanism, notably in the switch from ground to the first transition state, a Hammett-Brønsted analysis was performed using a series of aryl L-arabinofuranosides to probe leaving group (LG) effects (Table SI 7). Plotting $\log(k_{cat}/K_M)$ as a function of the pK_a of the LG (Figure 6A) revealed that while the glycosylation step in reactions catalyzed by *Tx*Abf is relatively insensitive to LG pK_a changes (i.e., β_{LG1} = -0.35 or -0.28), the mutants R69H, L352M and R69H-L352M were 1.3 to 2.1-fold more sensitive (Table SI 8), indicating that these mutants are characterized by modifications in the transition state at the glycosylation step. Besides, since the catalytic constant (k_{cat} = $k_2.k_3/(k_2+k_3)$) reflects the rate-limiting step, plotting $\log(k_{cat})$ as a function of LG pK_a (Figure 6B) reveals whether the LG (released during glycosylation) is involved in this step or not. If not, the rate-limiting step is the deglycosylation one. Such an investigation demonstrated that most of enzyme, including *Tx*Abf, the $\log(k_{cat})$ was rather insensitive to LG pK_a changes, with values of β_{LG2} being close to zero for a wide range of pK_a values up to pK_a = 9.99. Thereby, for the wild-type as well as R69H and R69H-L352M mutants, the deglycosylation (k_3) is the rate-limiting step, being consistent with results obtained for other Abfs.[35,36] The minor exception to this was L352M, which displayed a slight dependence of k_{cat} on LG pK_a (β_{LG2} = -0.22). Taken together with the previous observation, this result most likely indicates this mutation strongly affects the glycosylation step, as suggested by the increase in the K_M value. It is noteworthy that when $\log(k_{cat}/K_M)$ was plotted as a function of LG pK_a evolution, the linearity of the different datasets was poor, mainly because of the two lowest LG pK_a points. In this respect, it is pertinent to note that the electron-withdrawing character of the LG might not be the only parameter influencing the glycosylation step. While diffusion-related phenomena can be discarded ($\sim 10^5$ $s^{-1}.mM^{-1}$ >> k_{cat}/K_M),[38] steric hindrances for instance can also impact the efficiency of this catalytic step. Albeit such profile was observed for other glycosidases,[38–40] there is no rational or objective means available to omit from calculations either of the data points corresponding to NCP or 3,4-dNP.

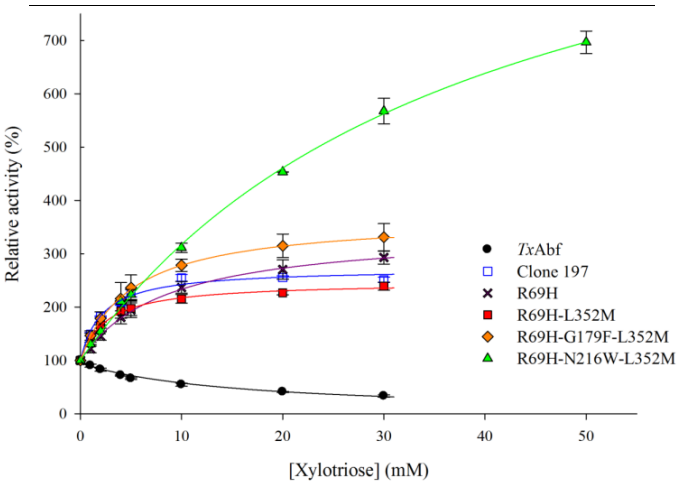


Fig. 3. Analysis of xylotriose acceptor impact on the global activity (*p*NP release) for *Tx*Abf, clone 197, R69H, R69H-L352M, R69H-G179F-L352M and R69H-N216W-L352M. For the latter one an additional xylotriose concentration point was assayed (50 mM) since the plateau-phase was not reached. Assays were carried out in duplicate at 45 °C in sodium phosphate buffer (50 mM, pH 7) with *p*NP-α-L-Ara*f* as donor substrate (5 mM). Apparent K_i and K_i' values for xylotriose and *Tx*Abf are 15.0 and 15.6 mM. For R69H, R69H-L352M, R69H-G179F-L352M and R69H-N216W-L352M mutants, apparent $K_M^{xylotriose}$ values for the acceptor are 8.6, 3.5, 5.9 and 45.5 mM.

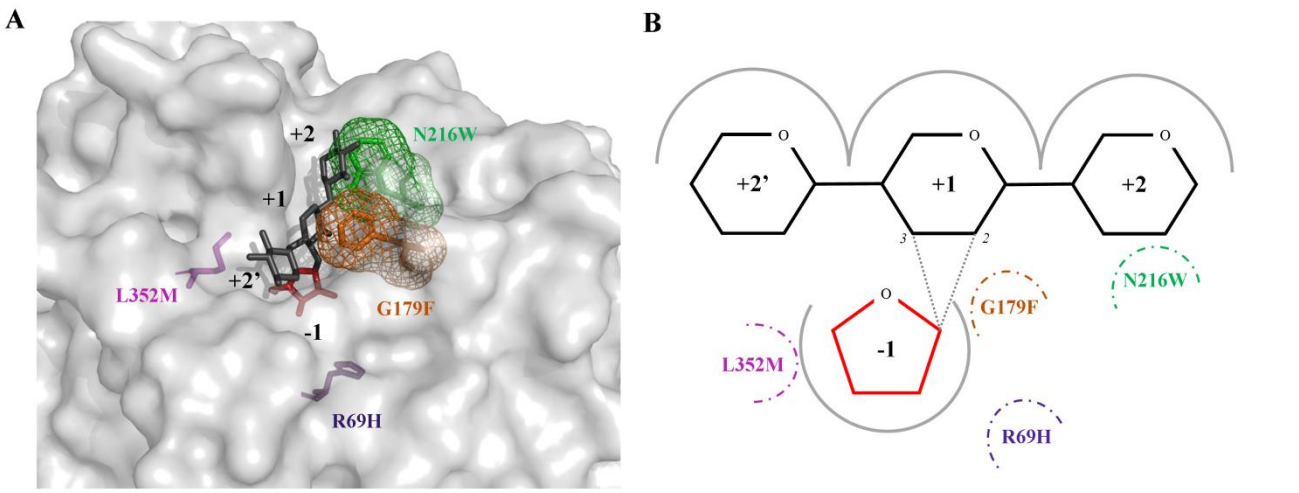


Fig. 4. Location of identified key mutations in -1 donor subsite (L352M and R69H) and +1 (G179F) or +2 (N216W) acceptor subsites. (A) 3D-view of co-crystallized *Tx*Abf-XA^{3}XX (PDB ID: 2VRQ)[37] X-ray structure in which mutations were introduced via PyMol software. One can observe a steric clash between N216W and +2 D-xylosyl unit, explaining (1→3) linkage deletion by N216W mutant and putative stacking interactions between G179F and +1 D-xylosyl unit. (B) 2D-schematic representation of *Tx*Abf active site.

3.3.2. Probing solvent kinetic isotope effects and the role of water

At the deglycosylation step, hydrolysis occurs when water acts as an acceptor, launching a nucleophilic attack on the anomeric carbon (Figures 1 and 2B). This catalytic sub-step is concomitant with the covalent bond breaking one, which leads to the decomposition of the β–L-Ara*f*-*Tx*Abf complex. Since it appears that *Tx*Abf and mutants thereof are deficient in the deglycosylation step, it was useful to investigate which of the two sub-steps was affected. For this, solvent kinetic isotope effects (sKIE) were measured using NCP-α-L-Ara*f* as the donor in reactions in which H_2O was replaced to different extents by D_2O (Fig. SI 9). In these experiments seeing as the LG reactivity of the donor, the apparent reaction rate reflects well the deglycosylation, including the water-mediated sub-step. As a result, no major changes in relative sKIE were observed, except for L352M, which was less sensitive to D_2O/H_2O exchange than *Tx*Abf. This indicates that in all cases, water nucleophilic attack *per se* is not a critical sub-step in this enzyme system. Nevertheless, an absolute sKIE effect was observed (80% residual activity), implying that H_2O to D_2O exchange does affect global catalytic efficiency, probably through modified protein dynamics or solvation,[41] as suggested by the exponential-like curve.[42]

3.3.3 Investigating the ionization state of catalytic residues

The pH-dependency of k_{cat}/K_M was investigated in order to access the apparent pK_a values of the two catalytic residues of the apo-enzyme (Fig. SI 10). The presence in mutants of amino acid substitutions R69H and L352M provided enzymes that were characterized by significant shifts (0.7 to 1.2 units) in the apparent pK_a value of the nucleophile E298, while the acid/base E176 ionization state remained globally unaffected (Table SI 9). Nevertheless, when the R69H was present as a single mutation, the basic limb of the pH profile displayed an unusual tail, which is characteristic of major ionization state modifications.[24]

4. DISCUSSION

Thus far, probably the most reproducible strategy to convert hydrolytic GHs into transglycosylases is the glycosynthase approach, which actually calls for the use of glycoside fluoride donors and catalytically-impotent (i.e., mutation of the catalytic nucleophile) versions of the parental enzymes.[43,44] Apart from this, no other strategies have come close to imitating the performance of natural transglycosylases, which are highly related to hydrolytic GH counterparts, but principally catalyze transglycosylation reactions. Unfortunately, attempts to unravel the results of natural evolutionary processes, which have furnished for example both transglycosylating XETs and hydrolytic XEHs, or trans-sialidase/sialidase pairs, has not yet provided the means to derive the generic knowledge that would be required to rationally convert any hydrolytic GH into a transglycosylase.[11,45,46] Nevertheless, some workers have successfully circumvented this conundrum, using artificial molecular evolution to convert GHs into transglycosylases,[33,47,48] while failing to fully elucidate the mechanisms that have driven this transition.

In the area of furanose chemistry, little attention has been focused on the development of furanose-acting transglycosylases.[14,49,50] Indeed, to our knowledge, the glycosynthase approach has not yet been successfully applied to such enzymes. For this reason and because furanose-acting transglycosylases would be a welcome addition to a subfield of glycochemistry that is widely recognized as being challenging, we attempted to engineer transfuranosylase activity into *Tx*Abf, a rather robust α-L-arabinofuranosidase from *Thermobacillus xylanyliticus* that already displays some ability to catalyze glycosynthetic reactions.

4.1. R69, a conserved residue in clan GH-A and a possible H/T modulator

Screening of the randomly-generated library, based on the L352M parent, was performed using a previously employed strategy that first seeks to identify clones that harbor hydrolytically-impaired mutants. Although rather simplistic, this approach has already proved to be efficient in the engineering of transglycosylase activity into a GH1 β-glycosidase.[33] In a second step, the activity of the selected mutants was measured in the presence of xylotriose, which can act as an acceptor, but also as a non-competitive inhibitor of *Tx*Abf, binding indifferently either to the *apo*-enzyme or to the covalent intermediate(Figure 3). This implies that the mutants that were selected at this stage were either no longer inhibited by xylotriose or that improved transglycosylation counterbalanced the inhibitory effects. Indeed, in the latter case, the rate enhancements observed in presence of xylotriose were underestimations of the actual transglycosylation improvements since the available active sites (i.e., not inhibited fraction) are less numerous than those in the acceptor free reaction. Overall, the two stage screening strategy pinpointed the mutation R69H. Interestingly, R69 is totally conserved among GH51 Abfs (Table SI 3) and, furthermore, is also one of only eight key residues that define clan GH-A, which is composed of 19 GH families.

Beyond sequence conservation, the spatial location of R69 is also significant, since its side chain points towards the nucleophile residue, and lies within H-bonding distance of the latter (Fig. SI 11). Consequently, it is plausible that the substitution of the equivalent arginine in any member of GH-A could lead to a similar modification of the H/T partition and thus a similar phenotype. In this regard, it is noteworthy that the recent engineering of the GH1 (a member of clan GH-A) *Thermus thermophilus* enzyme, *Tt*β-gly targeted R75, which is equivalent to R69in *Tx*Abf, and procured a similar transglycosylase phenotype.[46] In order to better understand how the mutation of R69 favours transglycosylation, it is important to recall the fact that the R69H mutation led to a drastic decrease in k_{cat} and, consequently, in the value of k_{cat}/K_M, both reflecting slower deglycosylation and glycosylation steps respectively. Moreover, the K_M value was also lower, suggesting that there was an accumulation of the glycosyl-enzyme intermediate. Taken together, these results point towards increases in the energy barriers of the transition states (TS) that are involved in each of the steps. One plausible explanation for these effects is the loss of a hydrogen bond between the guanidinium group (pK_a ~ 12.5) of R69 and the carboxyl oxygen of the catalytic nucleophile E298. This modification would alter the electron displacement system (Figure 2B) and, in particular affect the ionization state of catalytic residues as shown by the increase of E298 pK_a when R69 is mutated into histidine. It is known that along the double-displacement mechanism, pK_a cycling occurs,[51] with the nucleophile (E298) being negatively charged at the beginning of the reaction and the acid/base (E176) protonated. To achieve this in the case of *Tx*Abf, the acid/base catalyst (E176) displays an apparent pK_a ≈7.5, which is far higher than the pK_a value of a free glutamic acid (~ 4.5). This high pK_a is probably generated through charge-like repulsion with E176's close nucleophilic neighbor, E298, such interactions being necessary to avoid unfavorable Coulombic energetics.[51] Regarding the nucleophile catalyst, it has been demonstrated for a GH1 β-glycosidase from *Spodoptera frugiperda* (Sfβgly50) that the negative ionization state of its nucleophile catalyst, E399, is stabilized through hydrogen bond interactions with close neighbors, Y331 and R97, which are equivalent to Y242 and R69 in *Tx*Abf.[52]. Therefore, these observations are consistent with the idea that the mutation R69H alters the H/T partition through modifications of the nucleophile's ability to attack the anomeric carbon rather than a direct alteration of the efficiency of the catalytic acid/base. Interestingly, when Y242, which figures among the eight key residues of GH-A (Figure 2), is substituted by F, a phenotype (i.e., decreased H/T) similar to that of R69H was observed (data not shown). Overall, this is unsurprising, since Y242 also interacts with E298 and probably contributes to the definition of its operational charge.

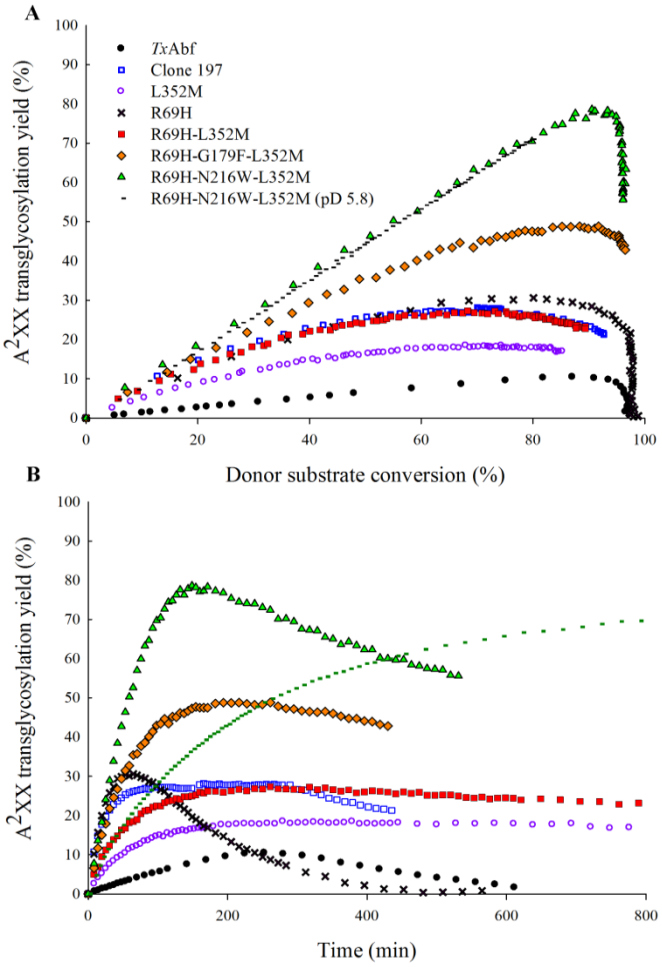


Fig. 5. A²XX major transglycosylation product evolution as a function of (A) donor conversion or (B) time for *Tx*Abf (pD 5.8), clone 197, L352M, R69H, R69H-L352M, R69H-G179F-L352M and R69H-N216W-L352M. All Assays were carried out at 45 °C in sodium phosphate buffer (25 mM, pD 7) or, when stated, in sodium acetate buffer (25 mM, pD 5.8), with *p*NP-α-L-Ara*f* as donor substrate (15 mM) and xylotriose as acceptor (30 mM), using 2000 nM of enzyme, except for *Tx*Abf and L352M or which 18 and 50 nM were employed, respectively.

Table 2. Transglycosylation yields[a] (%) for reactions catalyzed by *Tx*Abf and R69H derivatives.

Enzyme	XA^3X	A^3XX	(L-Ara*f*) + XA^2X	A^2XX	Auto 1[b]	Auto 2[b]
*Tx*Abf (pD 5.8)	7.5	11.8	3.3	10.5	6.7	3.2
R69H (pD 7)	23.2	25.2	4.9	30.6	3.9	2.2
R69H-L352M (pD 7)	20.4	22.2	4.6	27.3	4.7	2.2
R69H-G179F-L352M (pD 7)	35.3	17.3	8.4	48.8	2.1	1.6
R69H-N216W-L352M (pD 7)	4.6	7.7	8.0	78.6	3.0	7.7
R69H-N216W-L352M (pD 5.8)	0.7	1.2	5.5	71.0	2.3	8.1
R69H-N216W-L352M (H_2O-pH 7)	0.9	8.6	6.3	74.6	2.4	7.0
R69H-N216W-L352M (H_2O-pH 5.8)	1.4	1.4	5.4	71.9	2.4	8.1

[a] Maximum yields can be reached at different times explaining the fact that the sum of maximum yields can be superior to 100 %.
[b] Auto 1 and auto 2 correspond to two regioisomers of self-condensation products.[14]

It is now well recognized that an oxocarbenium ion is formed during TS of deglycosylation and that, at least in the case of retaining β-glycosidases, a partial positive charge is developed on the anomeric carbon,[53,54] the latter feature being important for catalysis since the incoming acceptor reacts with this species.[45,55] Therefore, the development of a charge of appropriate magnitude is critical and no doubt facilitates water-mediated deglycosylation. In this case, although water activation is a required featured, it is likely that there is no need for any energetic contribution from the water molecule itself. Indeed, this is consistent with the conclusions of recent *in silico* analyses performed on other GHs that indicate that the deglycosylation step is highly dissociative, meaning that the covalent bond that links the enzyme nucleophile to the anomeric carbon of the sugar moiety is almost broken when it is replaced by a bond with water.[56] In the light of these mechanistic details, it appears probable that the mutation R69H provokes a charge modification that impacts on TS, reducing the efficiency of water-mediated deglycosylation.

The Hammett-Brønsted analysis revealed that *Tx*Abf and most of the mutants are characterized by a rate-limiting deglycosylation over a large range of substrates, an observation that is consistent with data collected for other Abfs,[35,36] but different to those collected for glycopyranosidases.[38–40,57,58] For the latter, the Brønsted plot of $\log(k_{cat})$ is usually characterized by a concave breakdown for LGs whose pK_a values are above 8, this transition being indicative of a change in the rate-limiting step (from deglycosylation to glycosylation). Accordingly, for *Tx*Abf, and more generally FHs, it is possible that this difference is linked to the ring strain of furanoside substrates,[59] which makes the decomposition of the glycosyl-enzyme complex (i.e., deglycosylation) more energy demanding than its creation (i.e., glycosylation).

Taking the observation that deglycosylation remains the rate-limiting step in *Tx*Abf and mutants thereof, irrespective of the LG, together with the strong decrease in k_{cat} values for R69H mutants and the absence of solvent KIE changes suggests that the covalent bond breaking sub-step associated with deglycosylation step is highly impaired in the R69H mutants. However, if one also considers the rather low dependency of k_{cat}/K_M on LG pK_a values, it is also possible to deduce that only a small negative charge is developed on the interglycosidic oxygen at TS_1, which implies that either acid/base protonation is strong enough to compensate for the electronic character of the LG,[60,61] or that during glycosylation the major contribution comes from the catalytic nucleophile, which provides a strong 'nucleophilic push' towards the anomeric carbon. In this respect, the more negatives β_{LG_1} values measured for the mutants containing R69H and/or L352M provide further evidence for modifications of charge distribution within the TS, along the axis [nucleophile E298 ↔ anomeric carbon ↔ interglycosidic oxygen].

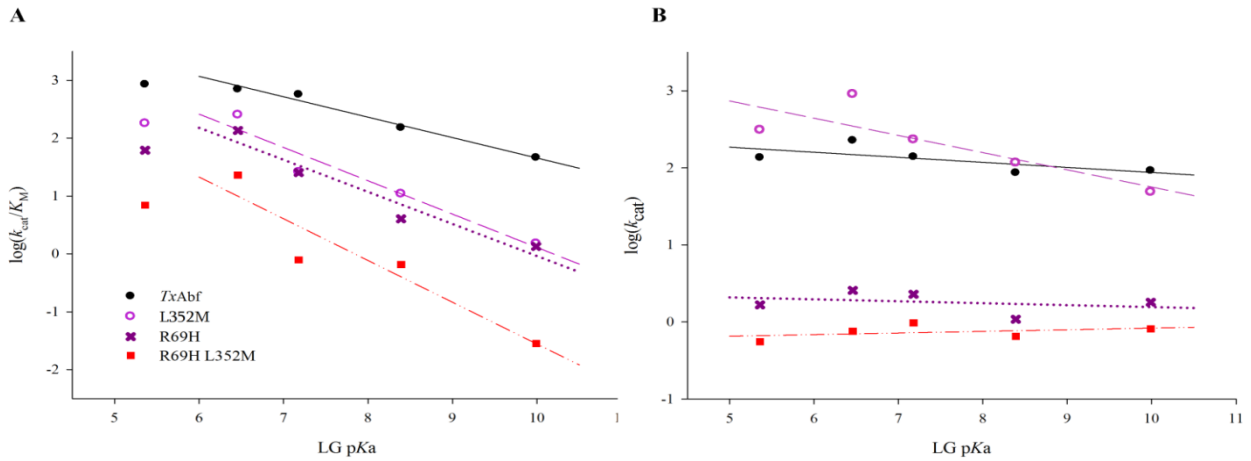


Fig. 6. Hammett relationship, represented as a Brønsted plot, for the hydrolysis of aryl α-L-arabinofuranosides catalyzed by *Tx*Abf (solid), L352M (dash), R69H (dot) and R69H-L352M (dash-dot-dot). (A) $\log(k_{cat}/K_M)$ and (B) $\log(k_{cat})$ are plotted as a function of the pK_a of the LG for a series of LG-α-L-Ara*f*.

Overall, all of the kinetics and sKIE data presented in this study point to the fact that the mutations R69H and L352M, disrupt both covalent glycosyl-enzyme bond formation and decomposition, and that this is caused by changes in the distribution of electronic charge.

4.2. Key factors that transform GHs into transglycosylases

4.2.1. Donor subsite mutations undermine enzyme activity, but are necessary to counter water activity

The fact that donor subsite mutations might be a prerequisite for the successful production of a transglycosylase appears to be rather counterintuitive, especially because data presented here clearly show that such mutations severely reduce enzyme activity. Nevertheless, the mutation of donor subsite determinants has recently been proposed as a semi-rational strategy to improve the transglycosylation ability of GHs,[46] despite the fact that it might appear more intuitive to suppose that mutations in the acceptor subsites could promote transglycosylation, by increasing the recognition of the incoming sugar moiety. However, bearing in mind the fact that GHs generally function in aqueous environments characterized by 55 M water concentration, it is clearly imperative for transglycosylases to somehow avoid water-mediated deglycosylation, reflected for instance in increased lifetime of the glycosyl-enzyme as observed for XET.[62]

In this study, we have shown that mutation of R69H, which had drastic effect on activity, had a very positive impact on the H/T ratio in favor of transglycosylation, by reducing water-mediated deglycosylation and thus hydrolysis (Fig. SI 12). However, even this mutation did not eliminate secondary hydrolysis, a reaction that procures products that can no longer act as substrates for further transglycosylation (Figure 1). Indeed, it was necessary to combine R69H with a second subsite -1 mutation, L352M, in order to sufficiently diminish secondary hydrolysis and thus preserve the transglycosylation products from decomposition. Clearly, the introduction of L352M affected the glycosylation step, a fact revealed by the marked sensitivity of L352M-containing mutants to LG pK_a variations (β_{LG}) and the very high K_M values measured on LG-α-L-Ara*f*. In this respect, it is noteworthy that, compared to *p*NP (pK_a = 7.18), the xylotriosyl moiety of the transglycosylation products displays much lower leaving group ability (pK_a glycosidic hydroxyl ~ 12), which might explain the reduced ability of L352M-containing mutants to hydrolyze the transglycosylation products.

4.2.2. Different effects result from acceptor subsite modifications

In this work, two acceptor subsite mutations contributed to the successful creation of a transarabinofuranosidase. The first one G179F, located in subsite +1, was selected using an *in silico* approach with the algorithm (*BindScan*) that identifies positions sensitive to a binding affinity gain for a given acceptor and putative positive mutations. Whereas the introduction of G179F into R69H-L352M double mutant did not impact the transfer rate, it clearly affects secondary hydrolysis, although the mechanism has yet to be clarified. Nevertheless, it is highly probable that the mutation G179F procures increased hydrophobicity in subsite +1 and a stacking interaction with *p*NP, thus specifically increasing the affinity of the mutant enzyme for *p*NP-α-L-Ara*f*. The implication of this modification is that *p*NP-α-L-Ara*f* is probably better recognized compared to the transglycosylation products and thus secondary hydrolysis is postponed.

Regarding N216W, when present as a single mutation this subsite +2 mutation only modestly increases the transfer rate and diverts activity away from the formation of (1→3) linkages while alleviating xylotriose inhibition. Therefore, it is probable that N216W affects both acceptor positioning and affinity. Additionally, as illustrated by its non-saturating acceptor profile (Figure 3), R69H-N216W-L352M triple mutant may reveal the acceptor-activation behavior that R69H-L352M should display without acceptor inhibition which restrains the apparent acceptor-mediated activity enhancement.

Since overall yields remained low for single mutants G179F and N216W, altering acceptor subsites appears insufficient to create good transglycosylase activity. However, when the effects of these mutations were combined with the crippling effect of R69H-L352M, the benefits of these mutations were fully visible, since almost perfect transfuranosylases were obtained (Figure 8).

4.2.3. Towards an understanding of evolutionary processes

Numerous structural studies, including the one performed on *Tx*Abf, have revealed that *exo*-acting GHs are characterized by extensive enzyme-substrate interaction networks in the donor subsite(s). Consequently, reactions catalyzed by *exo*-acting GH are characterized by a strong stabilization of the transition state energy barriers, which usually results in quite high catalytic efficiencies (k_{cat}/K_M values), linked to a fast glycosylation step, and reasonably efficient (albeit slower) water-mediated deglycosylation. The efficiency of the latter step is underpinned by the fact that the vast majority of the energy required to overcome the TS is procured by the donor subsite interactions (Fig. SI 12). This last point is vital to understand the domination of water-mediated deglycosylation in hydrolytic GHs operating in aqueous environments in which the concentration of water (55 M) always far outweighs that of other acceptor molecules. Nevertheless, despite such adverse conditions, *exo*-acting GHs have often been shown to perform some degree of transglycosylation (up to 10% yields for *Tx*Abf and sometimes more for other *exo*-GHs),[63] which is consistent with the fact that acceptor sugars can lower the TS2 energy level at deglycosylation step (Fig. SI 12 and Fig. SI 13). Therefore, acceptor subsite mutations that improve enzyme-(acceptor) substrate interactions can further exacerbate

this thermodynamic advantage and thus increase transglycosylation yields. When coupled to other mutations that deprive water-mediated deglycosylation of much of the energy that is required to overcome the TS (i.e., ones that perturb donor subsite interactions) potent transglycosylases are produced. This combinatorial strategy is possibly a general route that has been taken by Nature in order to avoid overwhelming competition from water molecules, and might partly explain why GHs that display strong transglycosylation are generally characterized by quite low overall activity. [64–66]

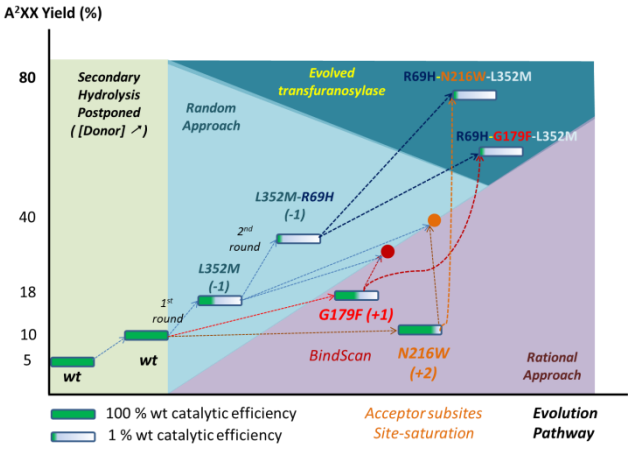


Figure 8. Evolutionary pathway followed by *Tx*Abf allowing its conversion in a non-Leloir transarabinofuranosylase. A subtle combination of random, semi-rational and *in silico* approaches was elaborated to reach this goal.

5. CONCLUSION AND PERSPECTIVES

To our knowledge, this work constitutes the very first description of the creation a non-Leloir transfuranosylase (80% yield, slight residual hydrolysis). In this respect, the result is significant, since it opens up a wide variety of options for the synthesis of hitherto difficult to obtain furanose-based sugars, including those containing D-galactofuranose, a sugar that is widespread in pathogenic microorganisms. Regarding the combinatorial method that was employed to successfully engineer transglycosylase activity, it is noteworthy that the random mutagenesis coupled to a quite simple screening method proved to be quite robust and allowed the selection of R69H. The detailed analysis of this mutation has provided strong evidence to explain how this mutation affects activity and, importantly, analysis of clan GH-A has revealed that this amino acid may well possess a functional homolog in every clan member, implying that the same strategy might produce highly similar results in other GHs. Similarly, concerning the other mutations that were identified using more rational approaches, it may well be possible to reproduce their beneficial effects in other GHs, by carefully targeting certain amino acid determinants of GH acceptor subsites. Overall, we believe that the lessons learned in this work provide the basis of a general GH engineering route that should allow the successful improvement of transglycosylase activity in other *exo*-acting GHs.

ASSOCIATED CONTENT

Supplementary tables and figures are available at the end of this manuscript: SI of part IV

AUTHOR INFORMATION

Corresponding Author

* michael.odonohue@insa-toulouse.fr

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

PhD fellowship of B. Bissaro was supported by the Institut National de la Recherche Agronomique (INRA) [CJS]. The NMR work carried out in this work at the Laboratory for BioSystems & Process Engineering (Toulouse, France) was performed with the equipments of MetaToul, the Toulouse Center for Metabolomics, which is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, the SICOVAL, the Centre National de la Recherche Scientifique (CNRS) and the Institut National de la Recherche Agronomique (INRA). We thank the ICEO facility dedicated to enzyme screening and discovery, and part of the Integrated Screening Platform of Toulouse (PICT, IBiSA) for providing access to the TECAN liquid handling equipment.

ABBREVIATIONS

Abfs, α-L-arabinofuranosidases; AXOS, arabinoxylo-oligosaccharides; FH, furanoside hydrolase; GH, glycoside hydrolase; LG, leaving group; *p*NP-α-L-Ara*f*, 4-nitrophenyl α-L-arabinofuranoside; sKIE, solvent Kinetic Isotope Effects; *Tx*Abf, α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*; H/T, hydrolysis/transglycosylation ratio; TS, transition state; X, donor substrate conversion rate; X-Ara*f*, 5-bromo-4-chloro-3-indolyl α-L-arabinofuranoside; XOS, xylo-oligosaccharides; D-Xyl*p*, D-xylopyranosyl; Y, yield.

REFERENCES

(1) Lowary, T. *Curr. Opin. Chem. Biol.* **2003**, *7*, 749–756.
(2) Peltier, P.; Euzen, R.; Daniellou, R.; Nugier-Chauvin, C.; Ferrières, V. *Carbohydr. Res.* **2008**, *343*, 1897–1923.
(3) Wang, L.-X.; Huang, W. *Curr. Opin. Chem. Biol.* **2009**, *13*, 592–600.
(4) Champion, E.; André, I.; Moulis, C.; Boutet, J.; Descroix, K.; Morel, S.; Monsan, P.; Mulard, L. a; Remaud-Siméon, M. *J. Am. Chem. Soc.* **2009**, *131*, 7379–7389.
(5) Fauré, R.; Saura-Valls, M.; Brumer, H.; Planas, A.; Cottaz, S.; Driguez, H. *J. Org. Chem.* **2006**, *71*, 5151–5161.

(6) Koshland Jr., D. E. *Biol. Rev. Cambridge Philos. Soc.* **1953**, *28*, 416–436.
(7) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Res.* **2014**, *42*, D490–D495.
(8) Street, I. P.; Kempton, J. B.; Withers, S. G. *Biochemistry* **1992**, *31*, 9970–9978.
(9) Ducros, M.-A.; Tarling, C. A.; Zechel, D. L.; Brzozowski, A. M.; Frandsen, T. P.; Ossowski, I. Von; Withers, S. G.; Davies, G. J. *Chem. Biol.* **2003**, *10*, 619–628.
(10) Zakariassen, H.; Hansen, M. C.; Jøranli, M.; Eijsink, V. G. H.; Sørlie, M. *Biochemistry* **2011**, *50*, 5693–5703.
(11) Eklöf, J. M.; Shojania, S.; Okon, M.; McIntosh, L. P.; Brumer, H. *J. Biol. Chem.* **2013**, *288*, 15786–15799.
(12) Saburi, W.; Kobayashi, M.; Mori, H.; Okuyama, M.; Kimura, A. *J. Biol. Chem.* **2013**, *288*, 31670–31677.
(13) Debeche, T.; Cummings, N.; Connerton, I.; Debeire, P.; O'Donohue, M. J. *Appl. Environ. Microbiol.* **2000**, *66*, 1734–1736.
(14) Rémond, C.; Plantier-Royon, R.; Aubry, N.; O'Donohue, M. J. *Carbohydr. Res.* **2005**, *340*, 637–644.
(15) Rémond, C.; Plantier-Royon, R.; Aubry, N.; Maes, E.; Bliard, C.; O'Donohue, M. J. *Carbohydr. Res.* **2004**, *339*, 2019–2025.
(16) Arab-Jaziri, F.; Bissaro, B.; Dion, M.; Saurel, O.; Harrison, D.; Ferreira, F.; Milon, A.; Tellier, Charles, O'Donohue, Michael J, Fauré, R. *New Biotechnol.* **2013**, *30*, 536–544.
(17) Arab-Jaziri, F.; Bissaro, B.; Tellier, C.; Dion, M.; Fauré, R.; O'Donohue, M. J. *Submitt. Results* **2014**.
(18) Kelly, M. A.; Sinnott, M. A.; Widdows, D. *Carbohydr. Res.* **1988**, *181*, 262–266.
(19) Jacobsson, M.; Malmberg, J.; Ellervik, U. *Carbohydr. Res.* **2006**, *341*, 1266–1281.
(20) Du, Y.; Pan, Q.; Kong, F. *Carbohydr. Res.* **2000**, *323*, 28–35.
(21) Bissaro, B.; Saurel, O.; Arab-Jaziri, F.; Saulnier, L.; Milon, A.; Tenkanen, M.; Monsan, P.; O'Donohue, M. J.; Fauré, R. *Biochim. Biophys. Acta Gen. Subj.* **2014**, *1840*, 626–636.
(22) Patrick, W. M.; Firth, A. E.; Blackburn, J. M. *Protein Eng. Des. Sel.* **2003**, *16*, 451–457.
(23) Joshi, M. D.; Sidhu, G.; Nielsen, J. E.; Brayer, G. D.; Withers, S. G.; McIntosh, L. P. *Biochemistry* **2001**, *40*, 10115–10139.
(24) Bicknell, R.; Knott-Hunziker, V.; Waley, S. G. *Biochem. J.* **1983**, *213*, 61–66.
(25) Glasoe, P. K.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188–190.
(26) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *3263*, 7512–7515.
(27) Gruppen, H.; Hoffmann, R. A.; Kormelink, F. J.; Voragen, A. G.; Kamerling, J. P.; Vliegenthart, J. F. *Carbohydr. Res.* **1992**, *233*, 45–64.
(28) Vietor, R. J.; Hoffmann, R. A.; Angelino, S. A. G. .; Voragen, A. G. J.; Kamerling, J. P.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1994**, *6215*, 245–255.
(29) Ferré, H.; Broberg, A.; Duus, J. O.; Thomsen, K. K. *Eur. J. Biochem.* **2000**, *267*, 6633–6641.
(30) Pastell, H.; Tuomainen, P.; Virkki, L.; Tenkanen, M. *Carbohydr. Res.* **2008**, *343*, 3049–3057.
(31) Pitkänen, L.; Tuomainen, P.; Virkki, L.; Tenkanen, M. *Int. J. Biol. Macromol.* **2011**, *49*, 963–969.
(32) Correia, M. A. S.; Mazumder, K.; Brás, J. L. A.; Firbank, S. J.; Zhu, Y.; Lewis, R. J.; York, W. S.; Fontes, C. M. G. A.; Gilbert, H. J. *J. Biol. Chem.* **2011**, *286*, 22510–22520.
(33) Feng, H.-Y.; Drone, J.; Hoffmann, L.; Tran, V.; Tellier, C.; Rabiller, C.; Dion, M. *J. Biol. Chem.* **2005**, *280*, 37088 –37097.
(34) Fauré, R.; Courtin, C. M.; Delcour, J. A.; Dumon, C.; Faulds, C. B.; Fincher, G. B.; Fort, S.; Fry, S. C.; Halila, S.; Kabel, M. A.; Pouvreau, L.; Quemener, B.; Rivet, A.; Saulnier, L.; Schols, H. A.; Driguez, H.; O'Donohue, M. J. *Aust. J. Chem.* **2009**, *62*, 533–537.
(35) Shallom, D.; Belakhov, V.; Solomon, D.; Shoham, G.; Baasov, T.; Shoham, Y. *J. Biol. Chem.* **2002**, *277*, 43667–43673.
(36) Wan, C.; Chen, W.; Chen, C.; Chang, M.; Lo, L.; Li, Y. *Biochem. J.* **2007**, *401*, 551–558.
(37) Paës, G.; Skov, L. K.; O'Donohue, M. J.; Rémond, C.; Kastrup, J. S.; Gajhede, M.; Mirza, O. *Biochemistry* **2008**, *47*, 7441–7451.
(38) Zechel, D. L.; Reid, S. P.; Stoll, D.; Nashiru, O.; Warren, R. A. J.; Withers, S. G. *Biochemistry* **2003**, *42*, 7195–7204.
(39) Bravman, T.; Zolotnitsky, G.; Belakhov, V.; Shoham, G.; Henrissat, B.; Baasov, T.; Shoham, Y. *Biochemistry* **2003**, *42*, 10528–10536.
(40) Kempton, J. B.; Withers, S. G. *Biochemistry* **1992**, *31*, 9961–9969.
(41) Combes, D.; Yoovidhya, T.; Girbal, E.; Willemot, R.-M.; Monsan, P. *Ann. NY. Acad. Sci.* **1987**, *501*, 59–62.
(42) Schowen, K. B.; Limbach, H. H.; Denisov, G. S.; Schowen, R. L. *Biochim. Biophys. Acta* **2000**, *1458*, 43–62.
(43) Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120*, 5583–5584.
(44) Malet, C.; Planas, A. *FEBS Lett.* **1998**, *440*, 208–212.
(45) Pierdominici-Sottile, G.; Palma, J.; Roitberg, A. E. *Proteins* **2013**, *82*, 424–435.
(46) Teze, D.; Hendrickx, J.; Czjzek, M.; Ropartz, D.; Sanejouand, Y.-H.; Tran, V.; Tellier, C.; Dion, M. *Protein Eng. Des. Sel.* **2014**, *27*, 13–19.
(47) Osanjo, G.; Dion, M.; Drone, J.; Solleux, C.; Tran, V.; Rabiller, C.; Tellier, C. *Biochemistry* **2007**, *46*, 1022–1033.
(48) Placier, G.; Watzlawick, H.; Rabiller, C.; Mattes, R. *Appl. Environ. Microbiol.* **2009**, *75*, 6312–6321.
(49) Almendros, M.; Danalev, D.; François-Heude, M.; Loyer, P.; Legentil, L.; Nugier-Chauvin, C.; Daniellou, R.; Ferrières, V. *Org. Biomol. Chem.* **2011**, *9*, 8371–8378.
(50) Chlubnova, I.; Filipp, D.; Spiwok, V.; Dvorakova, H.; Daniellou, R.; Nugier-Chauvin, C.; Kralova, B.; Ferrieres, V. *Org. Biomol. Chem.* **2010**, *8*, 2092–2102.
(51) Mcintosh, L. P.; Hand, G.; Johnson, P. E.; Joshi, M. D.; Ko, M.; Plesniak, L. A.; Ziser, L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1996**, *35*, 9958–9966.
(52) Marana, S. R.; Mendonça, L. M. F.; Andrade, E. H. P.; Terra, W. R.; Ferreira, C. *Eur. J. Biochem.* **2003**,

270, 4866–4875.
(53) Notenboom, V.; Birsan, C.; Nitz, M.; Rose, D. R.; Warren, R. a; Withers, S. G. *Nat. Struct. Biol.* **1998**, *5*, 812–818.
(54) Vocadlo, D. J.; Davies, G. J. *Curr. Opin. Chem. Biol.* **2008**, *12*, 539–555.
(55) Pierdominici-Sottile, G.; Horenstein, N. A.; Roitberg, A. E. *Biochemistry* **2011**, *50*, 10150–10158.
(56) Brás, N. F.; Ramos, M. J.; Fernandes, P. a. *J. Mol. Struct. THEOCHEM* **2010**, *946*, 125–133.
(57) Tull, D.; Withers, S. G. *Biochemistry* **1994**, *33*, 6363–6370.
(58) Vallmitjana, M.; Ferrer-Navarro, M.; Planell, R.; Abel, M.; Ausín, C.; Querol, E.; Planas, A.; Pérez-Pons, J. a. *Biochemistry* **2001**, *40*, 5975–5982.
(59) Taha, H. a; Richards, M. R.; Lowary, T. L. *Chem. Rev.* **2013**, *113*, 1851–1876.
(60) Ly, H. D.; Withers, S. G. *Annu. Rev. Biochem.* **1999**, *68*, 487–522.
(61) Kelly, M. a; Sinnott, M. L.; Herrchen, M. *Biochem. J.* **1987**, *245*, 843–849.
(62) Piens, K.; Fauré, R.; Sundqvist, G.; Baumann, M. J.; Saura-Valls, M.; Teeri, T. T.; Cottaz, S.; Planas, A.; Driguez, H.; Brumer, H. *J. Biol. Chem.* **2008**, *283*, 21864–21872.
(63) Dion, M.; Nisole, A.; Spangenberg, P.; André, C.; Glottin-Fleury, A.; Mattes, R.; Tellier, C.; Rabiller, C. *Glycoconjugate J.* **2001**, *18*, 215–223.
(64) Potocki de Montalk, G.; Remaud-Simeon, M.; Willemot, R. M.; Sarçabal, P.; Planchot, V.; Monsan, P. *FEBS Lett.* **2000**, *471*, 219–223.
(65) Baumann, M. J.; Eklöf, J. M.; Michel, G.; Kallas, A. M.; Teeri, T. T.; Czjzek, M.; Brumer, H. *Plant Cell* **2007**, *19*, 1947–1963.
(66) Ribeirão, M.; Pereira-Chioccola, V. L.; Eichinger, D.; Rodrigues, M. M.; Schenkman, S. *Glycobiology* **1997**, *7*, 1237–1246.

Supporting Information of Part IV: "Molecular Design of the first 'non-Leloir' transarabinofuranosylases".

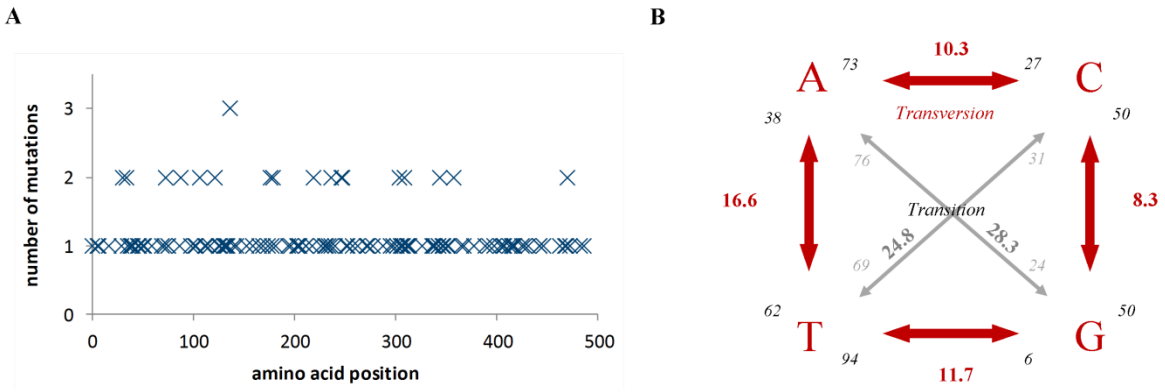


Fig. SI 1. L352M-based random library sequence analysis. (A) Mutations repartition along *Tx*Abf sequence and (B) transversion and transition rate analysis. (A) Among the 24 sequences (equivalent to 35,856 bases), a total of 145 mutations were found accounting for 30 silent mutations and 115 missense (including 4 STOP mutations). The mean mutation rate is of 3.2 amino acid mutations/kb (4.8 mutations for the whole gene). (B) Besides, a finer analysis of transition (grey) and transversion (red) rates revealed acceptable percentages, 53 and 47% respectively, in agreement with the ones expected for the chosen mutagenesis approach.(Rasila et al., 2009)[8]. Are also given for each possible transition or transversion between two bases the contribution of both (x→y) and (y→x) mutations: for instance, within the (A↔ T) transversion (which accounts for 16.6% of total transversion and transition events) 38% are T→A events whereas 62% are A→T.

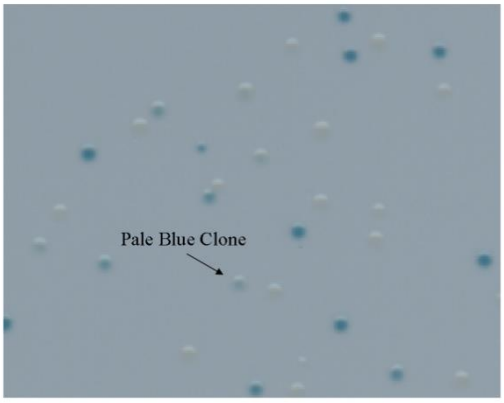


Fig. SI 2. Example of Q-tray for *in vivo* primary screening. Three main different behaviors can be observed, *(i)* white clones (inactive), *(ii)* pale blue (hydrolytically crippled) and *(iii)* dark blue (equivalent to L352M) clones.

[8] Rasilla, T.S., Pajunen, M.I., Savilahti, H. (2009). Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Echerichia coli mutator strain, and hydroxylamine treatment. *Anal. Biochem.*, 388, pp.71-80.

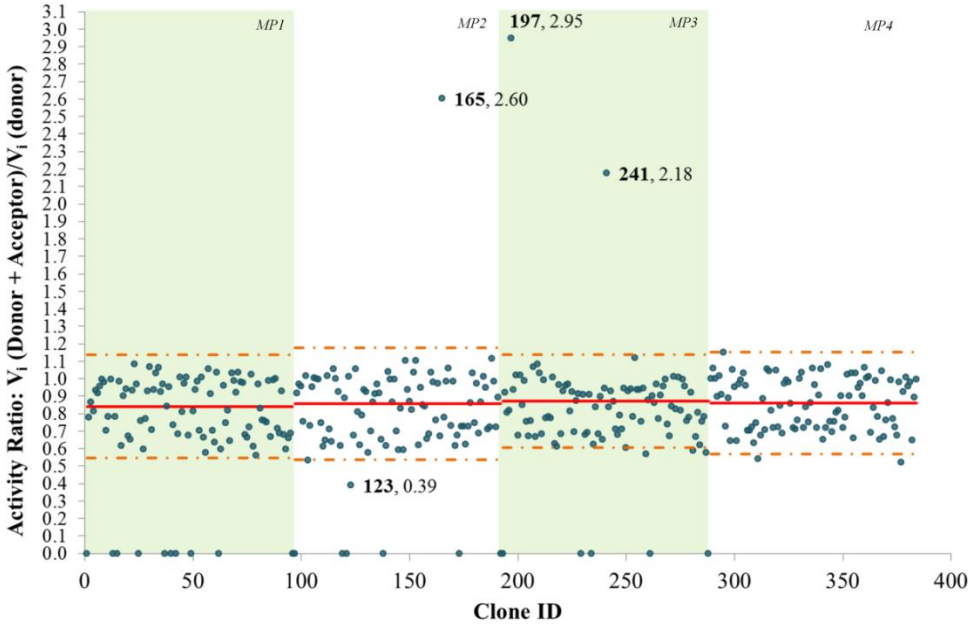


Fig. SI 3. Secondary screening of L352M-based random library using activity ratio measurements. Ratios were calculated from donor *p*NP-α-L-Ara*f* (15 mM) consumption rates in presence (10 mM) or absence of the acceptor xylotriose. Assays were conducted at 45 °C in sodium phosphate buffer (pH 7, 50 mM) with automated kinetics on TECAN robot. Each micro-plate (MP) activity ratio mean value is represented (red solid line) as well as 95% confidence interval (1.96 standard deviation in orange dash-dot line). Mean value (and standard deviation) are equal to 0.84 (0.15), 0.86 (0.16), 0.87 (0.14) and 0.86 (0.15) for MP1, MP2, MP3 and MP4, respectively. This screening revealed the presence of 3 positive and 1 negative clones, which appeared outside of the 95% confidence interval (μ ± 1.96 σ) and that are represented by their ID numbers followed by their associated activity ratio. Namely, positive clones are clones 165 (μ + 10.7 σ), 197 (μ + 15.3 σ) and 241 (μ + 9.6 σ) and the negative one is clone 123 (μ -2.9 σ).

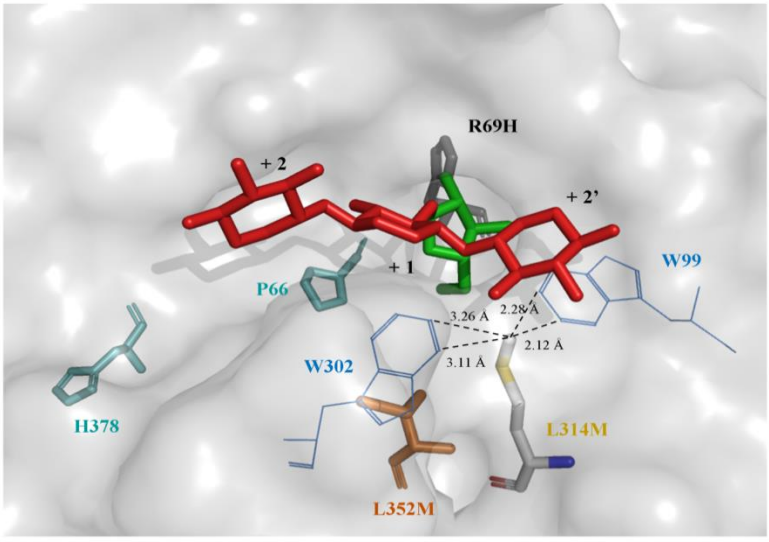


Fig. SI 4. Location of residues from the 'negative' clone 123. The *Tx*Abf-mutant encoding sequence of this clone revealed mutations H378Y, P66Q and L314M in addition to parental L352M. P66 and L314 are relatively conserved (~ 80% for both among GH51 Abfs) and Q66 or M314 substitutions never occur. On the contrary, H378 occurrence only reaches 47% and H378Y substitution is found at a level of 27%. The location far from the active site and its non-conserved nature allow us to discard this mutation as the putative responsible for low activity ratio (0.39). On the other hand, P66 is located on the same strand as R69 and the proline mutation may directly impact the active site interaction network. Similarly, L314 is close to W302 and W99, which were shown to be essential residues for transglycosylation to properly occur (unpublished results).

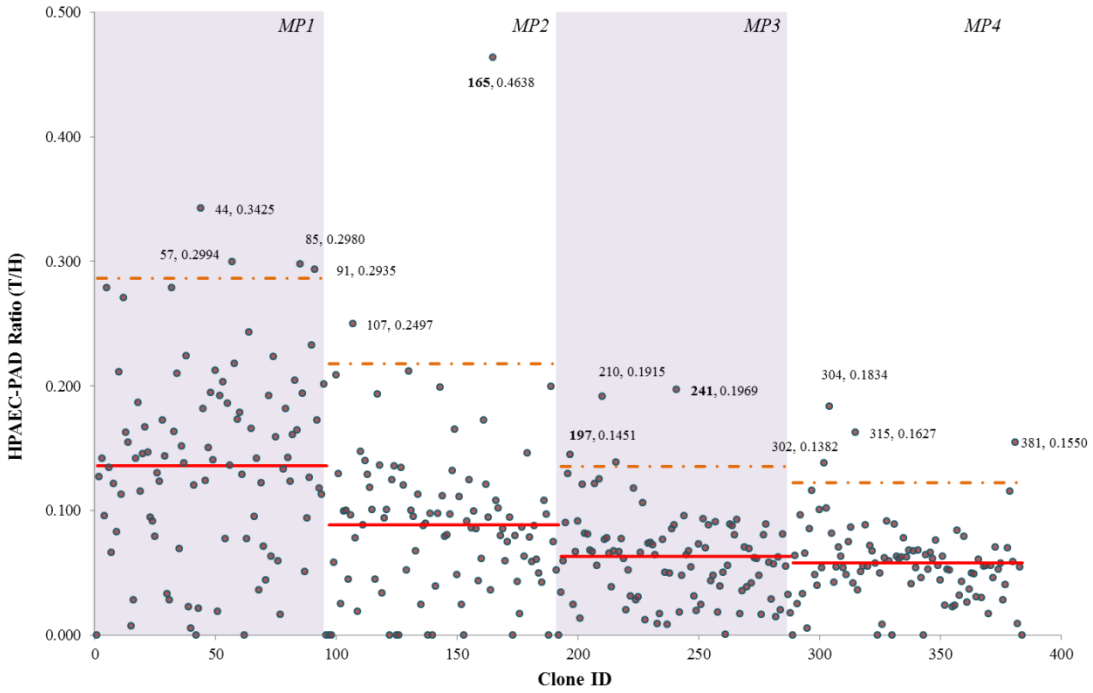


Fig. SI 5. HPAEC-PAD secondary screening of L352M-based random library. The ratio H/T was calculated as the sum of transglycosylation products signals area upon the height of arabinose signal.

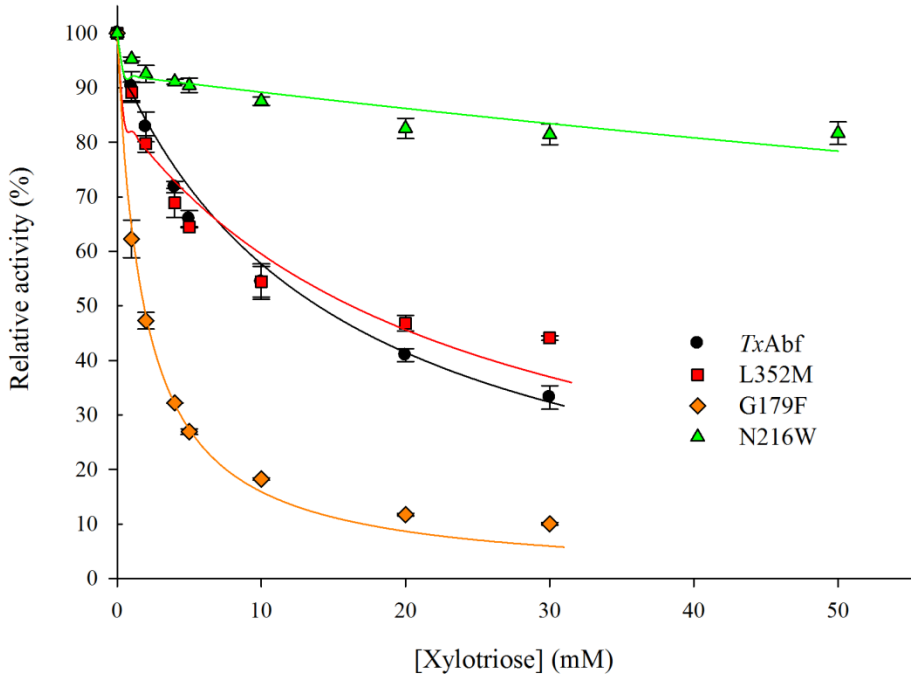


Fig. SI 6. Analysis of xylotriose acceptor impact on the global activity (*p*NP release) for *Tx*Abf and single mutants L352M, G179F and N216W. For the latter one an additional xylotriose concentration point was assayed (50 mM) as for R69H-N216W-L352M (Figure 3). Assays were carried out in duplicate at 45 °C in sodium phosphate buffer (50 mM, pH 7) with *p*NP-α-L-Ara*f* as donor substrate (5 mM). One can observe the strongest recognition of xylotriose by the rationally found (BindScan) mutant G179F (+1 subsite) and the alleviation of inhibition provoked by N216W (+2 subsite).

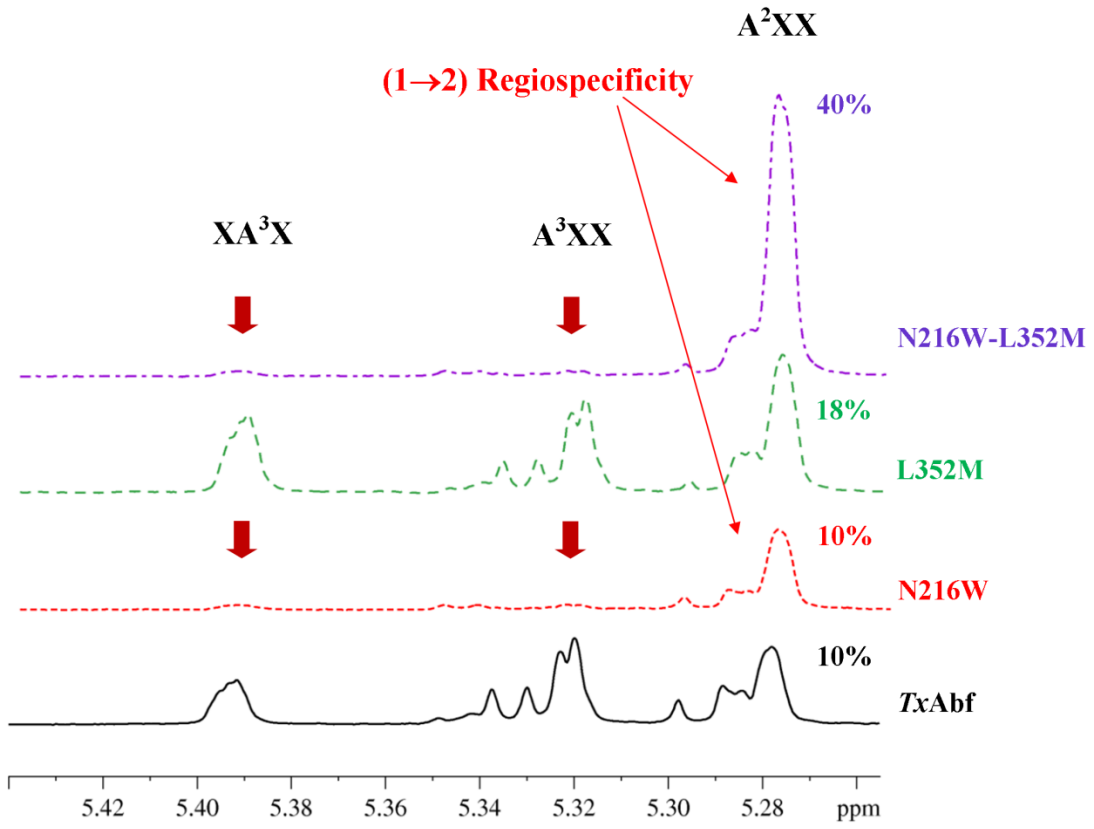


Fig. SI 7. Introduction of (1→2) regiospecific linkage formation phenotype into *Tx*Abf or L352M scaffold through N216W mutation. Is represented the anomeric region of 1H NMR spectrum of transglycosylation products obtained during reactions catalyzed by *Tx*Abf (black solid), N216W (red small dash), L352M (green medium dash) and L352M-N216W (purple, dash-dot) with *p*NP-α-L-Ara*f* (15 mM) as donor and xylotriose (30 mM) as acceptor.

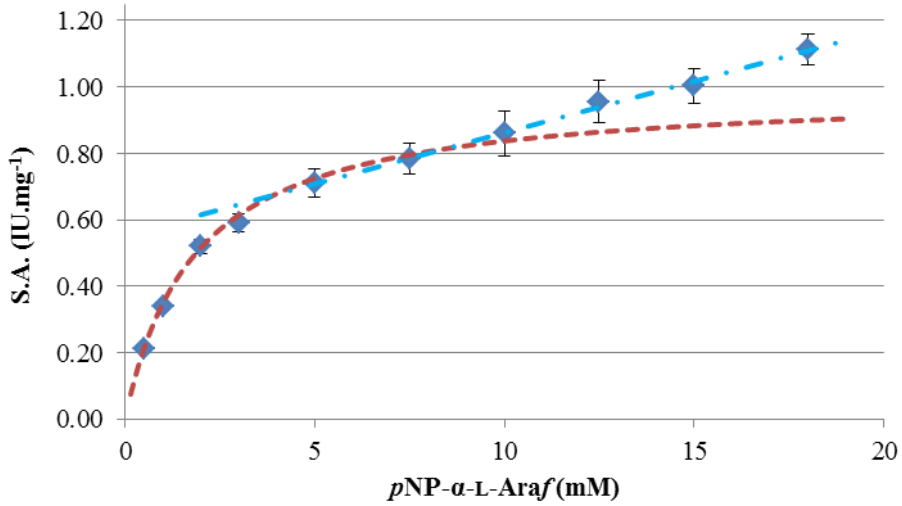


Fig. SI 8. Substrate-dependent activity profile of R69H-L352M. Two phases can be observed, for [*p*NP-α-L-Ara*f*] < 5 mM a Michaelis-Menten model (red dash) can be fitted to experimental points whereas for [*p*NP-α-L-Ara*f*] > 5 mM a linear relationship (blue dash-dot) demonstrates the effect of *p*NP-α-L-Ara*f* as an 'activator' (autocondensation). A mixed model SA = SA_{max}.[S]/(K_M + S) + Ns.[S] allows to represent such a phenomenon.

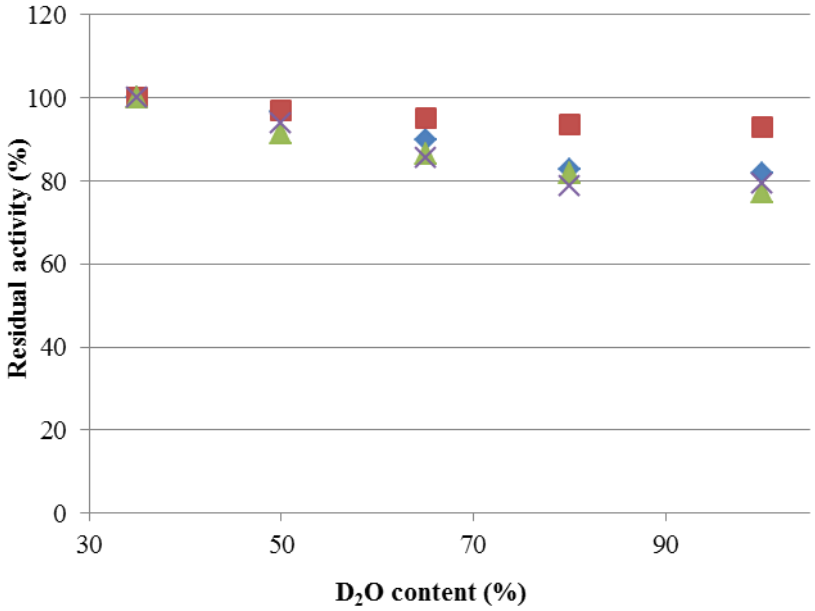


Fig. SI 9. Solvent kinetic isotope effect on specific activity of *Tx*Abf (◆), L352M (■), R69H(▲) and R69H-L352M (×). The assay were realized in triplicates at 45 °C and pH 7, with NCP-α-L-Ara*f* as donor substrate for which deglycosylation is rate-limiting, ensuring therefore that the apparent reaction rate reflects the water-mediated deglycosylation step.

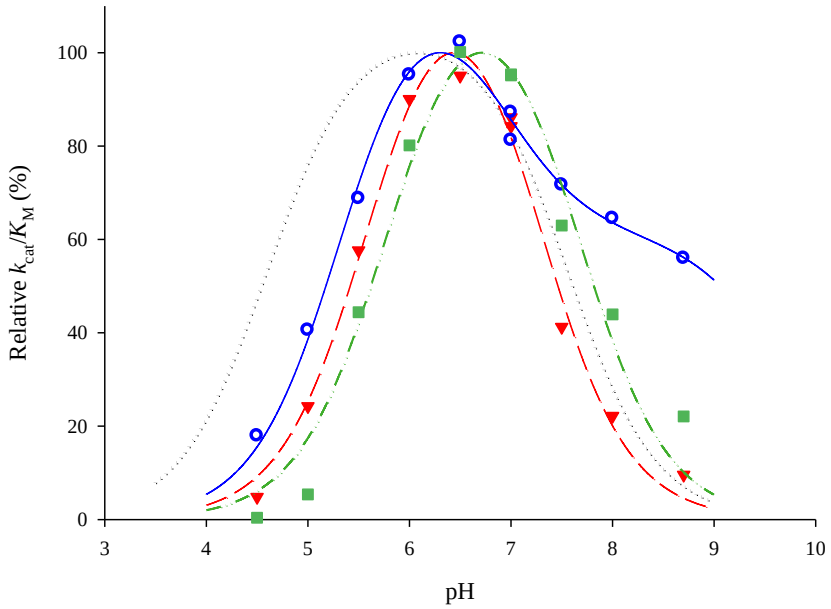


Fig. SI 10. pH-dependency of k_{cat}/K_M for *Tx*Abf (dotted black), L352M (▼, dashed red), R69H (○, solid blue) and R69H-L352M (■, dash-dot-dot green). For each enzyme, the relative k_{cat}/K_M represents a percentage of the maximum $(k_{cat}/K_M)_{max}$ obtained from the theoretical model, for experimental (symbols) and model-fitted points (lines).

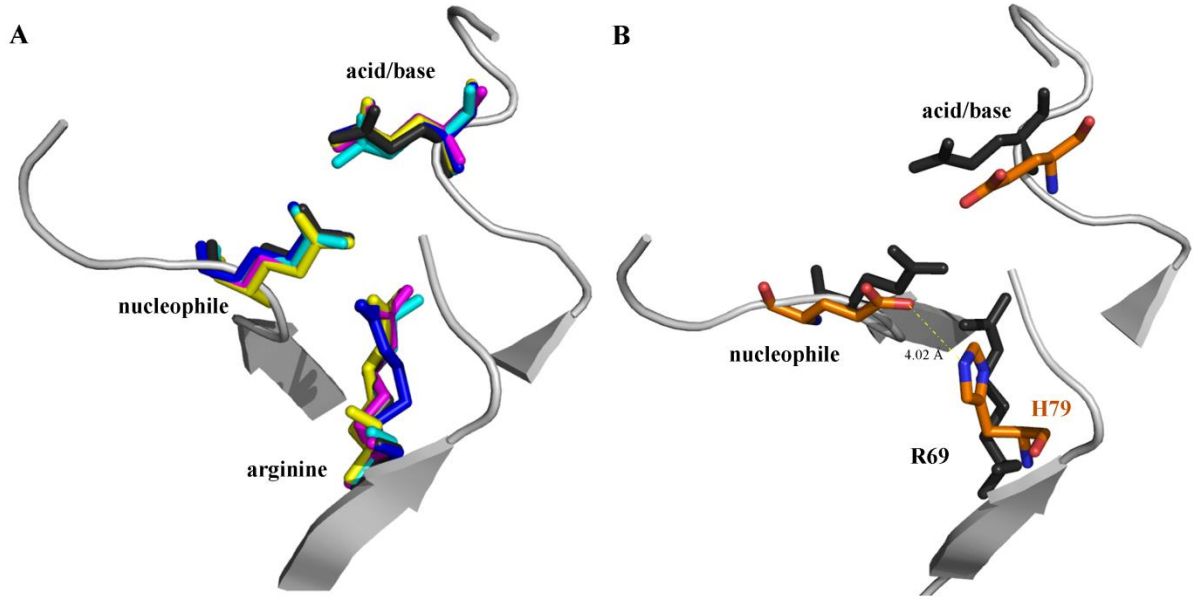


Fig. SI 11. Superimposition of conserved motifs [Arginine-acid/base-nucleophile residues] for several GHs of (A) different clan A enzymes from family GH1 (*Tt* β-glycosidase, PDB ID: 1UG6-[R75-E164-E338], cyan), GH5 (xyloglucanase, PDB ID: 2JEP-[R87-E182-E323], pink), GH39 (XynB, PDB ID: 2BS9-[R52-E278-E160], yellow), GH42 (*Tt* βGal, PDB ID: 1KWG-[R32-E141-E312], blue), GH51 (*Tx*Abf, PDB ID: 2VRQ-[R69-E176-E298], black). (B) Some variations can be observed as in the case of Xyn10A from GH10 (PDB ID: 1CLX-[H79-E127-E246]) for which a histidine is found in place of an arginine. The structures were aligned with *PyMol* software alignment tool using only this three amino-acid sequence as an object to be aligned against *Tx*Abf (2VRQ) motif, with RMSD values inferior to 0.52 Å.

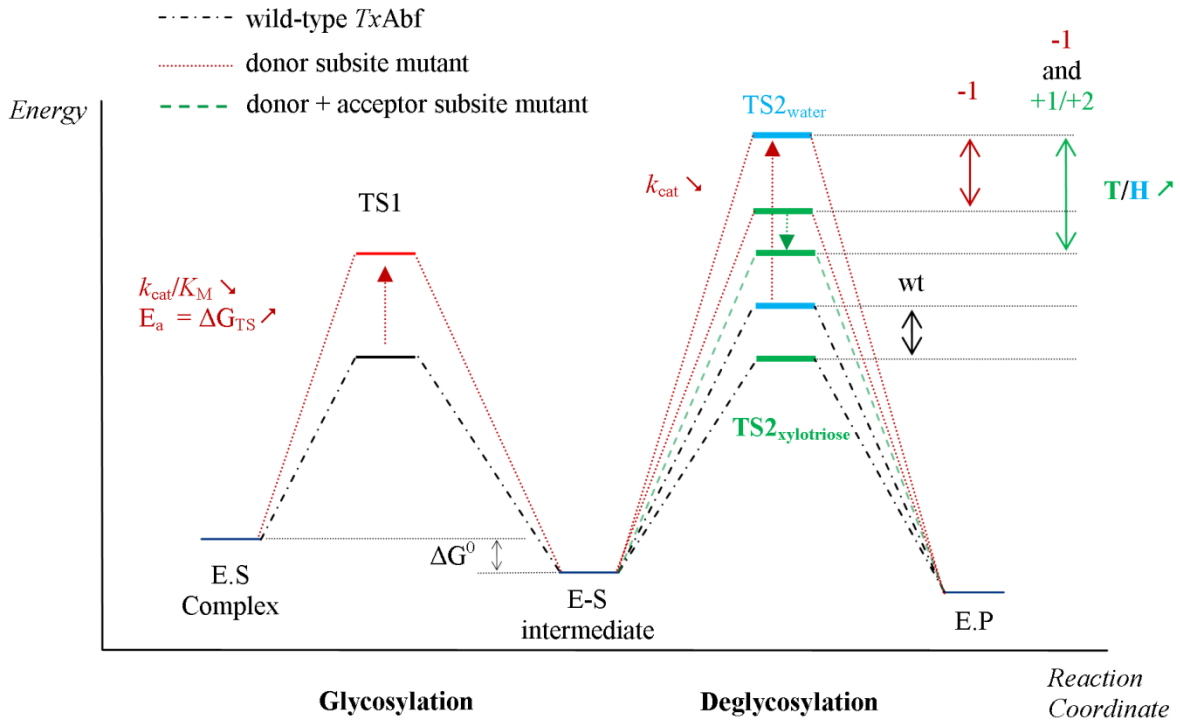


Fig. SI 12. Energetic diagram of the two step displacement mechanism of retaining GHs (black dash-dot) and putative alternative energetic pathways for evolved transglycosylases. Introduction of -1 subsite mutations would provoke a destabilization of the TS2 (red dotted line) mediated by water (—), mechanically mirrored in TS1 energy level increase (approximated by $\Delta\Delta G^{\ddagger} = -RT.\ln[(k_{cat}/K_M)^{mutant}/(k_{cat}/K_M)^{wt}] = 1.95$ kcal.mol^{-1} for R69H relative to *Tx*Abf). This increase in absolute energy levels of $TS2_{water}$ (—) as well as $TS2_{xylotriose}$ (—) necessarily comes with a greater energy difference between both TS2 since the T/H ratio increased (3 to 6 kcal.mol^{-1}, as predicted by Arrhenius law, Figure SI 13). This energy gap can be further widened by providing positive mutations in acceptor subsites. Besides, since the interaction network is altered, the E-S intermediate energy level could also be re-up (less stable) favoring as well an acceptor-mediated deglycosylation. A combination of both phenomena is not excluded to explain transglycosylases behavior.

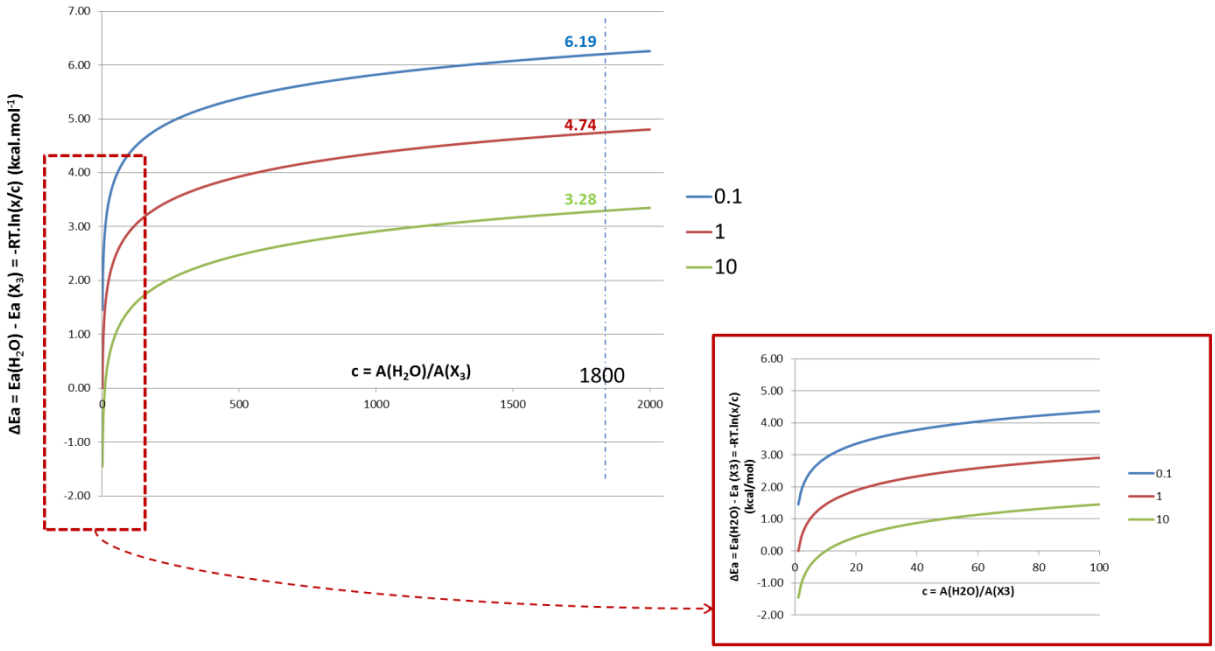


Fig. SI 13. Difference of activation energy barrier between water and xylotriose-mediated deglycosylation as a function of the attempt frequency ratio (c value). This difference was derived from Arrhenius' equation, for different $k_{H2O}/k_{xylotriose}$ values (denoted x) representative of hydrolytic (green, x = 10), 'mixed' (red, x = 1) or transglycosylating (blue, x = 0.1) enzymes.

For calculation details, according to the Arrhenius law:

$k_{H2O} = A_{H2O}.\ \exp(-Ea^{H2O}/RT)$

$k_{X3} = A_{X3}.\ \exp(-Ea^{X3}/RT)$

A_{H2O} and A_{X3} are the pre-exponential factors, also called attempt frequency, and represent the number of collisions that can occur (leading or not to reaction). Therefore A_{H2O}/A_{X3} = c = constant, whichever enzyme is considered (wild-type or mutant) and c > 1 since $A_{H2O} >> A_{X3}$.

Hence, $k_{H2O}/k_{X3} = x = c.\exp[-(Ea^{H2O}-Ea^{X3})/RT] = c.\exp[-(\Delta Ea^{H2O-X3})/RT]$

Also written: $\Delta Ea^{H2O-X3} = -RT.\ln(x/c)$

Therefore, for a given attempt frequency, the more x is decreased the higher ΔEa^{H2O-X3}, meaning an decrease of H/T ratio. Starting from an enzyme naturally evolved to perform hydrolysis(rather low Ea^{H2O}) decreasing k_{H2O} to significantly decrease x (i.e., k_{H2O}/k_{X3}) allows more room for maneuver on an energetic point of view.

Numerically, if one consider a ratio of attempt frequency of ~ 1800 (55,000 mM for H_2O over 30 mM for xylotriose), decreasing x value from 10 down to 0.1 is reflected by an increase of ΔEa^{H2O-X3} from 3.3 to 6.2 kcal.mol^{-1} (with T = 318 K and R = 1.9872.10^{-3} kcal.mol^{-1}.K^{-1}).

Table SI 1. Sequence analysis of the three selected clones from L352M-based random library.

Clone	Wild-type codon	Mutant codon	AA wt	AA mutant	Activity Ratio*
MP2-F9 (n°165)	**CGC (205-207)**	**CAC**	**R69**	**H**	2.44
	AAA (244-246)	AAT	K82	N	
	GGC (250-252)	GGC	G84	S	
	CTG (1054-1056)	ATG	L352	M	
	CTG (1252-1254)	ATG	L418	M	
	ACC (1327-1329)	ACT	T443	T	
MP3-A5 (n°197)	**CGC (205-207)**	**CAC**	**R69**	**H**	2.82
	GTG (673-675)	ATG	V225	M	
	CAC (991-993)	CAT	H331	H	
	CTG (1054-1056)	ATG	L352	M	
MP3-E1 (n°241)	**CGC (205-207)**	**CAC**	**R69**	**H**	2.32
	GTG (388-390)	GAG	V130	E	
	AGG (472-474)	GGG	R158	G	
	CTG (1054-1056)	ATG	L352	M	
	CTG (1093-1095)	TTG	L365	L	

* Activity ratios were measured from pre-treated crude extracts, on *p*NP-α-L-Ara*f* (15mM) and X_3 (10 mM), at 45 °C, 'on the lab bench'. Values from automated kinetics (mean relative error of 8%) were 2.60, 2.95 and 2.18 for F9, A5 and E1, respectively.

Table SI 2. Transglycosylation maximum yields and L-arabinofuranosyl transfer rates[a] comparison for single mutants.

Enzyme	Transglycosylation Parameters	Transglycosylation products				
		XA^3X 5.39 ppm	**A^3XX** 5.32 ppm	**A^2XX** 5.28 ppm	**Auto 1**	**Auto 2**
***Tx*Abf**	Y_{max} [b] (%)	7	12	10	7	3
	R_T [c]	7.3	12.1	13	11.9	nd
	[Enz] (nM)			18		
G179F	Y_{max} (%)	13	4	23	7	5
	R_T	12.0	nd	24.0	11.6	nd
	[Enz] (nM)			50		
N216W	Y_{max} (%)	np	np	9.1	4.8	21.8
	R_T	nd	nd	8.6	9.2	36.6
	[Enz] (nM)			19		
L352M	Y_{max} (%)	12.9	15.7	18.5	5.3	3.8
	R_T	16.8	15.9	43.9	8.61	nd
	[Enz] (nM)			50		

[a] Assays were carried out with *p*NP-α-L-Ara*f* as donor (15 mM) and xylotriose as acceptor (30 mM) at pD 5.8 and 45 °C
[b] Y_{max} is the maximum transglycosylation yield, which is not necessarily reached at the same time for all the products.
[c] R_T represents the number of µmoles of L-Ara*f* transfered on a given product for 100 µmoles of *p*NP-α-L-Ara*f* consumed. This rate is the slope of the plot of yield versus donor conversion.
nd: not determined because $R^2 < 0.97$ for the conversion-dependent plot of transglycosylation yield.
np: no product detected at this chemical shift.

Table SI 3. Conservation of amino acids R69, L352, G179 and N216 among GH51 Abfs[a] and natural occurences of possible substitutions[b].

	Occurrence frequency (%)			
Subsite location	**-1**	**-1**	**+1**	**+2**
Amino acid substitution	**R69**	**L352**	**G179**	**N216**
A	0	0	1.2	1.2
C	0	0	0	0.2
D	0	0	<1	7.0
E	0	0	0	1.6
F	0	0	0	0
G	0	0	97.0	1.2
H	0	0.2	0	1.0
I	0	44.7	0	0.6
K	0	0	0	0.2
L	0	52.3	0	0.4
M	0	0	0	0
N	0	0	0	**25.2**
P	0	0	0	0
Q	0	0	0	0.4
R	100	1.0	0	0.6
S	0	0.2	<1	53.9
T	0	0.4	0	4.5
V	0	0	0	0.6
W	0	0	0	**0**
Y	0	0	0	0.2

[a] 488 Abf sequences from GH51 family were used to calculate amino acid conservation.
[b] Are framed in green the occurrence of the wild-type amino acid found in *Tx*Abf and in red the occurrence of the mutation of interest among GH51 Abfs.

Table SI 4. Impact of xylotriose on the overall activity for *Tx*Abf and mutants thereof.

[Xylotriose] (mM)		0	1	2	4	5	10	20	30
Relative activity (%)	wt	100	90	83	72	66	54	41	33
Ratio* of relative activities mutant/wt	wt*	1	1	1	1	1	1	1	1
	Clone 197	1	1.7	2.2	2.9	3.4	4.7	6.2	7.6
	R69H	1	1.4	1.8	2.5	2.9	4.4	6.6	8.8
	R69H-L352M	1	1.6	2.0	2.7	3.0	4.0	5.5	7.2
	R69H-N216W-L352M	1	1.4	1.9	2.9	3.4	5.7	11.1	17.1
	R69H-G179F-L352M	1	1.6	2.2	3.0	3.6	5.1	7.7	10.0

* Ratios were calculated by dividing the relative activity of the mutant by the one of *Tx*Abf for a given xylotriose concentration (e.g., at 30 mM, Ratio = 568/33% = 17.1 for R69H-L352M-N216W relative to *Tx*Abf)

Table SI 5. Apparent A^2XX synthesis/hydrolysis rates for *Tx*Abf and R69H derivatives.

Enzyme	[nM]	ν_T (%/min)[a]	ν_{HII} (%/min)[a]	Rho[b]	Rho increase (-fold)
wt	18	0.0591	-0.0290	2	1
R69H	2000	1.0237	-0.1409	7	3.6
R69H-L352M	2000	0.401	-0.0073	55	27.0
R69H-G179F-L352M	2000	0.6348	-0.0281	23	11.1
R69H-N216W-L352M (pD 7)	2000	0.9037	-0.0633	14	7.0
R69H-N216W-L352M (pD 5.8)	2000	0.3582	none	**∞**	**∞**
R69H-N216W-L352M (H_2O-pH 7)	2000	1.5667	-0.1302	12	6
R69H-N216W-L352M (H_2O-pH 5.8)	2000	0.8988	-0.0152	59	29

[a] Apparent transglycosylation (ν_T, initial phase of the plot) and secondary hydrolysis (ν_{HII}, after maximum yield) rates are derived from the time-course monitoring of A^2XX transglycosylation yield (expressed as % of 15 mM/min).

[b] Rho is equal to (ν_T/ν_{HII}) and represent the apparent synthesis/hydrolysis rate of A^2XX.

Table SI 6. L-arabinofuranosyl transfer rates* (µmol of product/100 µmol of donor) for reactions catalyzed by *Tx*Abf and R69H derivatives.

Transglycosylation products		XA^{3}X	A^{3}XX	L-Ara*f* + XA^{2}X	A^{2}XX	Auto 1	Auto 2	R² values	X error (%)
Enzyme	**[Enzyme] (nM)**	5.40 ppm	5.32 ppm	5.30 ppm	5.28 ppm	5.96 ppm	5.89 ppm		
*Tx*Abf (pD 5.8)	18	7.3	12.1	nd	13.0	11.9	nd	> 0.99	1.6
R69H (pD 7)	2000	34.6	nd	8.9	55.9	nd	nd	> 0.99	1.2
R69H-L352M (pD 7)	2000	nd	nd	12.7	71.9	nd	nd	> 0.99	1.0
R69H-G179F-L352M (pD 7)	2000	nd	nd	15.1	71.5	nd	nd	> 0.98	1.0
R69H-N216W-L352M (pD 7)	2000	nd	nd	5.5	90.2	nd	nd	> 0.99	1.3
R69H-N216W-L352M (pD 5.8)	2000	nd	nd	6.2	89.2	nd	nd	> 0.98	0.9
R69H-N216W-L352M (H_2O-pH 7)	2000	nd	nd	5.0	82.9	nd	nd	> 0.98	3.6
R69H-N216W-L352M (H_2O-pH 5.8)	2000	nd	nd	nd	79.0	nd	nd	> 0.99	3.8

* L-arabinofuranosyl transfer rate (R_T) corresponds to the slope of the donor substrate conversion-dependent plot of the transglycosylation product yield (e.g., Fig. 5A). R_T represents therefore a selectivity factor for the transfer of the covalently bound L-Ara*f* moiety toward xylotriose acceptor (transglycosylation) *vs* H_2O (hydrolysis).
nd: not determined since the profile was not linear enough (i.e., $R^2 < 0.95$) to provide an accurate transfer rate value.

Table SI 7. Steady-state kinetics parameters of *Tx*Abf and mutants thereof determined for a series of aryl-α-L-Ara*f*, at pH 7 and 45 °C.

Enzyme	LG	pK_a	Analysis mode*	K_M (mM)		K_i (mM)		Ns ($s^{-1}.mM^{-1}$)		k_{cat} (s^{-1})		k_{cat}/K_M ($s^{-1}.mM^{-1}$)		Bronsted Correlation	
				value	error	value	error	value	error	value	error	value	error	$\log(k_{cat}/K_M)$	$\log(k_{cat})$
*Tx*Abf	3,4-dNP	5.36	C	0.16	0.01	20.15	1.32	-	-	135.9	1.8	849.5	48.3	2.93	2.13
	NCP	6.46	A	0.32	0.02	-	-	-	-	226.7	4.9	698.9	56.1	2.84	2.36
	4-NP	7.18	A	0.25	0.01	-	-	-	-	138.8	1.9	566.7	40.3	2.75	2.14
	3-NP	8.39	A	0.57	0.03	-	-	-	-	86.1	1.3	151.1	10.8	2.18	1.94
	Ph	9.99	A	2.00	0.08	-	-	-	-	92.0	1.2	46.1	2.4	1.66	1.96
L352M	3,4-dNP	5.36	C	1.72	0.08	55.80	5.79	-	-	310.2	6.9	179.9	12.3	2.26	2.49
	NCP	6.46	A	3.59	0.13	-	-	-	-	906.6	10.9	252.9	11.9	2.40	2.96
	4-NP	7.18	A	8.98	0.28	-	-	-	-	233.3	3.4	26.0	1.2	1.41	2.37
	3-NP	8.39	A	10.63	0.64	-	-	-	-	116.4	3.5	10.9	1.0	1.04	2.07
	Ph	9.99	A	32.13	2.36	-	-	-	-	48.7	2.6	1.5	0.2	0.18	1.69
R69H	3,4-dNP	5.36	C	0.027	0.002	21.61	6.22	0.037	0.005	1.7	0.02	61.8	5.4	1.79	0.22
	NCP	6.46	B	0.019	0.002	-	-	0.093	0.002	2.6	0.02	135.6	15.2	2.13	0.41
	4-NP	7.18	B	0.090	0.009	-	-	0.047	0.004	2.3	0.04	25.4	3.0	1.41	0.36
	3-NP	8.39	B	0.269	0.019	-	-	0.032	0.003	1.1	0.02	4.0	0.4	0.61	0.04
	Ph	9.99	B	1.337	0.079	-	-	null	-	1.8	0.1	1.3	0.1	0.13	0.25
R69H-L352M	3,4-dNP	5.36	B	0.08	0.01	-	-	0.014	0.001	0.56	0.01	6.95	0.72	0.84	-0.26
	NCP	6.46	B	0.03	0.00	-	-	0.052	0.001	0.76	0.01	22.98	3.08	1.36	-0.12
	4-NP	7.18	B	1.23	0.11	-	-	0.031	0.003	0.97	0.04	0.79	0.11	-0.10	-0.01
	3-NP	8.39	B	0.99	0.11	-	-	0.041	0.003	0.65	0.04	0.66	0.11	-0.18	-0.19
	Ph	9.99	A	28.68	4.40	-	-	-	-	0.81	0.09	0.03	0.01	-1.55	-0.09

* Steady-state kinetics parameters were determined using (A) the canonical Michaelis-Menten model $SA = SA_{max}.[S]/(K_M + [S])$, (B) a modified Michaelis-Menten model when activation was observed with $SA = SA_{max}.[S]/(K_M + [S]) + Ns.[S]$ and (C) an inhibition by excess substrate model $SA = SA_{max}.[S]/(K_M + [S] + [S]^2/K_i)$.

Table SI 8. Brønsted correlation coefficients

Enzyme	β_{LG1} [a]		Relative β_{LG1} [b]		β_{LG2} [c]
	with NCP	with 3,4-dNP	with NCP	with 3,4-dNP	
***Tx*Abf**	-0.35	-0.28	1	1	-0.07
L352M	-0.58	-0.44	1.66	1.57	-0.22
R69H	-0.55	-0.37	1.57	1.32	-0.03
R69H-L352M	-0.72	-0.48	2.06	1.71	0.02

[a] β_{LG1} is the slope of the pK_a-dependent plot of log(k_{cat}/K_M). Because of the bad linearity (Figure 6A) and since there is no rational or objective means available to omit from calculations either of the data points corresponding to NCP or 3,4-dNP, both of them were considered separately for β_{LG1} calculation.
[b] Relative β_{LG1} were calculated relatively to *Tx*Abf as ($\beta_{LG1}^{mutant}/\beta_{LG1}^{TxAbf}$)
[c] β_{LG2} is the slope of the pK_a-dependent plot of log(k_{cat})

Table SI 9. Steady-state kinetic parameter for the hydrolysis of *p*NP-α-L-Ara*f* by L352M and R69H mutants.

Enzyme	k_{cat}/K_M ($s^{-1}.mM^{-1}$)	Relative k_{cat}/K_m (%)	p$K_{a\ E298}$ [b]	p$K_{a\ E176}$ [b]	pK_{a3} [c]	pH_{opt} [d]
wt[e]	1671	100	4.6	7.6	-	6.1
L352M	31.6	1.9	5.6	7.3	-	6.5
R69H	12.8	0.8	5.3	6.9	9.7	6.3
R69H-L352M	0.8	0.05	5.8	7.7	-	6.7

[a] k_{cat}/K_M values are the one obtained at optimum pH for each enzyme, extrapolated from the calculated model.
[b] Apparent pK_a were determined by fitting experimental values to the bell-shaped model (Eq. 3).
[c] Compared to the usual bell-shaped model, an asymmetrical profile displaying a 'tail' on the basic side was observed for R69H and modeled according to equation 4.
[d] pH_{opt} was determined as the value of the pH when the derivative of the model, to which experimental points were fitted, is equal to zero.
[e] Data from (Bissaro et al., 2014).

Part V

Tailor-made Furanosylated Compounds:

Towards Artificial Plant Cell Wall Structures

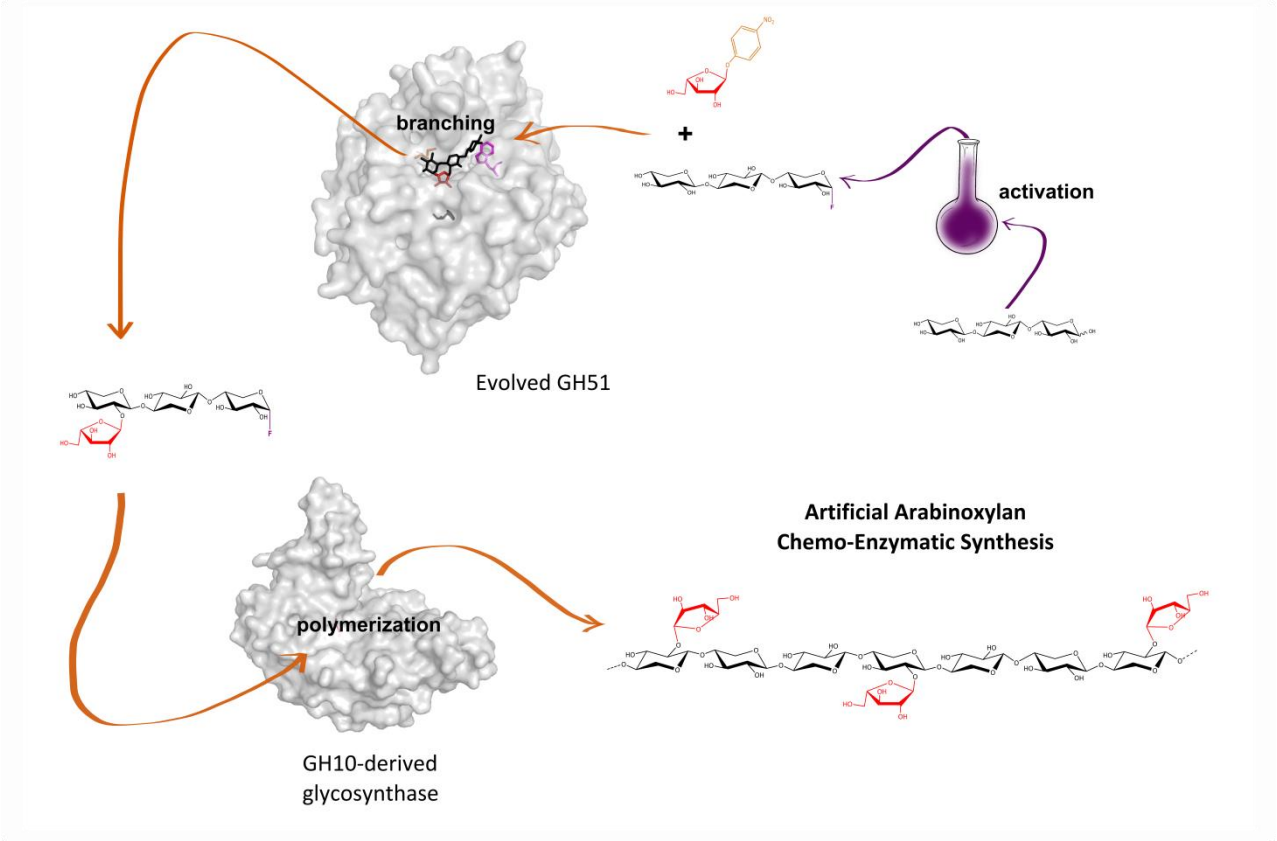
(Article in preparation)

Our interest in enzyme-catalyzed transglycosylation stems from the desire to design tools for the glycochemist. In this regard, it is rather logical that having successfully converted a FH into a transfuranosidase we wished to test the usefulness of the new enzymes for the development of chemoenzymatic strategies aimed at the *in vitro* synthesis of arabinoxylan-like structures. In the event of success, we estimate that the products of such reactions could be used as probes by plant cell wall physiologists and enzymologists alike or, from a more applied standpoint, as the basis for prebiotic molecules or bio-inspired materials. In the following section, we present some preliminary work that has been performed with this aim in mind. Bearing in mind the fact that this work is at an early stage, the perspectives and improvements that will no doubt be reached in the near future are outlined.

Towards Artificial Arabinoxylans

Bastien Bissaro[1,2,3], Yuval Shoham[4], Pierre Monsan[1,2,3,5], Régis Fauré[1,2,3] and Michael J. O'Donohue[1,2,3,*]

[1] *Université de Toulouse; INSA,UPS,INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France*
[2] *INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France*
[3] *CNRS, UMR5504, F-31400 Toulouse, France*
[4] *Department of Biotechnology and Food Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel*
[5] *Toulouse White Biotechnology, UMS INRA/INSA 1337, UMS CNRS/INSA 3582, 3 Rue des Satellites 31400 Toulouse, France*



Correspondance: M. J. O'Donohue, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, 135 Avenue de Rangueil, 31077 Toulouse cedex 4, France; Fax +33 5 6155 9400; Tel: +33 5 6155 9428; E-mail: michael.odonohue@insa-toulouse.fr

Keywords: glycosynthesis, enzymatic catalysis, glycoside hydrolase, pentoses-based carbohydrates, arabinoxylans

Abbreviations: Abfs, α-L-arabinofuranosidases; L-Ara*f*, L-arabinofuranosyl; AXs, arabinoxylans; AXOS, arabinoxylo-oligosaccharides; DP, degree of polymerization; GH, glycoside hydrolase; *p*NP-α-L-Ara*f*, *para*-nitrophenyl α-L-arabinofuranoside; *Tx*Abf, α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*; T/H, Transglycosylation/Hydrolysis ratio; *trans*-*Tx*Abf$^{(-1\ mutant)}$, *Tx*Abf-R69H-L352M, *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$, *Tx*Abf-R69H-N216W-L352M; X, donor substrate conversion rate; XOS, xylo-oligosaccharides; XT6, xylanase from *Geobacillus stearothermophilus*; D-Xyl*p*, D-xylopyranosyl; Y, yield.

1. Introduction

Glycosylation, leading to the formation of glyco-conjugates, oligo- and polysaccharides, is a fundamental reaction in biology that has far reaching implications in a myriad of biological events, such as cell-cell interactions or cell structuration (Varki et al., 2009). With regards to the formation of a covalent glycosidic bond between two monosaccharides, this is a well-established reaction (A. Michael, 1878), although some of the underlying chemical principles still remain to be fully elucidated (e.g. the electronic effects that control anomeric reactivity and product stereo-selectivity) (Jensen & Bols, 2006; Mydock & Demchenko, 2010). Despite this fact, synthetic glycochemistry is rather laborious, because the poly-hydroxylated nature of carbohydrates coupled to other inherent features of sugars (e.g. ring conformation and cyclization) makes glycosidic bond formation rather tricky, both in terms of the number of chemical steps involved and the precise control of the outcome of the reaction. Therefore, classical glycochemistry is ill-adapted for the synthesis of complex branched oligo- and poly-saccharides. In this regard, it is fortunate that the activities of glycosidic bond-forming enzymes have been harnessed, thus providing the glycoschemist with new tools. Carbohydrate-active enzymes are powerful tools, because they are stereo-specific and can be used to perform reactions that are often characterized by controllable regioselectivity. So far the *in vitro* chemo-enzymatic synthesis of complex carbohydrate structures has been demonstrated using both glycosyltransferases (Breton et al., 2012; Lairson et al., 2008) and glycoside hydrolases (GHs) (Schmaltz et al., 2011). Regarding the latter, those employed were either natural transglycosylases, such as glucansucrases (Monsan et al., 2010), or modified GHs such as glycosynthases, whose hydrolytic activity has been abolished (Mackenzie et al., 1998; Malet & Planas, 1998), or other GHs that were first been engineered using directed evolution approaches (Feng et al., 2005; Osanjo et al., 2007; Bissaro et al., 2014). Nevertheless, when one considers the chemoenzymatic synthesis of complex branched oligo- and poly-saccharides, there are actually very few examples in the literature. Among these rare examples, it is pertinent to cite a study that focused on the production of antigenic oligosaccharides (Champion et al., 2009) and another that targeted the synthesis of the primary plant cell wall motifs (Fauré et al., 2006).

Regarding the study of plant cell walls, interest in this field has been amplified by recent developments in biorefining of lignocellulosic biomass, a technology that requires in-depth knowledge of the secondary plant cell wall structures. In this respect, the study of heteroxylans and the interaction of these with other cell wall components are of particular interest, because it is widely held that these polysaccharides are important sources of biomass recalcitrance, notably because they bind both to cellulose and lignin via hydrogen bonding and covalent linkages (Buranov & Mazza, 2008). Moreover, from an industrial standpoint, heteroxylans constitute the principle source of pentose sugars, the conversion of which into value-added products is considered vital for

the economic viability of future biorefineries (Hermann et al., 2007; Dumon et al., 2012). Arabinoxylans (AXs) are a subclass of heteroxylans that constitute the main hemicelluloses in cereals and some other monocotyledons (Ebringerová, 2005). These polysaccharides display complex structures (Fig. 1) against which microorganisms (e.g. xylanolytic bacteria) have developed elaborate hydrolytic enzyme arsenals (Gilbert, 2010; Rakotoarivonina et al., 2012). Whereas the enzymes that are involved in the degradation of AXs are rather well known and characterized, those involved in the biosynthesis of AXs are still not all known, despite the fact the major glycosyltransferases responsible for AX biosynthesis have been identified over the last few years (Pauly et al., 2013) (Fig. 1). Likewise, the ability to synthesize *in vitro* arabinoxylan-like structures displaying precise structures and properties, such as controlled viscosity, gas permeability and gelation capacity (Saeed et al., 2011; Zhang et al., 2011; Niño-Medina et al., 2009) could be a source of innovation in the field of bio-based materials. Similarly, the synthesis of tailor-made arabinoxylo-oligo- and polysaccharides could open the way towards more potent AX-based prebiotics (Broekaert et al., 2011; Neyrinck AM et al., 2011) or anti-oxidative compounds (Noaman et al., 2008), or could even serve as smart ligands/substrates for the investigation of structure-function relationships in proteins and pentose-acting enzymes, notably those involved in AX biosynthesis and degradation.

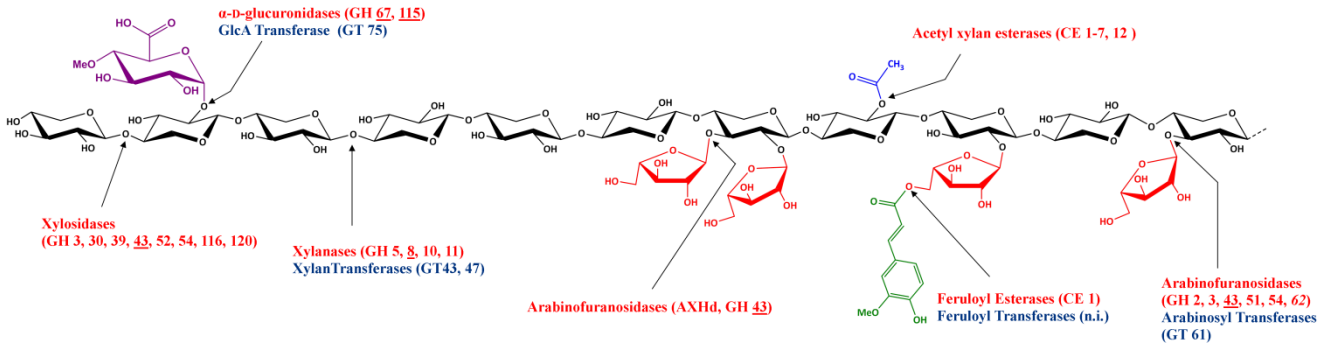


Fig. 1. Plant cell wall heteroxylan structure and enzymes involved in its hydrolysis (red) or biosynthesis (blue). Of note, acetyl groups can also be found at *O*-3 position of D-Xyl*p* residues. Besides, L-Ara*f* moieties can be also substituted by β-(1,2) linked D-Xyl*p* units (introduced by xylosyltransferases from GT61) or by α-(1,3) linked L-Ara*f*. Most of these hydrolytic enzymes act with a retaining mechanism unless otherwise stated (underlined = inverting, and *italic* = unknown mechanism). *n.i.*: not identified yet

The central aim of the present study was to provide a proof of concept for the *in vitro* synthesis of tailor-made arabinoxylo-oligosaccharides (AXOS). To achieve this, we have sought to combine engineered pentose-acting GHs in order to implement a chemo-enzymatic cascade leading to the production of target products. This work has been facilitated by the recent engineering of the GH51 α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*Tx*Abf), a hydrolytic enzyme that has been converted into a *bona fide* transarabinofuranosylase (Part IV). This engineered enzyme is able to add L-arabinofuranosyl (L-Ara*f*) moieties onto short xylo-oligosaccharides (xylobiose and xylotriose),

which can then be assembled using an engineered GH10 xylanase and appropriate chemistry (Ben-David et al., 2007).

2. Results and Discussion

2.1 Assaying the acceptor natural plasticity of *Tx*Abf active site

To ascertain the glycosynthetic potential of engineered transarabinofuranosylases, 25 different pyranosidic acceptors (Fig. SI 1) were assayed in reactions involving a L-arabinofuranoside donor moiety and engineered transfuranosylases. The enzyme chosen for this work is a mutant derivative of *Tx*Abf (denoted as *trans*-*Tx*Abf$^{(-1\ mutant)}$) whose catalytic behavior is that of a 'non-Leloir' transarabinofuranosylase (Part IV). The mutations that were introduced to create this activity are localized in the -1 subsite (R69H and L352M, Fig. SI 4) and thus were not expected to radically alter acceptor subsite topology.

To perform the assay, a previously developed screening procedure was adopted that relies on the comparison of *p*NP release from *p*NP-α-L-Ara*f* in the presence and absence of acceptor compounds (Koné et al., 2009; Arab-Jaziri et al., 2014) (Part IV). Although strictly speaking the release of *p*NP in this assay reflects enzyme glycosylation, a step which precedes and is common to both hydrolysis and transglycosylation, to some extent it also serves witness to the deglycosylation step, which is rate-limiting in the reaction (Fig. SI 2). In the case of *trans*-*Tx*Abf$^{(-1\ mutant)}$, this rate is accelerated in the presence of suitable acceptors provided that these are not also strong inhibitors.

Overall, the assay revealed that all compounds containing a D-xylose backbone are potentially suitable acceptors. However, according to the results, transfer onto α-L-Ara*f*-(1,5)-L-Ara*f* structures did not occur (Fig. 2), which means that *trans*-*Tx*Abf$^{(-1\ mutant)}$ cannot be used to prepare arabinan-type motifs. Interestingly, the introduction of an extra 6-hydroxymethyl function within the acceptor structure did not apparently hamper recognition by the enzyme, since cellobiose, cellotriose and even activated cellobiose provoked an increase in the rate of *p*NP release. In this respect, even the more complex β-D-Glc*p*-based trisaccharide α-D-Xyl*p*-(1,6)-β-D-Glc*p*-(1,4)-D-Glc, which is derived from plant cell wall xyloglucan, apparently acted as an activator of the reaction. Furthermore, the results revealed that when using β-D-Xyl*p* or β-D-Glc*p*-based acceptors, the acceleration of *p*NP release was stronger when disaccharides or aryl-glycosides were employed, indicating that occupation of at least two subsites is necessary for molecules to act efficiently as acceptors. According to our previous observations (Review and Part IV), this probably translates into the fact that quite extensive subsite interactions are required to sufficiently lower the transition state energy of deglycosylation.

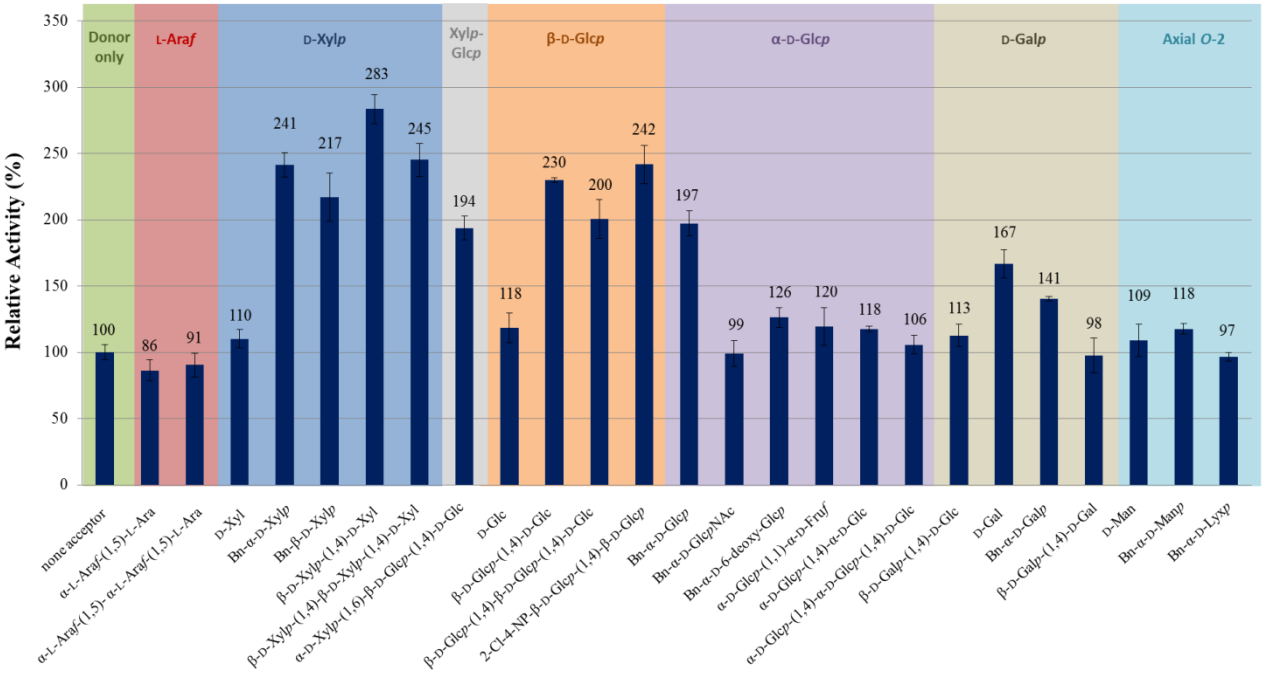


Fig. 2. Assessing positive subsites plasticity of *Tx*Abf through acceptor library screening. Reactions, performed in triplicates (error bars indicate standard deviation), were catalyzed by *trans*-*Tx*Abf$^{(-1\ mutant)}$ using *p*NP-α-L-Ara*f* as donor substrate (5 mM) and a given acceptor (30 mM), at pH 7.0 and 45 °C. As a control experiment, a reaction containing only the donor substrate was realized 7 times at different position within the microtiter plate, 100% corresponding to 0.64 ± 0.04 IU.mg^{-1}.

Regarding the impact of acceptor anomery, generally the α-configured substrates were inappropriate, since they did not procure significant acceleration of *p*NP release. Nevertheless, the use of either Bn-α-D-Xyl*p* or Bn-α-D-Glc*p* led to approximately 200% relative activity ratio. In the case of Bn-α-D-Xyl*p*, we have previously shown that this substrate docks into the active in a rather unexpected way, occupying subsites +1 and +2'. This alternative binding mode leads to a regiospecific (1,2) transglycosylation reaction (Part I), which is probably also the case when Bn-α-D-Glc*p* is used as the acceptor. Interestingly, the fact that the use of Bn-α-D-Glc*p*NAc does not accelerate *p*NP release is quite consistent with this analysis, because this acceptor's C-2 bears a *N*-acetyl group. Likewise, in the light of this analysis, it is also logical that acceptors displaying an axial configuration at *O*-2 (i.e. D-Man*p* or D-Lyx*p* derivatives) also failed to accelerate *p*NP release. In this respect, it is interesting to note that a recent study revealed that the GH51 α-L-arabinofuranosidase from *Clostridium thermocellum* (*Ct*Abf) is able to transfer D-Gal*f* moieties onto a variety of pyranosidic acceptor sugars, including α-configured *p*NP-D-Glc*p*, D-Gal*p*, D-Man*p* and L-Rha*p*, procuringto mixtures of (1,2), (1,3), (1,4) and (1,6) regioisomers in different proportions depending on the acceptor (Chlubnová et al., 2014).

In order to complete the preliminary analysis of acceptor promiscuity and to gain further insight into the reactions involving D-Glc*p*-based acceptors, the reaction involving *p*NP-α-L-Ara*f* and cellobiose acceptor was monitored using NMR. To perform the reaction, *trans*-*Tx*Abf$^{(-1\ mutant)}$, was replaced by *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$, which is a mutant that harbors an additional mutation (N216W in +2 subsite) relative to *trans*-*Tx*Abf$^{(-1\ mutant)}$ and that catalyzes a regiospecific reaction in which a (1,2) linkage is formed (Part I). This combination procured a 77% yield of a product that upon characterization was shown to be α-L-Ara*f*-(1,2)-β-D-Glc*p*-(1,4)-D-Glc (Fig. SI 3).

2.2. AXOS Assembly

2.2.1. Strategy

Even though we had demonstrated that mutant derivatives of *Tx*Abf are able to use a wide range of acceptors to perform transglycosylation, further work was nonetheless pursued using β-D-Xyl*p*-based acceptors which offer the possibility to access arabinoxylan-type structures (AX) that are representative of those found in plant cell walls. Therefore to synthesize AXs, the *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$ was used in tandem with the GH10 xylanase-derived glycosynthase, XT6-E265G (Ben-David et al., 2007). The choice of the latter was driven by the fact that, assuming that hydrolytic specificity and transglycosylation regioselectivity are related (Ajisaka & Yamamoto, 2002; Dion, Osanjo, et al., 2001; Dion, Nisole, et al., 2001), members of family GH10 reputedly display wider substrate selectivity and are able to accommodate decorated xylo-oligosaccharides, unlike members of family GH11 that are more active on undecorated xylans (Pollet et al., 2010) (Fig. SI 5). Regarding XT6-E265G, this mutant is a glycosynthase that has already been used for the chemo-enzymatic synthesis of xylan (Ben-David et al., 2007), catalyzing the autocondensation of activated xylobiose (α-X_2-F) and procuring D-Xyl*p*-based oligo- or polymers with DP up to 100. In the present work, the idea was to move one step further, using XT6-E265G to catalyze the polymerization of A^2XX, thus procuring a decorated xylan polymer. However, because it is necessary to supply XT6-E265G with fluorinated donor substrates, it was first necessary to determine at what stage in the process the fluoride moiety should be introduced, given the fact that two possibilities exist (Fig. 3)

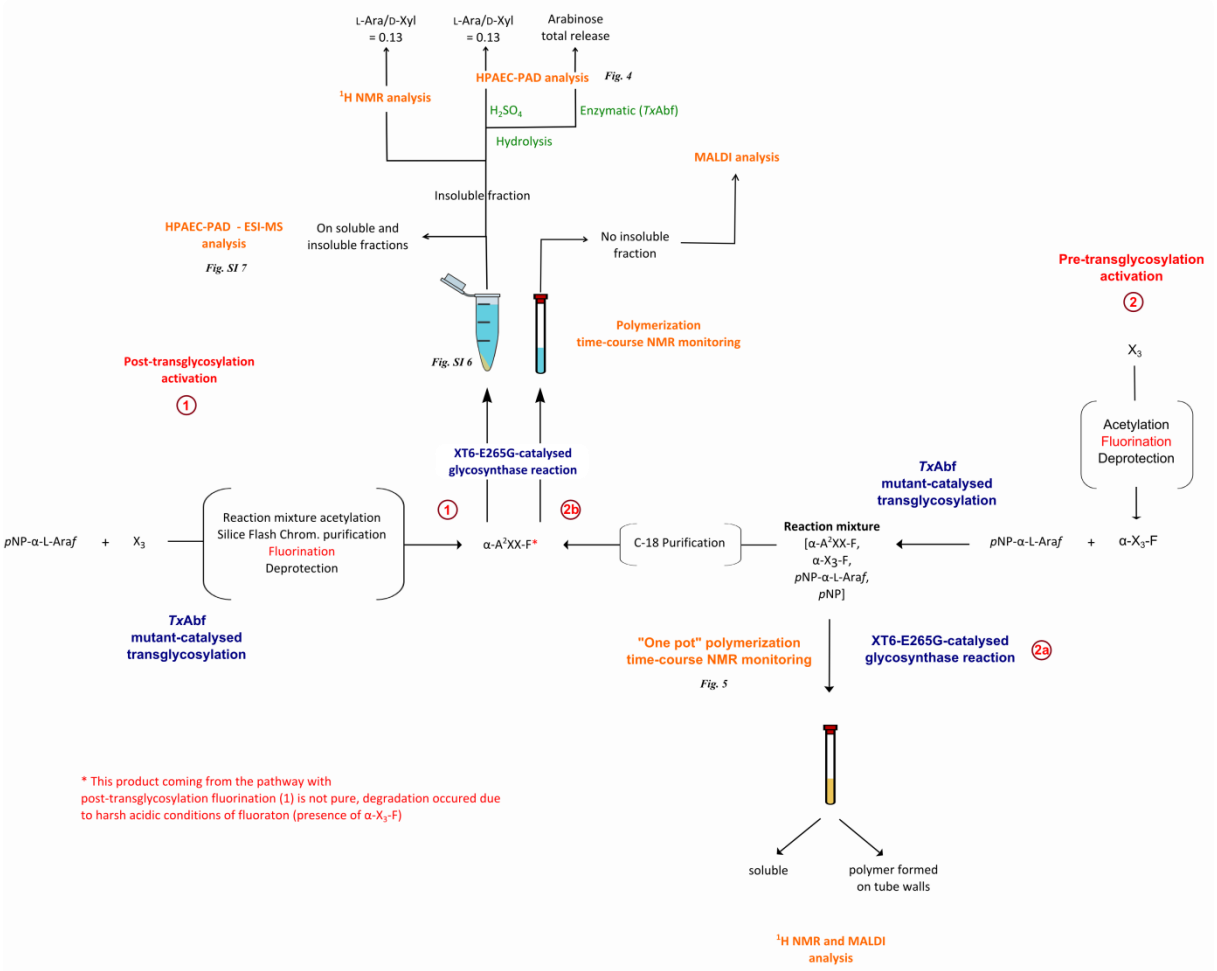


Fig. 3. Chemo-enzymatic pathway for AXOS assembly. This assembly involves the action of *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$, catalyzing the transfer of the L-Ara*f* moiety onto the non-reducing D-Xyl*p* of xylotriose. Then, the resulting A^{2}XX transglycosylation product was polymerized through the action of the glycosynthase XT6-E265G. To be employed as a donor by XT6-E265G, A^{2}XX was activated (fluorination) either following pathway 1 (i.e. after transarabinofuranosylation reaction) or pathway 2 (i.e. before transarabinofuranosylation, directly on xylotriose that serves as acceptor in *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$-catalyzed transglycosylation).

2.2.2. Sequential fluorination and polymerization of A^{2}XX

To fluorinate A^{2}XX, it was first necessary to introduce protecting groups. This was achieved directly after the enzyme-catalyzed synthesis of A^{2}XX, profiting from the per-*O*-acetylation to separate the reaction product from the acceptor, xylotriose. Therefore, the first pathway conveniently combined protection and purification and provided a direct route for fluorination of the pure acetylated A^{2}XX. However, based on TLC monitoring, some degradation occurred during the fluorination reaction (data not shown). It is likely that this was due to the harsh acidic conditions (hydrogen fluoride) that prevailed, which almost certainly led to partial hydrolysis of the linkage between the L-Ara*f* residue and the non-reducing D-Xyl*p* unit of xylotriose. Despite this drawback and thus the consequent presence of both free xylotriose and L-arabinose, the experiment was pursued in order to provide a first indication of whether XT6-E265G could polymerize the decorated sugar donor.

Satisfyingly, in both the reaction containing the α-fluorinated A^2XX (reaction C) and in the two control reactions A and B (containing α-X_2-F and α-X_3-F respectively) polymerization apparently occurred within 24 h, since the reaction medium quite rapidly (approximately 2 h) became turbid after the addition of the enzyme (Fig. SI 6). The analysis of the insoluble and soluble fractions of each reaction confirmed this preliminary conclusion, since HPAEC-PAD analysis revealed quite different profiles. In the control reactions, the soluble fractions were mainly composed of a range of low DP XOS (i.e. approximately X_2 to X_6), whereas the insoluble fractions were essentially composed of XOS displaying higher DP (X_4 to X_{15}). Similarly, the insoluble fraction from reaction C was characterized by a population that eluted over the 10 to 20 minutes, whereas the soluble fraction mainly contained molecules that eluted over the range 1 to 10 min, with A^2XX eluting at 7.3 min (Fig. SI 7). Interestingly, the median elution time of the product population in the insoluble fraction of reaction C was higher than that of the products from reactions A and B, which were rather similar. This is likely to represent a population displaying a higher average molecular weight, which in turn is consistent with the presence of L-Ara*f* moieties, which make the polymer population more soluble relative to their counterparts formed in reactions A and B. This higher solubility no doubt delayed product precipitation and thus prolonged the duration of the polymerization reaction, allowing the formation of higher molecular weight products. The ^{1}H NMR analysis of the insoluble fractions confirmed the presence of a characteristic signal of the anomeric proton of a substituting L-Ara*f* moiety (Fig. SI 8C), but the complexity of the molecular structure precluded complete structural characterization. Nevertheless, analysis of the spectra did reveal that reaction C provided a product population characterized by a L-Ara/D-Xyl ratio of 0.11, suggesting that on average one L-Ara*f* had been introduced for every nine D-Xyl*p* units (i.e. one A^2XX for two xylotriose) (Fig. SI 8).

When the insoluble fractions were submitted to acid- or enzyme-catalyzed hydrolysis, HPAEC-PAD revealed some very significant and quite logical modifications. First, acid hydrolysis of the insoluble fraction (Fig. 4C) generated species that were characterized by short retention times (1 to 3 min). According to HPAEC-PAD analyses, these latter were identified as L-Ara and D-Xyl (L-Ara/D-Xyl of 0.13), which are the logical end-products of acid hydrolysis of AX-like structures. Significantly, HPAEC-PAD analysis also revealed that incubation of the insoluble fraction of reaction C with *Tx*Abf led to a modification of product profile, with a shift of the median retention time from approximately 18.5 to 15 min. Accounting for the fact that *Tx*Abf is known to hydrolyze α-(1,2)- and α-(1,3)-linked L-Ara*f* units from moderately decorated AXs, this result strongly suggests that *Tx*Abf released α-(1,2)-L-Ara*f* moieties from the insoluble product, producing free L-Ara and a denuded heterogeneous xylan population.

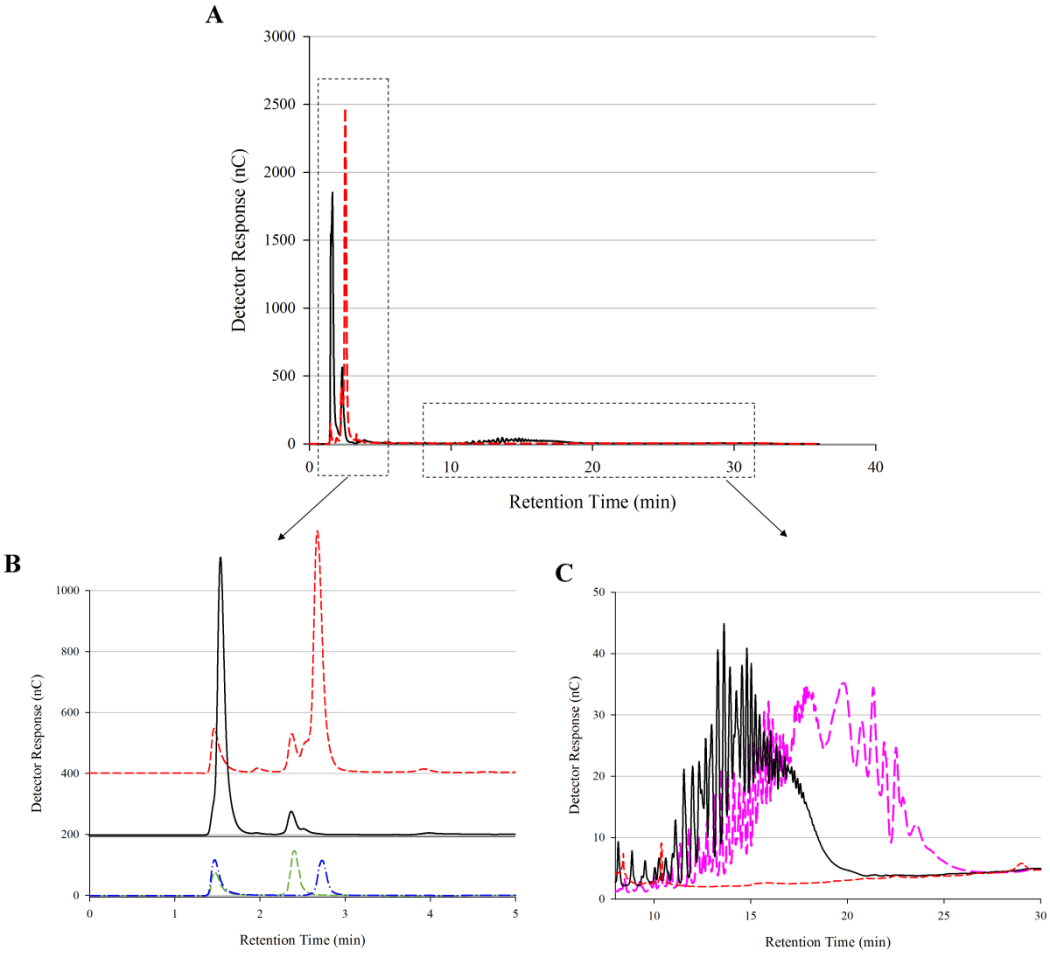


Fig. 4. HPAEC-PAD analysis of the insoluble product of the XT6-E265G-catalyzed polymerization of an α-A^{2}XX-F and α-X$_3$-F mixture. (A) Sulfuric acid (red) and *Tx*Abf-catalyzed (black) hydrolysis of insoluble products obtained from the XT6-E265G-catalyzed polymerization of impure α-A^{2}XX-F (pathway 1, reaction C). (B) Analysis of the monosaccharides (after 10-fold dilution) released after sulfuric acid (red) and *Tx*Abf-catalyzed (black) hydrolysis. The elution profiles of the products of hydrolysis are compared with those of L-Ara (green) and D-Xyl (blue). (C) Analysis of the polymer population before (pink) and after acid (red) or *Tx*Abf-catalyzed (black) hydrolysis, using an undiluted sample. All of the analyses were performed on a PA-100 column, which allowed the detection within the same run of both monosaccharides and longer polymer chains. The signal observed at t = 1.5 min corresponds to the void volume, with stronger intensity when Tris buffer (from enzyme stock solution) is present (black).

2.2.3. One-pot chemoenzymatic synthesis of artificial AX

To circumvent the problem of AXOS degradation during the fluorination step it was decided to first fluorinate xylotriose and then use α-X$_3$-F as an acceptor for the reaction catalyzed by *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$(Fig. 3). To monitor the progression of this reaction and to ensure that L-Ara*f* was successfully transferred onto α-X$_3$-F, ^{1}H NMR was employed especially to observe the H-1 atoms of both the product's L-Ara*f* moiety and that of *O*-2 substituted D-Xyl*p* unit, but also to monitor *p*NP release (Fig. 5). In this way it was possible to observe that in the first step (t < 950 min), α-A^{2}XX-F was successfully produced with a transfer rate of 0.83 µmoles of α-A^{2}XX-F / µmole of *p*NP-α-L-Ara*f*

consumed. This result demonstrates that the presence of α-fluorine on xylotriose did not hamper recognition of the acceptor by the enzyme. It is noteworthy that the acceptor α-X_3-F underwent slight autohydrolysis (confirmed by control experiments and being equal to approximately -0.22 mM/hour) during the reaction and that the significant H-1 proton signal of the fluorinated carbon of α-X_3-F was indistinguishable from that of α-A^2XX-F. This latter point is important because it precluded the quantification of α-X_3-F consumption. Advantageously, it was possible to preserve the transglycosylation product, α-A^2XX-F, from secondary hydrolysis by exploiting a previous observation that showed that secondary hydrolysis is subject to leaving group inhibition when the reaction is operated at pH 5.8 (Bissaro et al., 2014).

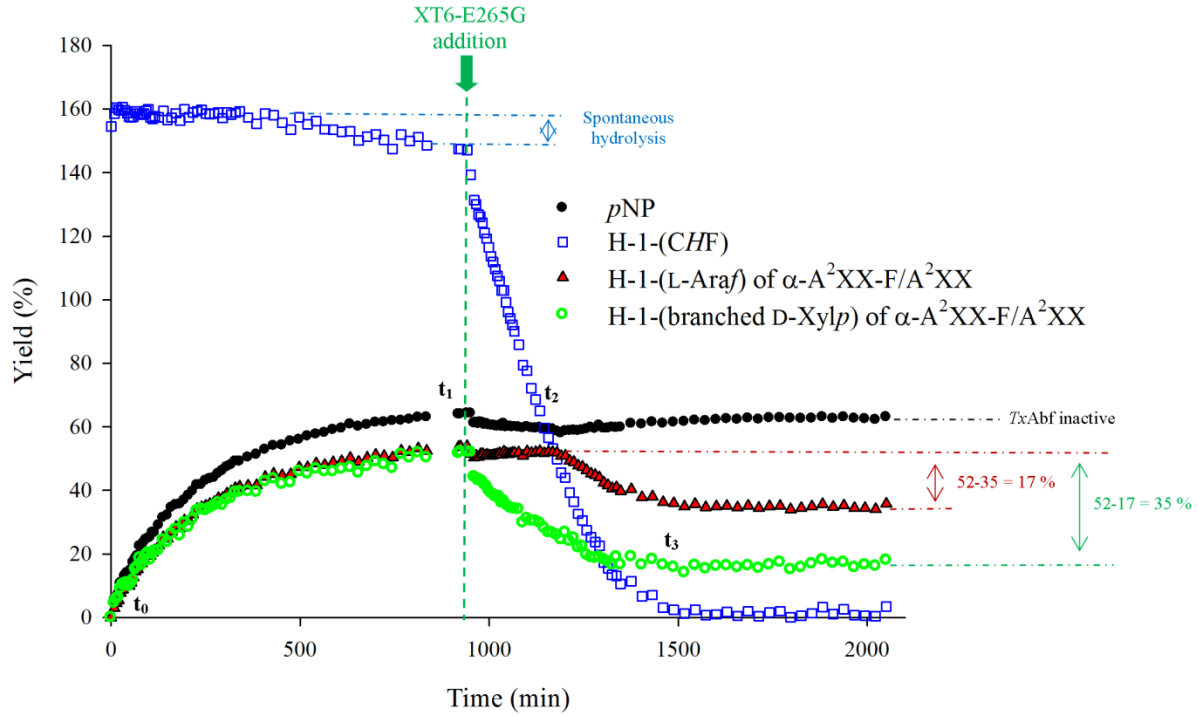


Fig. 5. Time-course ^{1}H NMR monitoring of "one pot" reaction for tailored AX synthesis. In a first phase of the reaction (t<950 min), *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$-mediated transarabinofuranosylation leads to the synthesis of A^2XX derivatives (t_o - t_1), followed by XT6-265G xylanosynthase-catalyzed polymerization (t_1 – t_3). At the end of the reaction a film was formed on the wall of the NMR tube, constituting an insoluble fraction. The soluble fraction was removed whereas the insoluble film was suspended in NaOD after D_2O "washing". Both soluble and insoluble fractions were then analyzed through ^{1}H NMR (Fig. SI 9).

Since ^{1}H NMR monitoring clearly revealed that the formation of α-A^2XX-F was successful, the reaction was pursued by adding XT6-E265G into the same NMR tube and monitoring was pursued, focusing on the evolution of pertinent signals, especially the anomeric H-1 signals of the fluorinated species (Fig. 6). During the second phase of this one-pot reaction, several observations were made. First, the *p*NP concentration remained constant, indicating that *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$ had ceased to be active. This observation is important, because it means that any further changes to the L-Ara*f* signal could not be attributed to the continued formation/degradation of α-A^2XX-F or any other

product that involves transfer of L-Ara*f* moieties by the *Tx*Abf mutant. Second, the H-1-(C*H*F) belonging to α-X_3-F and/or α-A^2XX-F decreased rapidly, an observation that is consistent with the fact that α-X_3-F is a good donor substrate for XT6-E265G. Interestingly, the H-1 signal of the substituted D-Xyl*p* unit (4.55 ppm) also decreased as soon as XT6-E265G was added, which reflects the fact that the L-Ara*f*-substituted D-Xyl*p* moiety no longer constitutes a reducing end, since it now occupies an internal position within an oligomer of growing polymer. A symmetric increase was observed at 4.59 ppm that might be attributed to the H-1 signal corresponding to internal branched D-Xyl*p* units. Furthermore, unlike the H-1 of the branched D-Xyl*p* unit, the H-1 of the L-Ara*f* branching moiety (from the same molecule α-A^2XX-F) remained unaltered in the early stages of polymerization, which is logical if one considers that polymerization *per se* should not involve any direct changes in the environment of this proton (Ferré et al., 2000). Consequently, the apparent concentration of bound L-Ara*f* remained constant until precipitation occurred at a later stage, an event that was accompanied by an apparent decrease (from 52 to 35%) in the yield of soluble branching L-Ara*f* units (Fig. 5). At the end of the reaction, the insoluble precipitate was clearly visible, forming a film on the wall of the NMR tube. Thereby, the supernatant was removed and the aforementioned insoluble precipitate was thoroughly washed and then suspended in deuterated sodium hydroxide, thus allowing further ^{1}H NMR analysis of both the insoluble and soluble products of the reaction. Gratifyingly, these analyses confirmed the presence of branching L-Ara*f* moieties in the insoluble fraction (Fig. SI 9). Finally, it is noteworthy that despite the fact that the measurement of H-1-(C*H*F) consumption indicated that the reaction had run to completion, the H-1 signal of branched terminal β-D-Xyl*p* did not completely disappear as one might expect and represents approximately 50% of final soluble products (i.e. oligosaccharides or unincorporated A^2XX). This observation can be explained either by the presence of a large number of non-fluorinated reducing end L-Ara*f*-branched D-Xyl*p* units (i.e. formed by the motif A^2XX located at the reducing end of soluble growing oligosaccharides) or by the presence of free A^2XX. In either case, the absence of the fluoride group is indicative of hydrolysis, a probable occurrence if one considers that α-X_3-F was shown to be prone to spontaneous hydrolysis. Unfortunately, the H-1 signal of branched L-Ara*f* moieties is highly similar when a fluoride group is present or absent at the "reducing" end of the tetrasaccharide, and this is also the case for the H-1 signal of branched D-Xyl*p* units (Fig. 6). Therefore, it is impossible to accurately monitor the evolution of the different species. Nevertheless, if one considers the autohydrolysis rate that was determined for α-X_3-F, it is possible to postulate that after a 10 h-reaction approximately 15% of α-A^2XX-F should be hydrolyzed. This is consistent with our results (obtained in duplicate, Fig. SI 10) that reveal a 17% yield of soluble products (after precipitation of the insoluble products) containing reducing end branched D-Xyl*p* units.

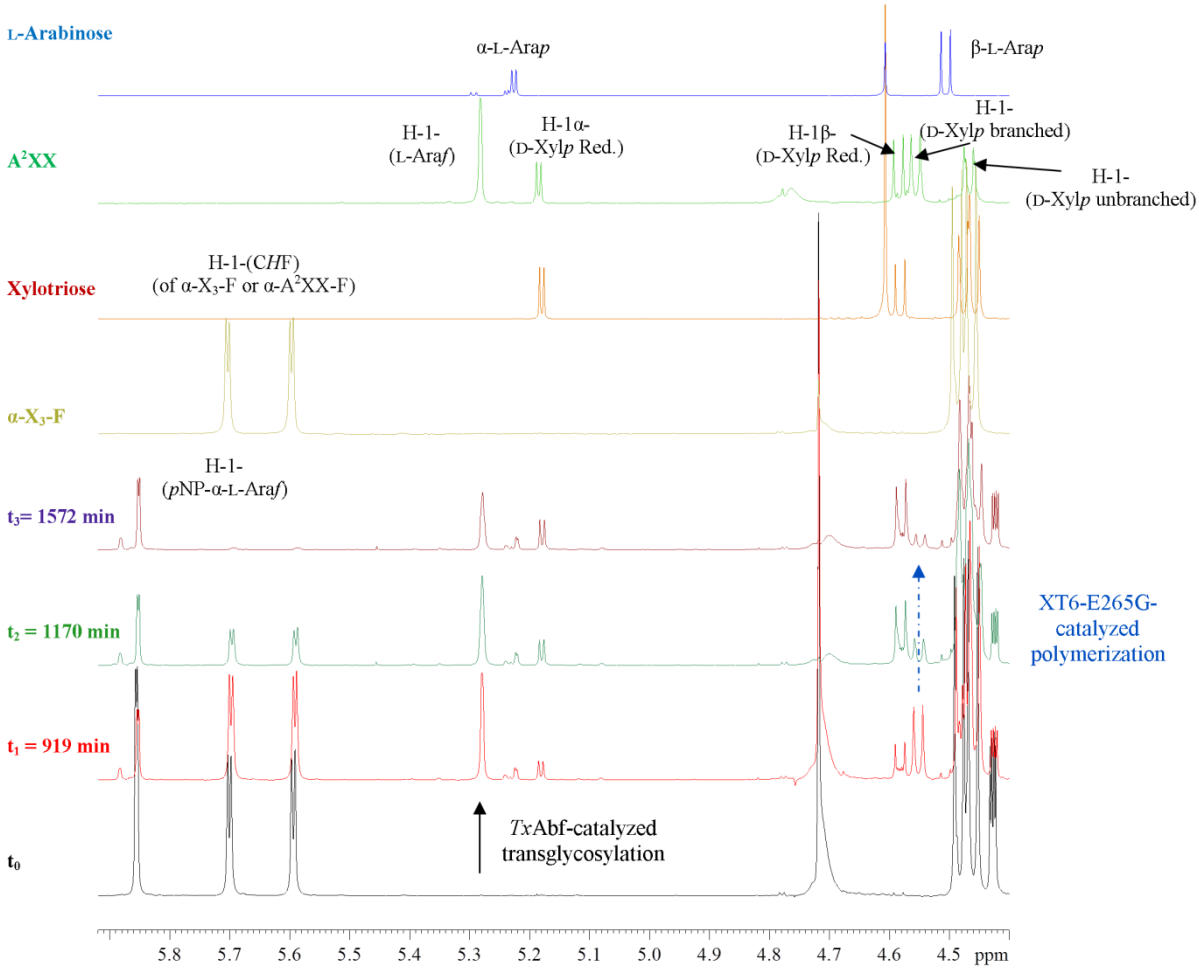


Fig. 6. ^{1}H NMR spectra superimposition showing the relevant anomeric H-1 signals being a feature for the monitoring of the different activities. Spectra corresponding to the different key moments reported on Fig. 5 were represented (i.e. t_0 = before *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$ transglycosylation start, t_1= before XT6-E265G addition, t_2 = before A^{2}XX-containing polysaccharide precipitation, t_3 = after H-1-(C*H*F) total consumption).

Despite the relative success of the one-pot reaction described above, we nevertheless tried to improve upon the reaction by separating the two reaction steps (i.e. L-arabinofuranosylation of α-X_3-F and polymerization), introducing a purification step to first isolate α-A^{2}XX-F. However, this extra step was not particularly successful, because both α-X_3-F and A^{2}XX (resulting from autohydrolysis) were detected in the α-A^{2}XX-F sample after purification. Interestingly, the use of this sample for polymerization revealed that all of the fluorinated substrates were consumed/degraded, since the H-1-(C*H*F) totally disappeared, but the H-1 signals of bound L-Araf units and branched D-Xylp units remained unchanged. Mass spectrometry analyses of the reaction product are currently ongoing. Nevertheless this experiment aroused a new question regarding the potential of A^{2}XX to act as an acceptor, since the pertinent H-1 signals of A^{2}XX did not evolve, suggesting that only α-X_3-F was used as donor and acceptor in the course of this reaction. Therefore, to further investigate this point, a glycosynthase reaction was carried out in presence of α-X_3-F (donor) and A^{2}XX (potential acceptor),

that can only fit in subsites +1 to +3 to be integrated in a growing chain. As a matter of fact, this experiment showed that A^2XX is actually poorly recognized as a direct acceptor by XT6-E265G since a very low decrease of the signal H-1-(Araf) of A^2XX was observed (approximately by 3%), probably due due to precipitation upon integration in longer chains (Fig. SI 11). A decrease of the H-1-(branched D-Xyl*p*) of A^2XX was noticed (ca. 20%) revealing that A^2XX is effectively incorporated in the growing chain, albeit to a moderate extent. As a result, a film was formed on the walls of the NMR tube and a quantitative ^1H NMR analysis consistently showed that 3% of initial soluble L-Ara*f* (from A^2XX) were present in this film. Furthermore, in the latter an L-Ara/D-Xyl ratio of 0.042 is observed, corresponding to 1 L-Ara*f* for 23 D-Xyl units. In other words, on average, one A^2XX motif is integrated for seven xylotriose molecules. The low incorporation rate of A^2XX in the insoluble fraction (i.e. long chains) coupled to the modest decrease in H-1-(branched D-Xyl*p*) of A^2XX shows that A^2XX is essentially used for short (soluble) oligosaccharide synthesis. Moreover, given the favorable equimolar ratio of (α-X_3-F:A^2XX), as opposed to the one pot reaction in which a ratio (α-X_3-F:α-A^2XX-F) of approximately (2:1) was observed at glycosynthase addition (Fig. 5), the poor A^2XX integration (20% max. versus 66%, respectively), suggests thar A^2XX is a far better donor substrate than acceptor for XT6-E265G-catalyzed reaction. This point is however rather difficult to rationalize in the light of 3-D structure observation (Fig. SI 5).

3. Conclusions and future prospects

By way of conclusion, the present study proved that the GH51 *Tx*Abf displays a wide acceptor selectivity which makes this enzyme highly interesting for acceptor subsite engineering to expand the repertoire of compounds that could be furanosylated (Dumon et al., 2012). Besides, small libraries could be constructed from evolved scaffold proteins, such as the *trans*-*Tx*Abf$^{(-1\ mutant)}$, and be easily screened using the activity ratio-based method in order to find mutants with new acceptor specificities. As an example it was shown that L-Ara*f* moieties can be transferred onto “unnatural” D-Glc*p*-based acceptors such as cellobiose, opening the way to chemo-enzymatic syntheses of L-arabinofuranoglucans. Arabinoglucans are molecules scarcely found in Nature (Basu et al., 1986) and a few structures were chemically synthesized such as the α-D-Glc*p*-(1,4)-[α-L-Ara*f*-(1,6)]-D-Glc (Fujiwara et al., 1985) involved in compounds showing anti-tumor effects.

From a fundamental standpoint, this work showed that the hydrolytic arsenal naturally devoted to AX degradation can be rerouted, through enzyme engineering, to provide a sort of basic artificial synthesis pathway. It is conceivable that such system might find its interest in the study of naturally-occurring plant cell wall biosynthetic pathways, notably by providing complex probe or model substrates. Furthermore, it is worth mentioning that this preliminary study constitutes the first

account of an *in vitro* chemo-enzymatic synthesis of simple AXs, displaying exclusively (1,2)-linked α-L-Ara*f* moieties.

However, *in planta*, AXs are much more complex than those that can be synthesized using the simple methods described in the present study. In future work it would perfectly feasible to introduce hydroxycinnamic acids such as feruloyl groups at the *O*-5 of the polymer bound L-Ara*f* moieties. Nevertheless, the strategy required to achieve feruloylation is not yet clear (de O Buanafina, 2009; Lenucci et al., 2009) and feruloyltransferases are far from being available. Lipase-catalyzed feruloylation is a well-known method to regioselectively transfer ferulic acids onto the hydroxyl functions of carbohydrates. Indeed, the feruloylation of methyl L-arabinofuranosides using Lipolase 100T operating in an organic solvent (acetonitrile, *tert*-butanol) and using a vinyl ferulate donor has been demonstrated (Mastihubová et al., 2006). Similarly, more recently, lipase-catalyzed transesterification of the glycoside arbutin has been achieved using a non-activated ferulic acid donor (Ishihara et al., 2010). Therefore, in preliminary trials we have also attempted to perform lipase-catalyzed feruloylation of the *O*-5-position L-Ara*f* moiety of the trisaccharide A^2X, using an unactivated ferulic acid donor. Unfortunately this strategy failed when using either Lipolase 100T or Novozyme 435, but no doubt there is considerable scope to optimize the reaction conditions or to use other enzymes feruloyl esterases operating in *trans*-esterification mode (Topakas et al., 2007; Vafiadi et al., 2009).

Beyond feruloylation of synthetic AXs, it is also possible envisage the formation of intermolecular crosslinking, for instance using peroxidases. It is reported in the literature that crosslinking between feruloylated L-Ara*f* moieties participates to specific properties of AXs and different diferulates linkages can also be found (Vismeh et al., 2013). The most obvious way to achieve this would be to first introduce ferulate groups and then catalyze the formation of diferulate dihydrodimers (Fig. 7). However, other methods might include the introduction of alternative 'non-natural' chemical functions that could then be used for crosslinking or other purposes, notably with material applications in mind.

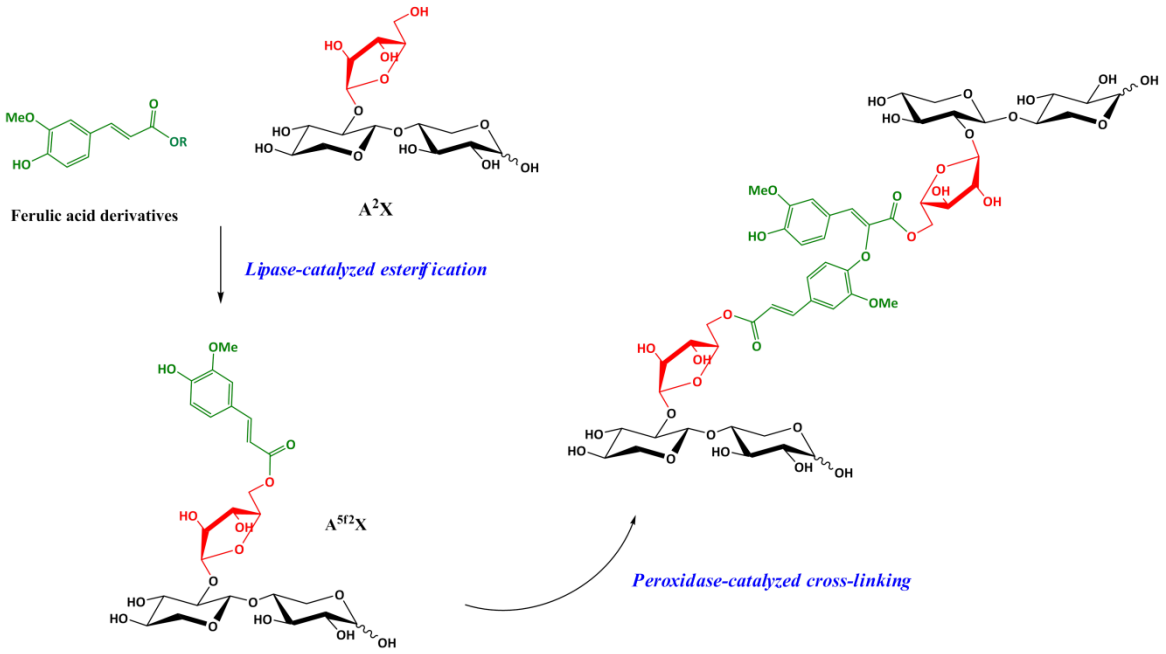


Fig. 7. Further strategy to functionalize AXOS issued from evolved *Tx*Abf-catalyzed transglycosylation.

Regarding the actual enzyme-catalyzed preparation of the artificial AXs, there is also scope for improvement, notably by circumventing the requirement for fluoride activated donors. Using sucrose as a naturally-occurring donor sugar, it has previously been possible prepare a variety of compounds using potent transglycosylases, such as amylosucrases and dextran sucrases (Monsan et al., 2010; Daudé et al., 2012). Therefore, it is reasonable to imagine that if *trans*-xylanosylases were available it would be possible to synthesize AXs, starting from simple glycosides, such as *p*NP-L-Araf on one hand and xylo-oligosaccharides or *p*NP-xylo-oligosaccharides on the other. However, so far no naturally-occurring *trans*-xylanases have been described. Therefore, it is likely that these will arise from enzyme engineering strategies, rather like the one that provided the transglycosylating *Tx*Abf mutants. Finally, accounting for the fact that the exact linkage pattern (i.e. (1,2) or (1,3)-bonds) of L-Ara*f* substitutions in natural AXs is different depending on their botanic (Saulnier et al., 2007) pattern it is possible to imagine that further engineering of *Tx*Abf will procure an enzyme that will synthesize A^3XX and related compounds in a regiospecific manner. The availability of such an enzyme would considerably extend the structural variety of synthetic AXs as would the availability of an enzyme capable of transferring a second L-Ara*f* onto an already branched L-Ara*f*-branched D-Xyl*p* group.

4. Experimental section

4.1. Substrates and reagents

Unless otherwise stated, routine experimental work was performed using chemicals purchased from Sigma-Aldrich (Illkirch, France) and molecular biology reagents purchased from New England Biolabs (Evry, France). The substrate, *p*NP-α-L-Ara*f*, was purchased from CarboSynth (Compton, UK) and xylotriose and XOS were purchased from Wako Chemicals GmbH (Neuss, Germany).

4.2. Proteins expression and purification

To produce larger batches of wild-type and mutated recombinant *Tx*Abf, protein expression and purification were performed as previously described (Bissaro et al., 2014). Regarding the E265G mutant (XT6-E265G) of the xylanase from *Geobacillus stearothermophilus* (XT6), its gene is cloned in pET9d without its signal peptide as a N-term His-tag fusion (Bravman et al., 2003; Khasin et al., 1993). The protein production (1 L) and purification protocol is similar to the one of *Tx*Abf with the following variations: cultures were performed in a selective (kanamycin, 40 µg.mL^{-1}), Terrific Broth medium (24 g.L^{-1} yeast extract, 12 g.L^{-1} tryptone, 9.4 g.L^{-1} K_2HPO_4, 2.2 g.L^{-1} KH_2PO_4 and 8 mL.L^{-1} glycerol) instead of LB medium. For purification, the cell pellet were leave to thaw on ice, before sonication (Bioblock Scientific vibracell 72434) employing cycles of (1 s 'on', 1 s 'off') during 10 min, on ice, with a probe operating at 30% of maximal power. Soluble cell content is then separated from cell debris by centrifugation (5,000 x g, 10 min, 4 °C), heat-treated (60 °C, 30 min) and clarified using centrifugation step (30 min at 11,000 x g, 4 °C). Enzyme crude extract is then purified using a Cobalt resin (TALON® Metal Affinity Resin, Clontech Laboratories, Inc.) contained in a disposable column (Bio-Spin column, Bio-Rad) and elution is achieved using a gradient of imidazole in TALON buffer as described for *Tx*Abf (Bissaro et al., 2014).

4.3. Acceptor screening using automated kinetics

The acceptor library screening was performed according to a protocol previously described (Part IV). Briefly, the activity of the *trans*-*Tx*Abf$^{(-1\ mutant)}$ (i.e. containing R69H and L352M mutations) (7.9 µg.mL^{-1} final concentration) was assayed in the presence and absence of the 25 acceptors (Fig. SI 1), in triplicates, from which a mean activity ratio (R) was derived, which is equal to $Vi_{(donor+acceptor)}/Vi_{(donor)}$. The employed acceptors belong to diverse carbohydrate series: L-arabinofuranose ((1): α-L-Ara*f*-(1,5)-L-Ara, (2): α-L-Ara*f*-(1,5)-α-L-Ara*f*-(1,5)-L-Ara), D-xylosopyranose ((3): D-Xyl, (4): Bn-α-D-Xyl*p*, (5): Bn-β-D-Xyl*p*, (6): β-D-Xyl*p*-(1,4)-D-Xyl, (7): β-D-Xyl*p*-(1,4)-β-D-Xyl*p*-(1,4)-D-Xyl), D-glucopyranose (α- and β-anomeries, (8): α-D-Xyl*p*-(1,6)-β-D-Glc*p*-(1,4)-D-Glc, (9): D-Glc,

(10): β-D-Glc*p*-(1,4)-D-Glc, (11): β-D-Glc*p*-(1,4)-β-D-Glc*p*-(1,4)-D-Glc, (12): 2-Cl-4-NP-β-D-Glc*p*-(1,4)-β-D-Glc*p*, (13): Bn-α-D-Glc*p*, (14): Bn-α-D-Glc*p*NAc, (15): Bn-α-6-deoxy-D-Glc*p*, (16): α-D-Glc*p*-(1,1)-α-D-Fru*f*, (17): α-D-Glc*p*-(1,4)-α-D-Glc, (18): α-D-Glc*p*-(1,4)-α-D-Glc*p*-(1,4)-D-Glc), D-galactopyranose ((19): β-D-Gal*p*-(1,4)-D-Glc, (20): D-Gal; (21): Bn-α-D-Gal*p*; (22): β-D-Gal*p*-(1,4)-D-Gal) and axial *O*-2-containing sugars ((23): D-Man; (24): Bn-α-D-Man*p*; (25): Bn-α-D-Lyx*p*). In practice, using a liquid handling system (Genesis RSP 200, Tecan), reactions were prepared in wells of a 96-well polypropylene PCR microtiter plate (Corning®, Sigma Aldrich). The basic reaction mixture used to determine $Vi_{(donor)}$ values was composed of 135 µL substrate solution (5.55 mM *p*NP-α-L-Ara*f* in 50 mM sodium phosphate, pH 7.0) and 15 µL (10% of final volume) of purified enzyme. To determine the value of $Vi_{(donor+acceptor)}$, the kinetic was performed using the same basic reaction mixture containing additional acceptor (30 mM). Three wells, randomly allotted within the microplate, were prepared for each "acceptor reaction". To perform reactions, the substrate solution was first incubated at 45 °C for 15 min in a Peltier heating device, before initiating the reaction through the addition of pre-diluted enzyme solution (sodium phosphate buffer, 50 mM, pH 7, 0.1 $mg.mL^{-1}$ of BSA). Once the reactions were initiated, the progression of reactions was monitored by removing 20 µL aliquots every 6 min (4 points in total, each kinetic lasting 24 min). These were immediately combined with 180 µL of 1 M Na_2CO_3 and absorbance (405 nm) was determined. Using a simple, purpose-designed Excel macro the activity ($IU.mg^{-1}$) of the different clarified lysates was calculated in the presence and absence of acceptor and activity ratios were derived for each one. The relative mean error of this method was inferior to 15% for any single acceptor reaction.

4.4. Production and chemical modification of AXOS

4.4.1. L-Arabinofuranosylated oligosaccharides production

Enzymatic transglycosylation, per-O-acetylation and purification

In a 100 mL round bottom flask, the donor substrate *p*NP-α-L-Ara*f* (123 mg, 0.45 mmol, 15 mM, 1 eq.) and acceptor (0.90 mmol, 30 mM, 2 eq.) were dissolved in sodium acetate buffer (25 mM, pH 5.8, 30 mL final volume). After incubation at 45 °C, the reaction was initiated by the addition of R69H-N216W-L352M (1 µM). After 14 h of reaction (45 °C, under magnetic stirring) the reaction was stopped by inactivating the enzyme with a cycle of liquid nitrogen freezing/95 °C, 10 min/liquid nitrogen freezing. The reaction mixture was then freeze-dried. The acceptor employed were either xylobiose (254 mg), xylotriose (373 mg) or cellobiose (308 mg) to produce through transglycosylation α-L-Ara*f*-(1,2)-β-D-Xyl*p*-(1,4)-D-Xyl (A^2X), α-L-Ara*f*-(1,2)-β-D-Xyl*p*-(1,4)-β-D-Xyl*p*-(1,4)-D-Xyl (A^2XX) or α-L-Ara*f*-(1,2)-β-D-Glc*p*-(1,4)-D-Glc*p* (L-Ara*f*-(1,2)-G_2), respectively.

Acetylation of the residue (in a mixture acetic anhydride/pyridine, 1:1 v/v, 50 mL) in the presence of

catalytic amount of dimethylaminopyridine (DMAP) was carried out at room temperature and quenched after 16 h by adding MeOH (35 mL) at 0°C. After concentration in vacuo, the residue was dissolved in CH_2Cl_2 and washed with saturated aqueous $KHSO_4$, $NaHCO_3$ (twice) and brine. The aqueous solutions were back extracted with CH_2Cl_2 (2 x). The organic layers were dried (Na_2SO_4), concentrated and purified on a silica column using the Reveleris flash chromatography System (Grace, Epernon, France), with a gradient of ethyl acetate in petroleum ether, with elution occurring between at 55 and 70%. The per-*O*-acetylated trisaccharides per-*O*-Ac-[A^2X] (137 mg, 0.183 mmol), per-*O*-Ac-[L-Ara*f*-(1,2)-G_2] (121 mg, 0.135 mmol) and tetrasaccharide per-*O*-Ac-[A^2XX] (147 mg, 0.153 mmol) were isolated in 41, 30 and 34% yields, respectively. Besides, per-*O*-Ac-[X_3] (312 mg, 0.416 mmol) issued from purification of acetylated reaction mixture of A^2XX production batch was also recovered with 46 % yield.

De*-O-*acetylation.

Compounds per-*O*-Ac-[A^2XX] (114 mg, 0.118 mmol) and per-*O*-Ac-[A^2X] (94 mg, 0.125 mmol) were treated in anhydrous MeOH/CH_2Cl_2 (1:1, v/v, 5 mL) with sodium methoxide (0.5 eq., 1 M in methanol), under nitrogen atmosphere, 1 h at 0 °C and 4 h at room temperature. The solution was neutralized with Amberlite IR120 (H^+), filtered and then concentrated under reduced pressure. The residue dissolved in deionized water was freeze-dried to give A^2X and A^2XX in quantitative yields.

4.4.2. A^2XX activation post-transglycosylation (pathway 1)

In a plastic vessel, hydrogen fluoride/pyridine (7:3, v/v, 3 mL) was added to a solution of per-*O*-Ac-[A^2XX] (75 mg, 0.078 mmol) dissolved in anhydrous CH_2Cl_2 (3 mL) and the reaction mixture was stirred at 0 °C for 5 h. Then, the solution was diluted with CH_2Cl_2 (~50 mL) and poured into a plastic beaker containing an ice-cooled ammonia in water (3 M, ~200 mL). The organic layer was washed with saturated aqueous $NaHCO_3$ (twice, for neutralization), KH_2SO_4 (to extract pyridine) and $NaHCO_3$, dried and concentrated under reduced pressure. The residue was purified on a silica column using the Reveleris flash chromatography System (Grace, Epernon, France), with a gradient of ethyl acetate in petroleum ether, with elution occurring at ca. 50 %. Per-*O*-Ac-[α-A^2XX-F] (40 mg, 0.043 mmol) was isolated in 55% yield, albeit it was not totally pure due to product instability. De-*O*-acetylation was performed according to exactly the protocol described hereinbefore, with a phase at room temperature lasting 2 h. α-A^2XX-F (22 mg) was not obtain has a pure compound precluding any yield calculation.

4.4.3. A^2XX activation pre-transglycosylation (pathway 2)

Per-*O*-Ac-[α-X_3-F] (177 mg, 0.249 mmol) was isolated in 60% yields starting from per-*O*-Ac-[X_3]

(312 mg, 0.416 mmol) using the same protocol employed for per-*O*-Ac-[A^2XX] fluorination. The deprotected product α-X_3-F (78 mg, 0.187 mmol), obtained like α-A^2XX-F, was isolated with 77% yield.

α-X_3-F (50 mg, 0.12 mmol, 30 mM, 2eq.) was then employed as an acceptor for a transglycosylation batch reaction, in a 15 mL tube with *p*NP-α-L-Ara*f* (16.3 mg, 0.06 mmol, 15 mM, 1 eq.) as donor substrate dissolved in sodium acetate buffer (25 mM, pH 5.8, 4 mL final volume). After incubation at 30 °C, the reaction was initiated by the addition of *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$ (i.e. containing R69H, N216W and L352M mutations) (1 µM). After 19h of reaction (30 °C, under magnetic stirring) the reaction mixture was freeze-dried. The lyophilized reaction mixture was puriried on C-18 column using the Reveleris flash chromatography System (Grace, Epernon, France), with elution by pure water allowing to separate α-A^2XX-F from α-X_3-F, ready to use by the glycosynthase.

4.5. Glycosynthase reactions

For the investigation of pathway 1, three reaction mixtures (500 µL) were prepared with 50 mM of substrate (A) α-X_2-F (7.1 mg), (B) α-X_3-F (10.4 mg) and (C) α-A^2XX-F (13.7 mg). Each reaction was prepared in sodium phosphate buffer (50 mM pH 7.0) and catalyzed by the glycosynthase XT6-E265G (0.5 mg.mL^{-1}). Following 24 h incubation at 25 °C (Thermomixer confort, Eppendorf) under agitation (1100 rpm), the xylo-oligosaccharides for (A) and (B) and arabinoxylo-oligosaccharides for (C) precipitated at the bottom of the tube. The precipitate was washed three consecutive times by centrifugation/resuspension in water cycle. Supernatant and pellet of each reaction were then lyophilized. Each sample was then resuspended in NaOH (0.2 M) for HPAEC-PAD to study the polymerization profile. Then, after freeze-drying and resuspension in D_2O, ^{1}H NMR analyses were also performed.

Regarding the investigation of pathway 2, “glycosynthase” reactions were monitored via ^{1}H NMR as described in 4.7.2 section.

4.6. Polymer hydrolysis

The insoluble fraction of reaction C (C_{pellet}), suspended in NaOH (0.2 M), was submitted to acid and enzymatic hydrolysis after ^{1}H NMR and HPAEC-PAD were performed. Regarding the enzymatic hydrolysis, the pH of an aliquot (100 µL) of sample C_{pellet} was adjusted to ~6 with HCl and *Tx*Abf wild-type (~ 1mg of protein/g of substrate) was added to the mixture (volume completed up to 200 µL) and the reaction was incubated at 45 °C, 1400 rpm (Thermomixer confort, Eppendorf), during 14 h. Reactions were stopped by heat denaturation at 95 °C for 5 min. Regarding acid hydrolysis, H_2SO_4 (2M, 100 µL) was added to sample C_{pellet} (100 µL). The mixture was incubated at 100 °C during 2 h in eppendorf tubes. The pH of both samples (enzymatic and acid hydrolysis) is made alkaline by addition of NaOH (5 M, ca. 30 µL) and stored at -20 °C till HPAEC-PAD analyses.

4.7. Analytics

4.7.1. NMR product characterization

Per-*O*-Acetyl α-L-Ara*f*-(1,2)-β-D-Glc*p*-(1,4)-Glc*p* ^{1}H NMR ($CDCl_3$, 500 MHz, 298K): δ = 6.28 (d, $J_{1,2}$ = 3.8 Hz, H-1 of $Glc^{I\alpha}$), 5.75 (d, $J_{1,2}$ = 8.4 Hz, H-1 of $Glc^{I\beta}$), 5.43 (dd, $J_{3,4}$ = 9.6 Hz, H-3 of $Glc^{I\alpha}$), 5.25 (dd, $J_{3,4}$ = 9.0 Hz, H-3 of $Glc^{I\beta}$), 5.15 (dd, $J_{3,4}$ = 9.5 Hz, H-3 of Glc^{II}), 5.13 and 5.10 (br. s, H-1 of Ara*f*$^{(Glc\text{-}I\alpha \text{ and } Glc\text{-}I\beta)}$) 5.08 (dd, $J_{2,3}$ = 9.8 Hz, H-2 of $Glc^{I\beta}$), 5.03 (dd, $J_{2,3}$ = 10.4 Hz, H-2 of $Glc^{I\alpha}$), 4.99-4.93 (m, H-4 of Glc^{II}), 4.94-4.93 (m, H-2 of Ara*f*), 4.91-4.88 (m, H-3 of Ara*f*), 4.56-4.52 (m, H-6a of $Glc^{I\alpha}$ and $Glc^{I\beta}$), 4.43-4.38 (m, H-5a of Ara*f*), 4.41-4.37 (m, H-6a of Glc^{II}), 4.36-4.33 (m, H-6b of $Glc^{I\alpha}$ and $Glc^{I\beta}$), 4.35 (dd, $J_{1,2}$ = 7.8 Hz, H-1 of Glc^{II}), 4.34-4.31 (m, H-4 of Ara*f*), 4.18 (ddd, *J* = 6.7 Hz, H-5b of Ara*f*), 4.08-4.05 (m, H-5 of $Glc^{I\alpha}$), 4.05-4.03 (m, H-6b of Glc^{II}), 3.93-3.87 (m, H-4 of $Glc^{I\beta}$), 3.88-3.84 (m, H-5 of $Glc^{I\beta}$), 3.86-3.84 (m, H-4 of $Glc^{I\alpha}$), 3.65-3.61 (m, H-5 of Glc^{II}), 3.60-3.56 (m, H-2 of Glc^{II}).

^{13}C NMR ($CDCl_3$, 125 MHz, 298K): 107.3-107.1 (C-1 of Ara*f*), 101.8-101.3 (C-1 of Glc^{II}), 91.6 (C-1 of $Glc^{I\beta}$), 89.0 (C-1 of $Glc^{I\alpha}$), 81.6-80.9 (C-4 of Ara*f*), 81.1 (C-2 of Ara*f*), 77.0 (C-2 of Glc^{II}), 76.9 (C-3 of Ara*f*), 75.0 (C-5 of $Glc^{I\beta}$), 74.7 (C-4 of $Glc^{I\beta}$), 73.4 (C-3 of Glc^{II}), 73.3 (C-4 of $Glc^{I\alpha}$), 72.3 (C-3 of $Glc^{I\beta}$), 71.8 (C-5 of Glc^{II}), 70.8 (C-5 of $Glc^{I\alpha}$), 70.3 (C-2 of $Glc^{I\beta}$), 69.3 (C-2 of $Glc^{I\alpha}$), 69.2 (C-3 of $Glc^{I\alpha}$), 68.4 (C-4 of Glc^{II}), 63.8-63.4 (C-5 of Ara*f*), 61.7 (C-6 of Glc^{II}), 61.5 and 61.3 (C-6 of $Glc^{I\alpha}$ and $Glc^{I\beta}$).

4.7.2. Transglycosylation time-course NMR monitoring

Reactions were prepared in 600 µL of buffered D_2O (final volume), containing donor and acceptor substrates when applicable and all employed enzymes were concentrated in D_2O (99%) beforehand. Deuterated acetate (Euriso-Top, France) was used to prepare a buffer displaying a pD value of 5.9, while deuterated phosphate was used to attain a value of pD 7.1. Deuterated sodium phosphate was prepared in-house by dissolving sodium phosphate in D_2O, followed by lyophilisation. This two-step protocol was repeated three times to achieve sufficient deuteration. Values of pD were measured by determining pH using a glass pH electrode and then applying the equation pD = pHelectrode + 0.41 (Glasoe & Long, 1960). For the cellobiose L-arabinofuranosylation reaction, the mixture was composed of *p*NP-α-L-Ara*f* (15 mM), cellobiose (30 mM) and *trans*-*Tx*Abf$^{(-1/+2 \text{ mutant})}$ (2 µM), incubated at 45 °C, in deuterated sodium phosphate buffer (25 mM, pD 7.1).

For the assessment of α-X_3-F recognition by R69H-N216W-L352M a transglycosylation reaction was carried out at 30 °C, in sodium acetate buffer (25 mM, pD 5.8) with *p*NP-α-L-Ara*f* (15 mM), α-X_3-F (30 mM) and *trans*-*Tx*Abf$^{(-1/+2 \text{ mutant})}$ (2 µM). Following overnight reaction a plateau phase was observed, marker of *Tx*Abf mutant inhibition. In the same reaction tube was then added XT6-E265G (0.5 mg.mL^{-1}) to monitor polymerization in real time during 18 h.

The polymerization reaction on purified α-A^{2}XX-F (pathway 2) was performed in deuterated sodium phosphate buffer (25 mM, pD 7.1) with α-A^{2}XX-F (8 mg) and XT6-E265G (0.5 mg.mL^{-1}), at 30 °C for 8 h.

^{1}H NMR spectra were accumulated semi-continuously over the reaction duration by accumulating a series of 5.52 min scan periods (i.e. 32 scans) that were interspersed by delay periods of 6 s. The exact full monitoring period was dependent on the enzyme employed. Each ^{1}H NMR spectrum was acquired using an excitation flip angle of 30° at a radiofrequency field of 29.7kHz, and the residual water signal was pre-saturated during the repetition delay (with a radiofrequency field of 21 Hz). The following acquisition parameters were used: relaxation delay (6 s), dummy scans (4). For each reaction, before enzyme addition, a ^{1}H NMR spectrum of the reaction mixture was acquired, serving as the starting point of the reaction, from which integrals were then corrected according to the small dilution factor induced by the enzyme addition (< 5% of total volume).

Data processing was performed as previously described (Bissaro et al., 2014). Briefly, the time-dependent evolution of donor (*p*NP-α-L-Ara*f*) and acceptor (xylotriose) concentrations were quantified by integrating the relative anomeric proton signals. Molar balances, based on acceptor and donor consumption, were used to convert the transglycosylation product signal integral into concentration, and to evaluate concentrations of hydrolysis products, respectively. From this, the donor substrate conversion rate (X in %) was calculated as the mean of three distinct signals; those of the *ortho* and *meta* aromatic protons (*p*NP) and that of the anomeric proton (L-Ara*f*), with X accounting for both reaction outcomes (i.e., hydrolysis and transglycosylation), and identifying self-condensation. For all time-course NMR kinetics, the absolute error mean value on X ranged from 1 to 3%. The transglycosylation yield, $Y_{(product)}$, was determined by relative integration of product anomeric proton signals. Relevant signals were those of H_{meta} (8.27 or 8.18 ppm, for linked and free state respectively) or H_{ortho} (7.24 or 6.98 ppm) from *p*NP, H-1 of *p*NP-α-L-Ara*f* (5.85 ppm), H-1-(C*H*F) of α-X$_3$-F or α-A^{2}XX-F (5.65 ppm, dd, $J_{1,2}$ = 2.8 Hz and $J_{H,F}$ = 53.3 Hz), H-1-(L-Ara*f*) (5.28 ppm) and H-1-(D-Xyl*p*) (4.55 ppm) of α-A^{2}XX-F or A^{2}XX.

4.7.3. HPAEC-PAD analysis

For HPAEC-PAD analysis, each sample was injected as such or diluted 10-fold in ultrapure water (MilliQ) and 20 µL was removed for injection on a Carbopac PA100 (4 x 250 mm) column, equipped with a Carbopac PA100 pre-column (4 x 50 mm) and mounted on an ICS-3000 system (Thermo scientific Dionex). For the analysis of soluble and insoluble fractions coming from polymerization reactions, the elution was performed at 1 mL.min^{-1} and 10 °C using the following eluants (A): NaOH (200 mM) and (B): sodium acetate (1 M) in NaOH (200 mM). A gradient from 0 to 60% of (B) was applied over 30 min followed by 100% (A) for 5 min. Regarding samples issued from polymer

hydrolysis, the following eluants were employed (A): NaOH (150 mM) and (B): sodium acetate (0.5 M) in NaOH (150 mM). A gradient from 0 to 80% of (B) was applied over 30 min followed by 100% (A) for 5 min. A gold probe was used for detection and data were recorded and analyzed using Chromeleon software (Dionex).

4.7.4. Mass spectrometry

High-resolution mass spectra (HRMS) were acquired at the CRMPO (Rennes University, France) using a Waters Q-Tof 2 or Q-Exactive spectrophotometer, equipped with an electrospray ionization source, operating in positive ion detection mode. ESI-HRMS for transarabinofuranosylation products: A^2X ($C_{15}H_{26}O_{13}$) *m/z* 437.127 ($M+Na^+$) calcd 437.1271, A^2XX ($C_{20}H_{34}O_{17}$) *m/z* 569.1689 ($M+Na^+$) calcd 569.16937, Per-*O*(Ac)-Ara-Glc_2 ($C_{37}H_{50}O_{25}$) *m/z* 917.2529 ($M+Na^+$) calcd 917.2533.

For polymerization reactions, MALDI analyses are currently on going.

References

Ajisaka, K. & Yamamoto, Y. (2002) Control of the regioselectivity in the enzymatic syntheses of oligosaccharides using glycosidases. *Trends Glycosci. Glycotechnol.*, 14 (75), pp.1–11.

Arab-Jaziri, F., Bissaro, B., Tellier, C., Dion, M., Fauré, R. & O'Donohue, M.J. (2014) Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 a-L-arabinofuranosidase using a directed evolution approach. *Submitted Results*.

Basu, N.G., Ghosal, P.K. & Thakur, S. (1986) Some structural features of an arabinoglucan from the fruits of *Cordia dichotoma* Forst. *Carbohydr. Res.*, 146, pp.350–351.

Ben-David, A., Bravman, T., Balazs, Y.S., Czjzek, M., Schomburg, D., Shoham, G. & Shoham, Y. (2007) Glycosynthase activity of *Geobacillus stearothermophilus* GH52 β-xylosidase: efficient synthesis of xylooligosaccharides from α-D-xylopyranosyl fluoride through a conjugated reaction. *ChemBioChem*, 8 (17), pp.2145–51.

Bissaro, B., Saurel, O., Arab-Jaziri, F., Saulnier, L., Milon, A., Tenkanen, M., Monsan, P., O'Donohue, M.J. & Fauré, R. (2014) Mutation of a pH-modulating residue in a GH51 α-L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products. *Biochim. Biophys. Acta Gen. Subj.*, 1840 (1), pp.626–636.

Bravman, T., Belakhov, V., Solomon, D., Shoham, G., Henrissat, B., Baasov, T. & Shoham, Y. (2003) Identification of the catalytic residues in family 52 glycoside hydrolase, a β-xylosidase from *Geobacillus stearothermophilus* T-6. *J. Biol. Chem.*, 278 (29), pp.26742–26749.

Breton, C., Fournel-Gigleux, S. & Palcic, M.M. (2012) Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.*, 22 (5), pp.540–549.

Broekaert, W.F., Courtin, C.M., Verbeke, K., Van de Wiele, T., Verstraete, W. & Delcour, J.A. (2011) Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides. *Crit. Rev. Food Sci. Nutr.*, 51 (2), pp.178–194.

Buranov, A.U. & Mazza, G. (2008) Lignin in straw of herbaceous crops. *Ind. Crop Prod.*, 28 (3), pp.237–259.

Champion, E., André, I., Moulis, C., Boutet, J., Descroix, K., Morel, S., Monsan, P., Mulard, L. a & Remaud-Siméon, M. (2009) Design of alpha-transglucosidases of controlled specificity for programmed chemoenzymatic synthesis of antigenic oligosaccharides. *J. Am. Chem. Soc.*, 131 (21), pp.7379–7389.

Chlubnová, I., Králová, B., Dvořáková, H., Hošek, P., Spiwok, V., Filipp, D., Nugier-Chauvin, C., Daniellou, R. & Ferrières, V. (2014) The versatile enzyme Araf51 allowed efficient synthesis of rare pathogen-related β-D-galactofuranosyl-pyranoside disaccharides. *Org. Biomol. Chem.*, 12 (19), pp.3080–3089.

Daudé, D., Remaud-Siméon, M. & André, I. (2012) Sucrose analogs: an attractive (bio)source for glycodiversification. *Nat. Prod. Rep.*, 29 (9), pp.945–960.

Dion, M., Nisole, A., Spangenberg, P., André, C., Glottin-Fleury, A., Mattes, R., Tellier, C. & Rabiller, C. (2001) Modulation of the regioselectivity of a *Bacillus* α-galactosidase by directed evolution. *Glycoconjugate J.*, 18 (3), pp.215–223.

Dion, M., Osanjo, G., André, C., Spangenberg, P., Rabiller, C. & Tellier, C. (2001) Identification by saturation mutagenesis of a single residue involved in the α-galactosidase AgaB regioselectivity. *Glycoconjugate J.*, 18 (6), pp.457–464.

Dumon, C., Song, L., Bozonnet, S., Fauré, R. & O'Donohue, M.J. (2012) Progress and future prospects for pentose-specific biocatalysts in biorefining. *Process Biochem.*, 47 (3), pp.346–357.

Ebringerová, A. (2005) Structural diversity and application potential of hemicelluloses. *Macromol. Symp.*, 232 (1), pp.1–12.

Fauré, R., Saura-Valls, M., Brumer, H., Planas, A., Cottaz, S. & Driguez, H. (2006) Synthesis of a library of xylogluco-oligosaccharides for active-site mapping of xyloglucan *endo*-transglycosylase. *J. Org. Chem.*, 71 (14), pp.5151–5161.

Feng, H.-Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C. & Dion, M. (2005) Converting a β-glycosidase into a β-transglycosidase by directed evolution. *J. Biol. Chem.*, 280 (44), pp.37088 –37097.

Ferré, H., Broberg, A., Duus, J.O. & Thomsen, K.K. (2000) A novel type of arabinoxylan arabinofuranohydrolase isolated from germinated barley analysis of substrate preference and specificity by nano-probe NMR. *Eur. J. Biochem.*, 267 (22), pp.6633–6641.

Fujiwara, T., Takeda, T. & Ogihara, Y. (1985) Synthesis of trisaccharides related to an arabinoglucan. *Carbohydr. Res.*, 141 (1), pp.168–171.

Gilbert, H.J. (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol.*, 153 (2), pp.444–455.

Glasoe, P.K. & Long, F.A. (1960) Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.*, 64, pp.188–190.

Hermann, B.G., Blok, K. & Patel, M.K. (2007) Producing bio-based bulk Chemicals using industrial biotechnology saves energy and combats climate change. *Environ. Sci. Technol.*, 41 (22), pp.7915–7921.

Ishihara, K., Katsube, Y., Kumazawa, N., Kuratani, M., Masuoka, N. & Nakajima, N. (2010) Enzymatic preparation of arbutin derivatives: lipase-catalyzed direct acylation without the need of vinyl ester as an acyl donor. *J. Biosci. Bioeng.*, 109 (6), pp.554–556.

Jensen, H.H. & Bols, M. (2006) Stereoelectronic substituent effects. *Acc. Chem. Res.*, 39 (4), pp.259–265.

Khasin, A., Alchanati, I. & Shoham, Y. (1993) Purification and characterization of a thermostable xylanase from *Bacillus stearothermophilus* T-6. *Appl. Environ. Microbiol.*, 59 (6), pp.1725–1730.

Koné, F.M.T., Le Béchec, M., Sine, J.-P., Dion, M. & Tellier, C. (2009) Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng. Des. Sel.*, 22 (1), pp.37 –44.

Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.*, 77, pp.521–555.

Lenucci, M.S., Piro, G. & Dalessandro, G. (2009) *In muro* feruloylation and oxidative coupling in monocots. *Plant Signal. Behav.*, 4 (3), pp.228–230.

Mackenzie, L.F., Wang, Q., Warren, R.A.J. & Withers, S.G. (1998) Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.*, 120 (22), pp.5583–5584.

Malet, C. & Planas, A. (1998) From beta-glucanase to beta-glucansynthase: glycosyl transfer to alpha-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS letters*, 440 (1-2), pp.208–212.

Mastihubová, M., Mastihuba, V., Bilaničová, D. & Boreková, M. (2006) Commercial enzyme preparations catalyse feruloylation of glycosides. *J. Mol. Catal. B: Enzym.*, 38 (1), pp.54–57.

Monsan, P., Remaud-Siméon, M. & André, I. (2010) Transglucosidases as efficient tools for oligosaccharide and glucoconjugate synthesis. *Curr. Opin. Microbiol.*, 13 (3), pp.293–300.

Mydock, L.K. & Demchenko, A. V (2010) Mechanism of chemical *O*-glycosylation: from early studies to recent discoveries. *Org. Biomol. Chem.*, 8 (3), pp.497–510.

Neyrinck, A.M., Possemiers S., Druart C., Van de Wiele, T., De Backer, F., Cani, P.D., Larondelle, Y. & Delzenne N.M. (2011) Prebiotic effects of wheat arabinoxylan related to the increase in *bifidobacteria, roseburia and bacteroides/prevotella* in diet-induced obese mice. *PLoS One*, 6 (6), p.e20944.

Niño-Medina, G., Carvajal-Millán, E., Rascon-Chu, A., Marquez-Escalante, J.A., Guerrero, V. & Salas-Muñoz, E. (2009) Feruloylated arabinoxylans and arabinoxylan gels: structure, sources and applications. *Phytochem. Rev.*, 9 (1), pp.111–120.

Noaman, E., Badr El-Din, N.K., Bibars, M.A., Abou Mossallam, A.A. & Ghoneum, M. (2008) Antioxidant potential by arabinoxylan rice bran, MGN-3/biobran, represents a mechanism for its oncostatic effect against murine solid Ehrlich carcinoma. *Cancer Lett.*, 268 (2), pp.348–359.

De O Buanafina, M.M. (2009) Feruloylation in grasses: current and future perspectives. *Molecular plant*, 2 (5), pp.861–872.

Osanjo, G., Dion, M., Drone, J., Solleux, C., Tran, V., Rabiller, C. & Tellier, C. (2007) Directed Evolution of the α-L-fucosidase from *Thermotoga maritima* into an α-L-transfucosidase. *Biochemistry*, 46 (4), pp.1022–1033.

Paës, G., Skov, L.K., O'Donohue, M.J., Rémond, C., Kastrup, J.S., Gajhede, M. & Mirza, O. (2008) The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry*, 47 (28), pp.7441–7451.

Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A. & Xiong, G. (2013) Hemicellulose biosynthesis. *Planta*, 238 (4), pp.627–642.

Pollet, A., Delcour, J. a & Courtin, C.M. (2010) Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. *Crit. Rev. Biotechnol.*, 30 (3), pp.176–191.

Rakotoarivonina, H., Hermant, B., Monthe, N. & Rémond, C. (2012) The hemicellulolytic enzyme arsenal of *Thermobacillus xylanilyticus* depends on the composition of biomass used for growth. *Microb. Cell Fact.*, 11 (159), pp.1–12.

Saeed, F., Pasha, I., Anjum, F.M. & Sultan, M.T. (2011) Arabinoxylans and arabinogalactans: a comprehensive treatise. *Crit. Rev. Food Sci. Nutr.*, 51 (5), pp.467–476.

Saulnier, L., Sado, P.-E., Branlard, G., Charmet, G. & Guillon, F. (2007) Wheat arabinoxylans: Exploiting variation in amount and composition to develop enhanced varieties. *J. Cereal Sci.*, 46 (3), pp.261–281.

Schmaltz, R.M., Hanson, S.R. & Wong, C.-H. (2011) Enzymes in the synthesis of glycoconjugates. *Chem. Rev.*, 111 (7), pp.4259–4307.

Topakas, E., Vafiadi, C. & Christakopoulos, P. (2007) Microbial production, characterization and applications of feruloyl esterases. *Process Biochem.*, 42 (4), pp.497–509.

Vafiadi, C., Topakas, E., Nahmias, V.R., Faulds, C.B. & Christakopoulos, P. (2009) Feruloyl esterase-catalysed synthesis of glycerol sinapate using ionic liquids mixtures. *J. Biotechnol.*, 139 (1), pp.124–129.

Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. & Etzler, M.E. (2009) *Essentials of Glycobiology*. *Cold Spring* Harbor Laboratory Press.

Vismeh, R., Lu, F., Chundawat, S.P.S., Humpula, J.F., Azarpira, A., Balan, V., Dale, B.E., Ralph, J. & Jones, a D. (2013) Profiling of diferulates (plant cell wall cross-linkers) using ultrahigh-performance liquid chromatography-tandem mass spectrometry. *Analyst*, 138 (21), pp.6683–6692.

Zhang, Y., Pitkänen, L., Douglade, J., Tenkanen, M., Remond, C. & Joly, C. (2011) Wheat bran arabinoxylans: Chemical structure and film properties of three isolated fractions. *Carbohydr. Polym.*, 86 (2), pp.852–859.

Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G. & Shoham, Y. (2004) Mapping glycoside hydrolase substrate subsites by isothermal titration calorimetry. *Proc. Natl. Acad. Sci. U. S. A.*, 101 (31), pp.11275–11280.

Supporting Information of Part V: "Towards Artifical Arabinoxylans".

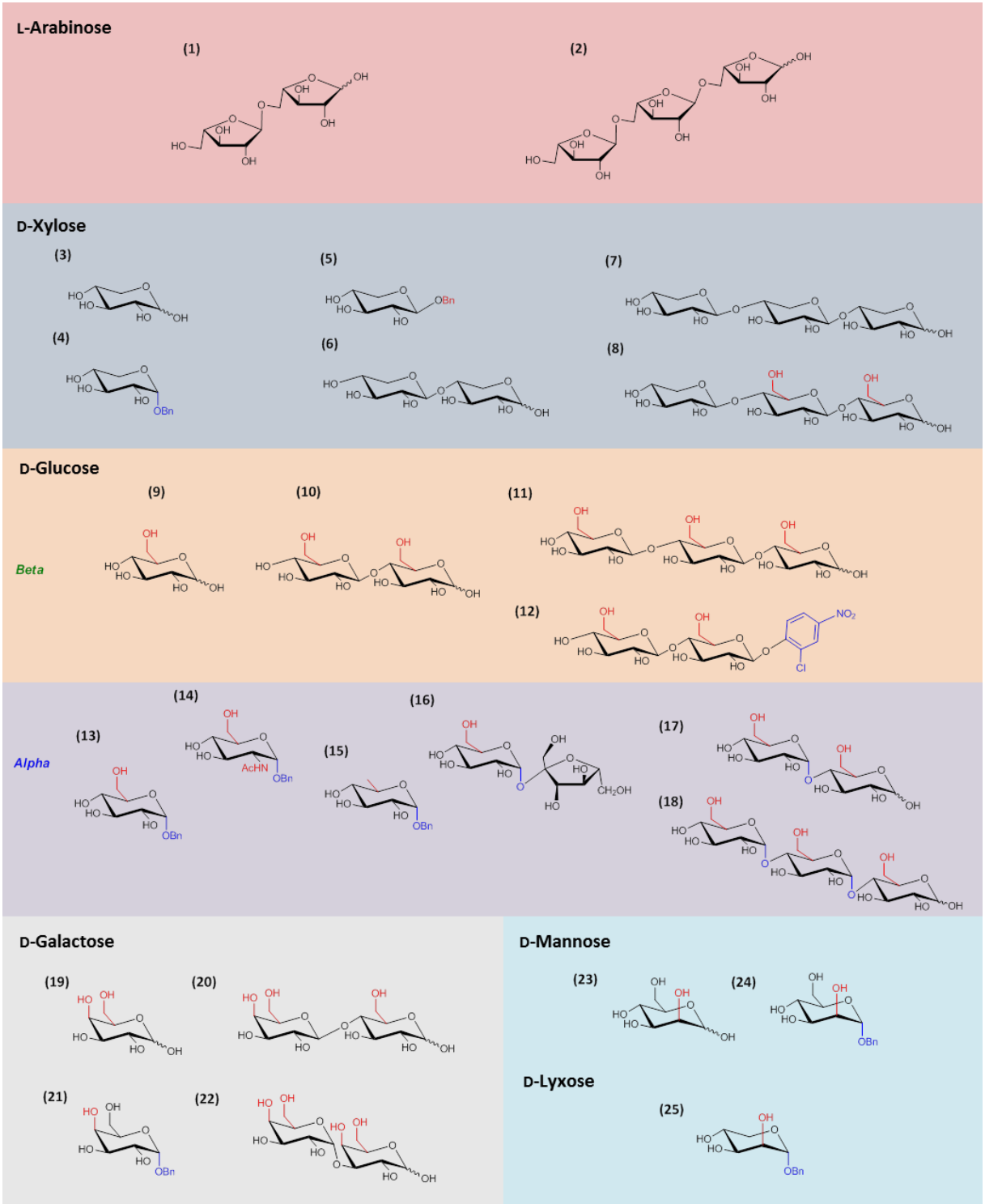


Fig. SI 1. Acceptor library screened to assess *Tx*Abf acceptor subsite plasticity

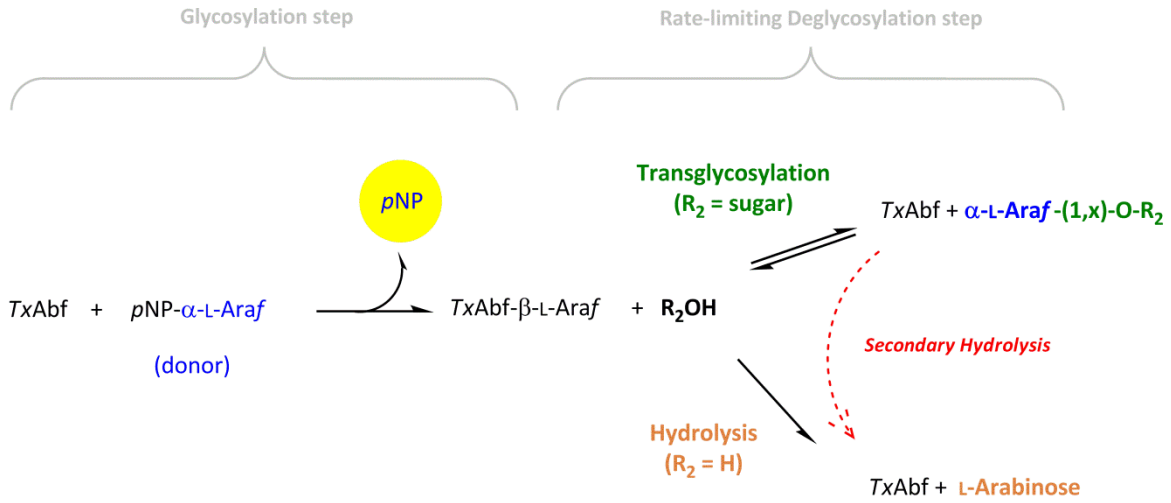


Fig. SI 2. *Tx*Abf catalytic mechanism proceeding through the two-step displacement of retaining GHs. In the course of the acceptor library screening (cf. Fig. 2) the *trans*-*Tx*Abf$^{(-1\ mutant)}$ was employed since displaying an acceptor-mediated activation phenotype (Part IV).

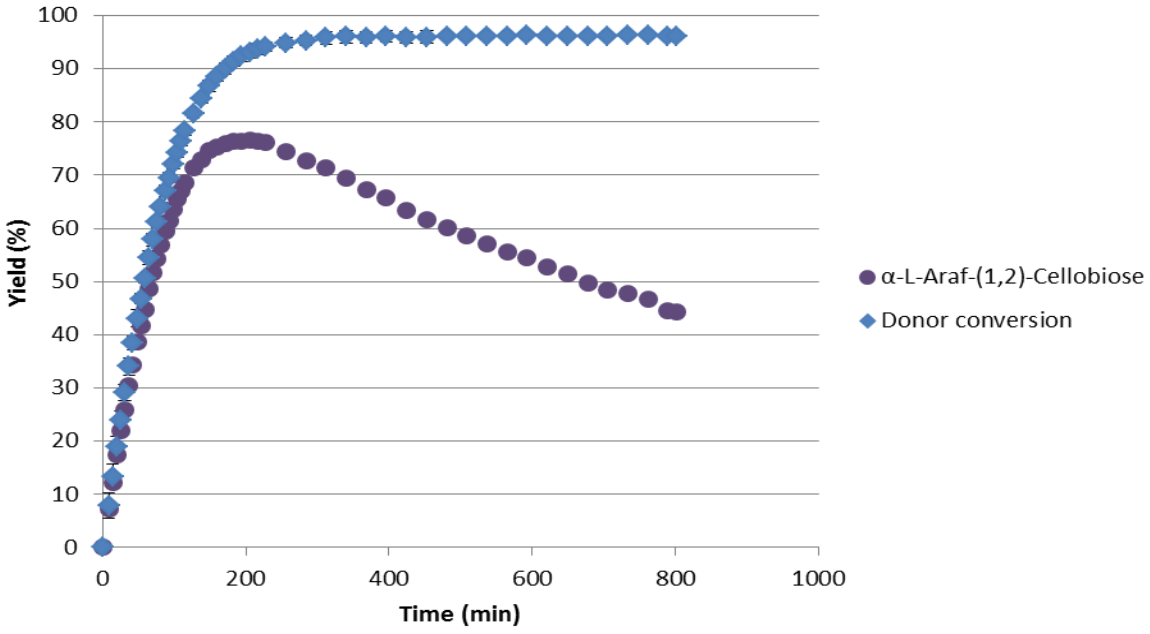


Fig. SI 3.Transglycosylation time-course NMR monitoring of reaction catalyzed by *trans*-*Tx*Abf$^{(-1/+2\ mutant)}$ (2 µM) in presence of *p*NP-α-L-Ara*f* as donor substrate (15 mM) and cellobiose as acceptor (30 mM). The reaction was performed at 45 °C in sodium phosphate buffer (25 mM, pD 7.1).

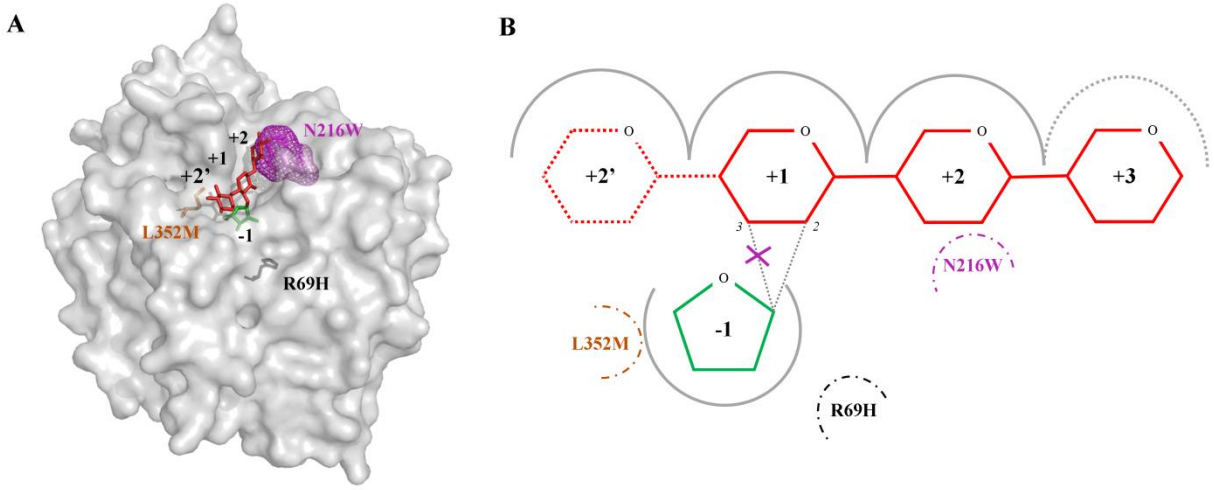


Fig. SI 4. Location of key mutations in -1 donor subsite (L352M and R69H) and +2 (N216W) acceptor subsites of *Tx*Abf. (A) 3D-view of co-crystallized *Tx*Abf-XA^{3}XX (PDB ID: 2VRQ) (Paës et al., 2008) X-ray structure in which mutations were introduced via *PyMol* software. One can observe a steric clash between N216W and +2 D-Xyl*p* unit, explaining (1,3) linkage deletion by N216W mutant. (B) 2D-schematic representation of *Tx*Abf active site showing the positioning of A^{2}XX product (solid line) with D-Xyl*p* units spanning from +1 to +3 subsites.

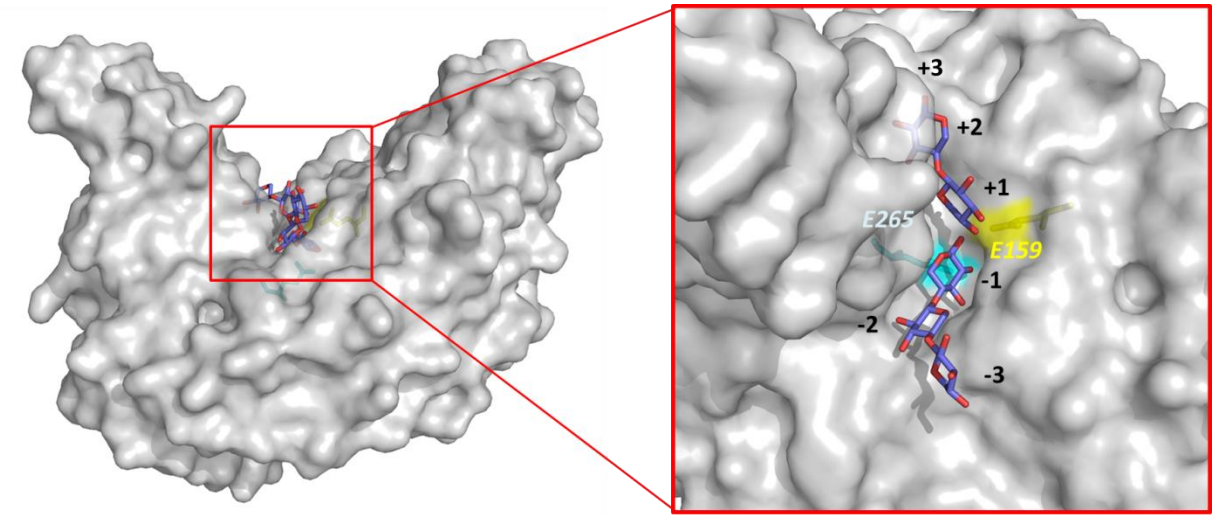


Fig. SI 5. 3D view of XT6-E265G active site (PDB1R87) in complex with xylopentaose (Zolotnitsky et al., 2004). E265 and E159 are the nucleophile and acid/base catalytic residues, respectively. It can be observed that the active site topoplogy of this GH10 would permit the accomodation of a α-(1,2)-linked L-Ara*f* moiety onto the non-reducing end D-Xyl*p* unit located either in the +1 or -3 subsites (if α-A^{2}XX-F acts as acceptor or donor respectively).

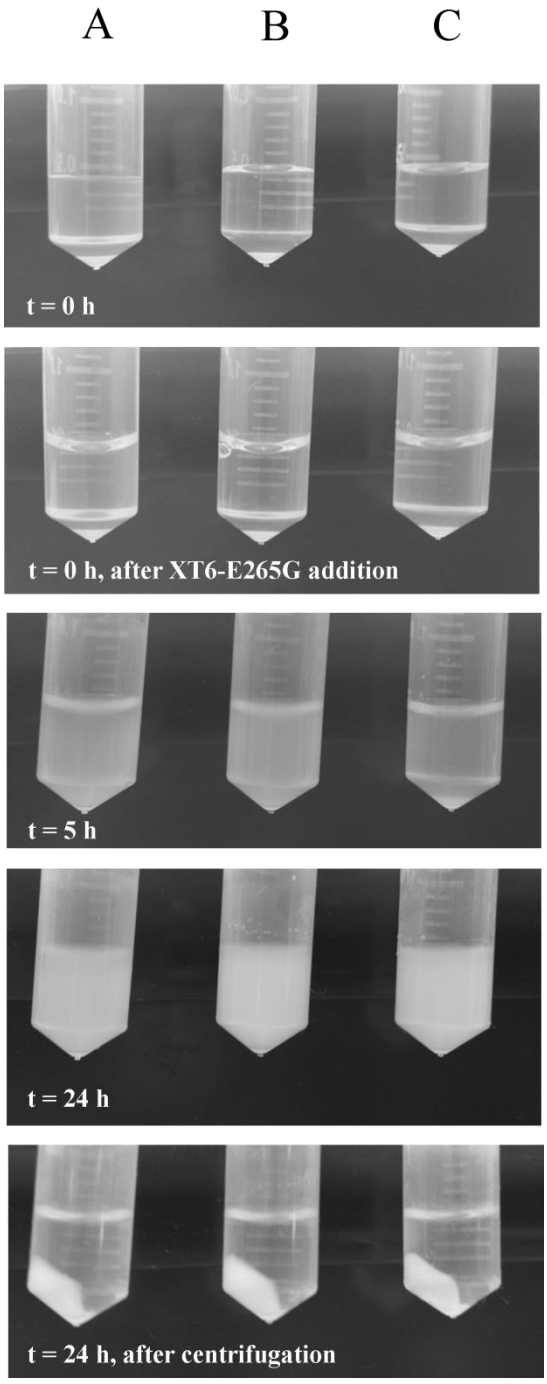


Fig. SI 6. Visual monitoring of XT6-E265G glycosynthase-catalyzed reaction using (A) α-X_2-F, (B) α-X_3-F and (C) not pure α-A^2XX-F donor substrates. The reaction with α-X_2-F substrate was realized as a control reaction since being the one described in the literature (Ben-David et al., 2007) and the one with α-X_3-F was employed to ensure that a backbone substrate composed of three D-Xyl*p* units would not hamper the glycosynthase reaction. Reactions were performed at 25 °C, in sodium phosphate buffer (pH 7.0, 50 mM) with XT6-E265G (0.5 $mg.mL^{-1}$) and substrate (50 mM).

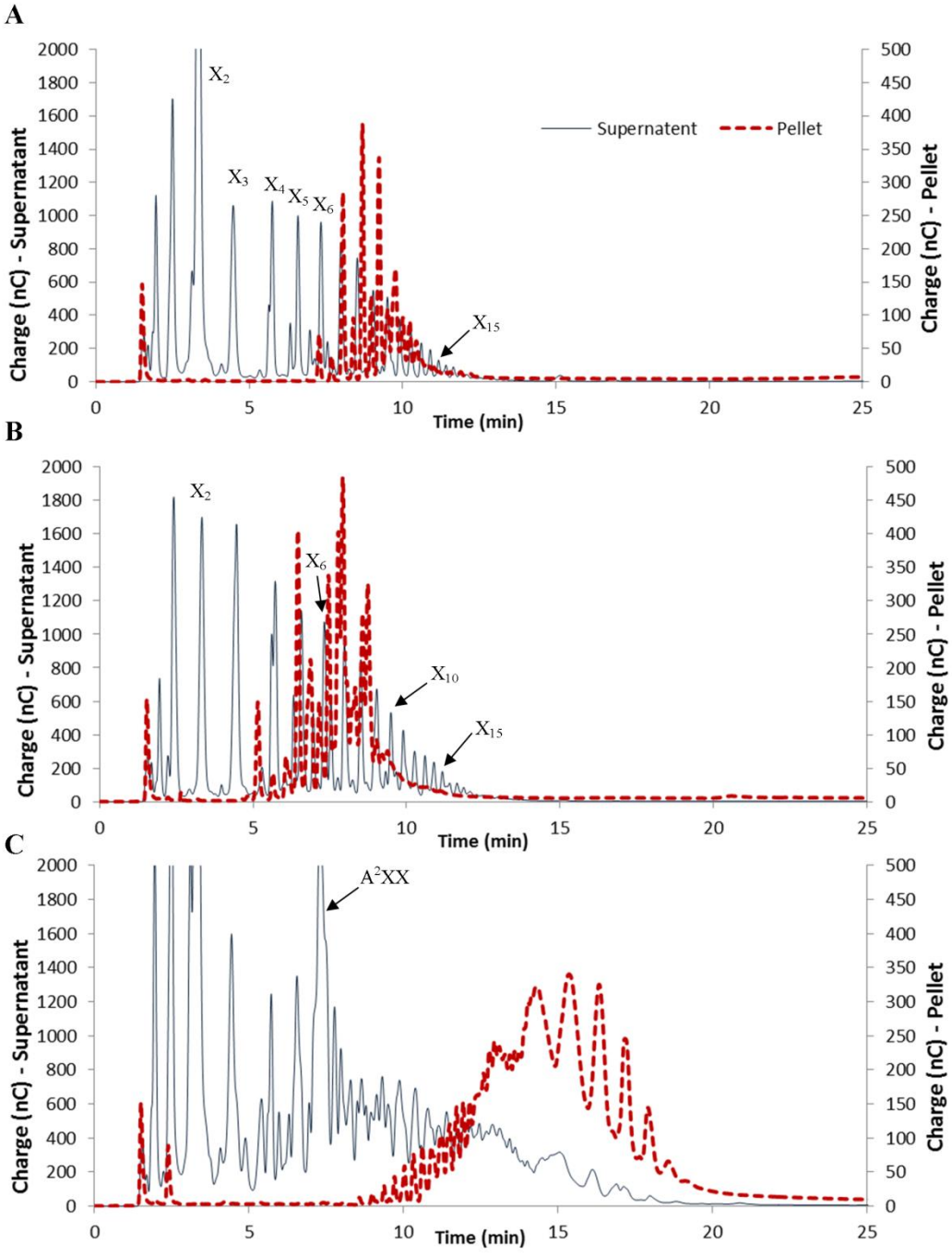


Fig. SI 7. HPAEC-PAD analysis of polymerization reactions catalyzed by XT6-E265G in presence of α-X_2-F (A), α-X_3-F (B) and α-A^2XX-F (C). Supernatants (solid dark blue) and pellets which were washed with water three times, were lyophilized and resuspended in NaOH (0.2 M) (dashed red line) prior to analysis.

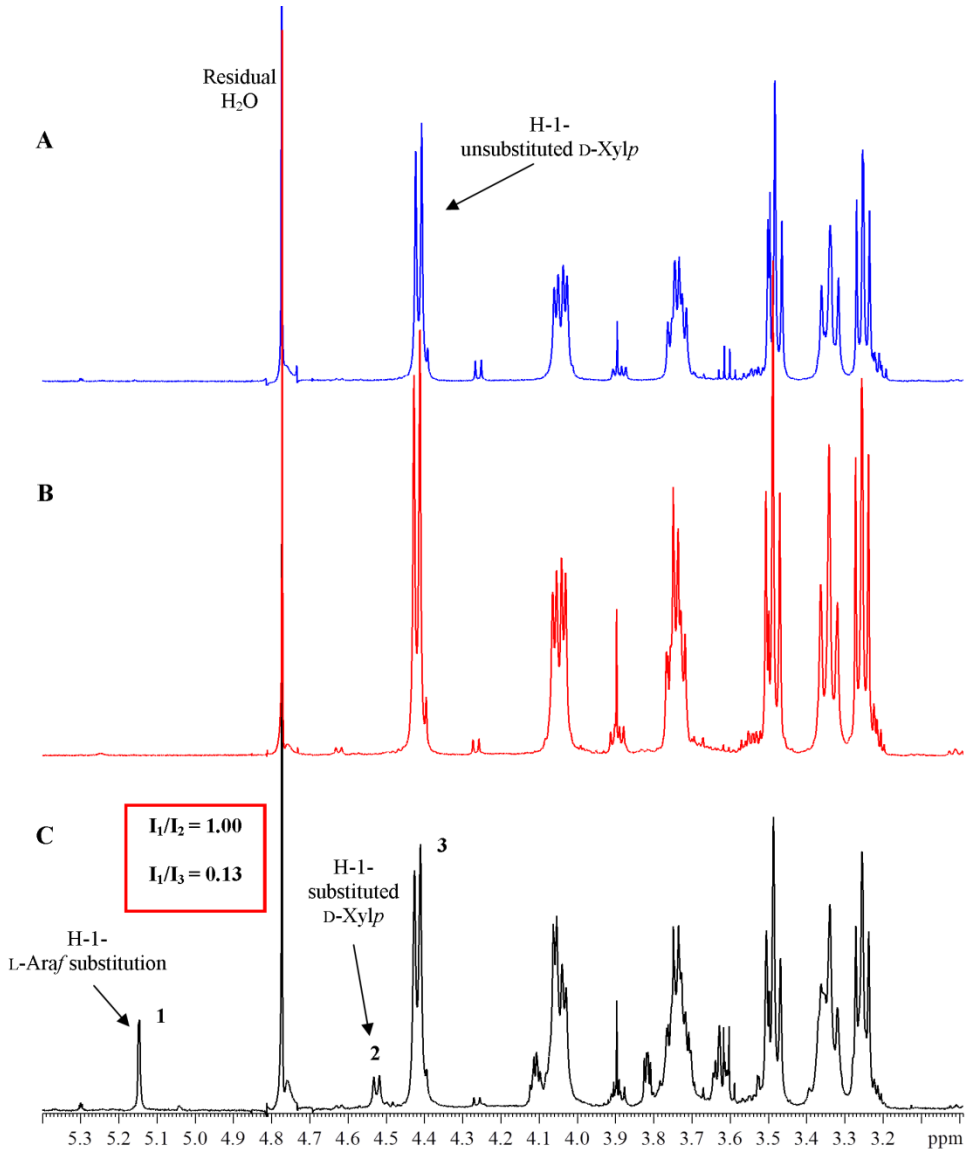


Fig. SI 8. ^{1}H NMR analysis of insoluble fraction (pellets) from polymerization reactions catalyzed by XT6-E265G in presence of α-X_2-F (A, blue), α-X_3-F (B, red) and α-A^2XX-F (C, black). Pellets which were resuspended in NaOH (0.2 M) for HPAEC-PAD analyses were freeze-dried and resuspended in D_2O for ^{1}H NMR analysis. Spectra were calibrated on residual H_2O signal (4.77 ppm at 25 °C) although the chemical shift is probably influenced by the presence of sodium hydroxide (NaOD, 0.2 M). Of note, the high basicity of the medium leads to almost unique β-anomery at the reducing end (the α-anomer is not observable). The L-Ara/D-Xyl ratio is equal to $1/(I_2/I_1 + I_3/I_1) = 0.11$, corresponding to one L-Ara*f* for nine D-Xyl*p* units.

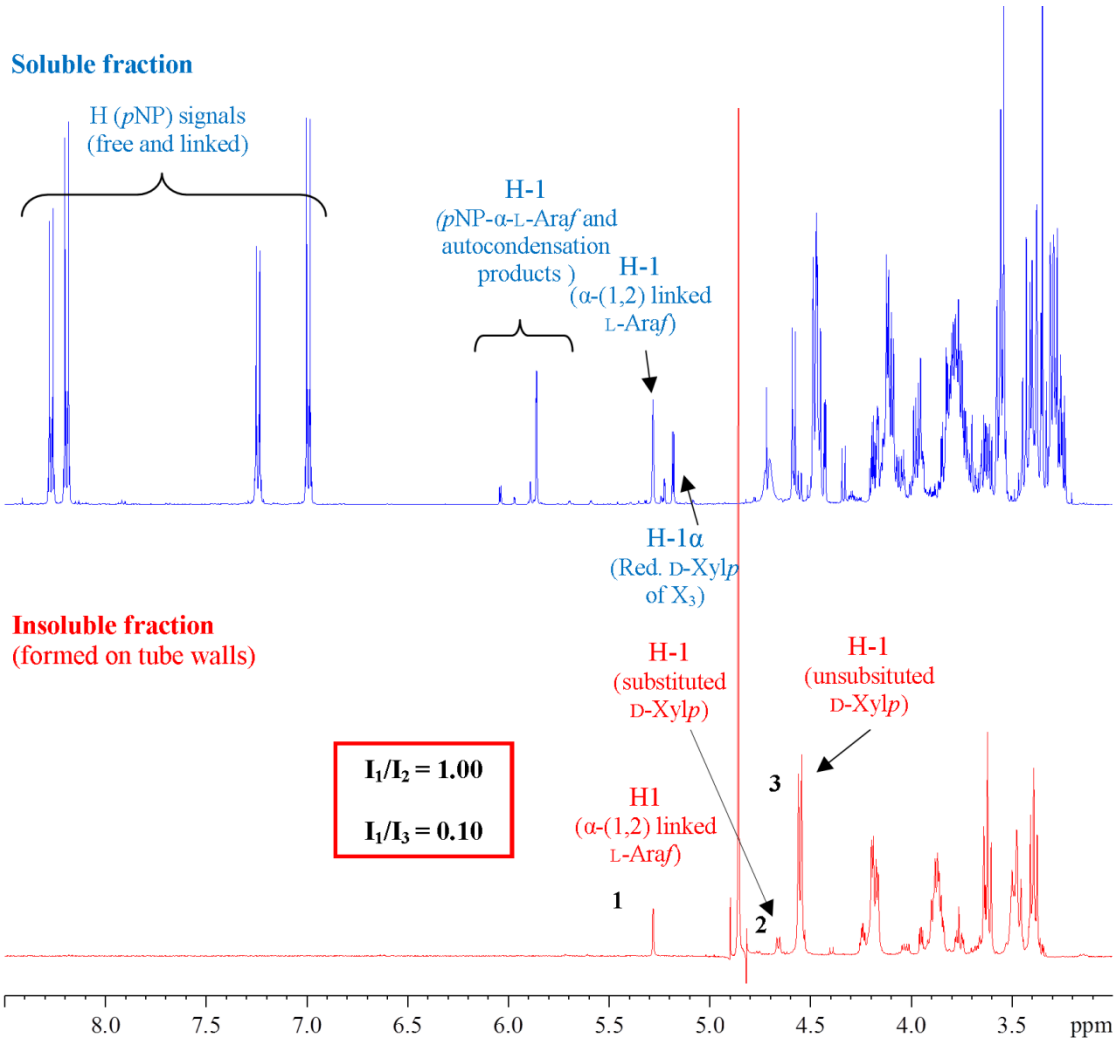


Fig. SI 9. ^{1}H NMR analysis of soluble and insoluble fractions issued from the "one pot" polymerization reaction (pathway 2a). The insoluble fraction, present as film formed on the NMR tube walls was resuspended in NaOD (0.2 M) after three D_2O "washing" steps. The soluble fraction represents the reaction mixture supernatant (sodium acetate buffer, pD 5.8, 25 mM). The L-Ara/D-Xyl ratio is equal to $1/(I_2/I_1 + I_3/I_1) = 0.09$, corresponding to one L-Ara*f* for eleven D-Xyl*p* units, on average.

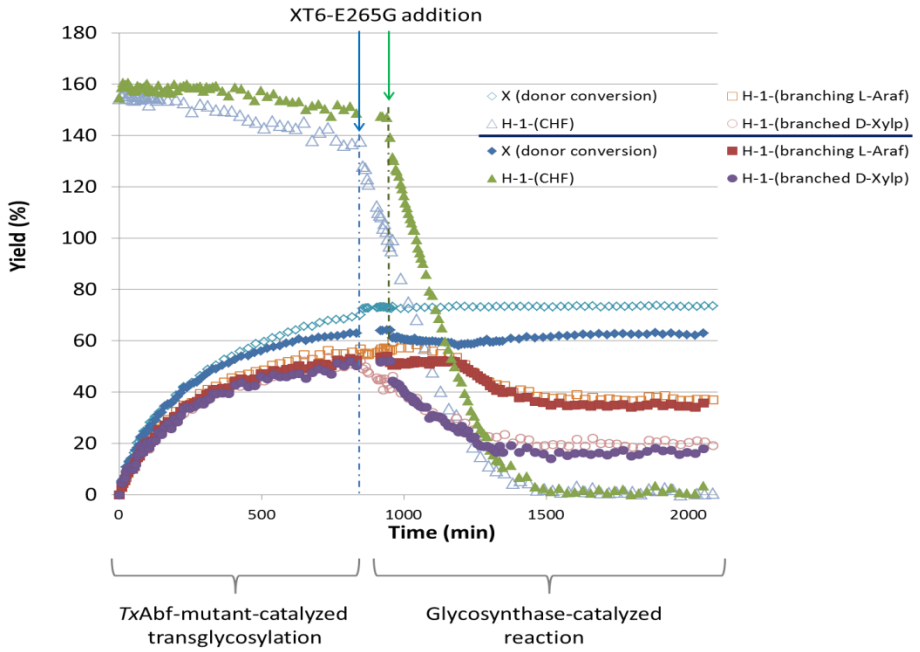


Fig. SI 10. Duplicate (open or closed symbols) of the "one-pot" two-steps polymerization reaction, including *(i) trans-Tx*Abf$^{(-1/+2\ mutant)}$-catalyzed transarabinofuranpsylation reaction (α-X_3-F as acceptor and *p*NP-α-Ara*f* as donor substrate), then, *(ii)* the GH10-derived glycosynthase XT6-E265G is added (0.5 mg.mL^{-1}) to perform polymerization. The reaction is carried out at 30 °C, in deuterated sodium acetate buffer (25 mM, pD 5.8).

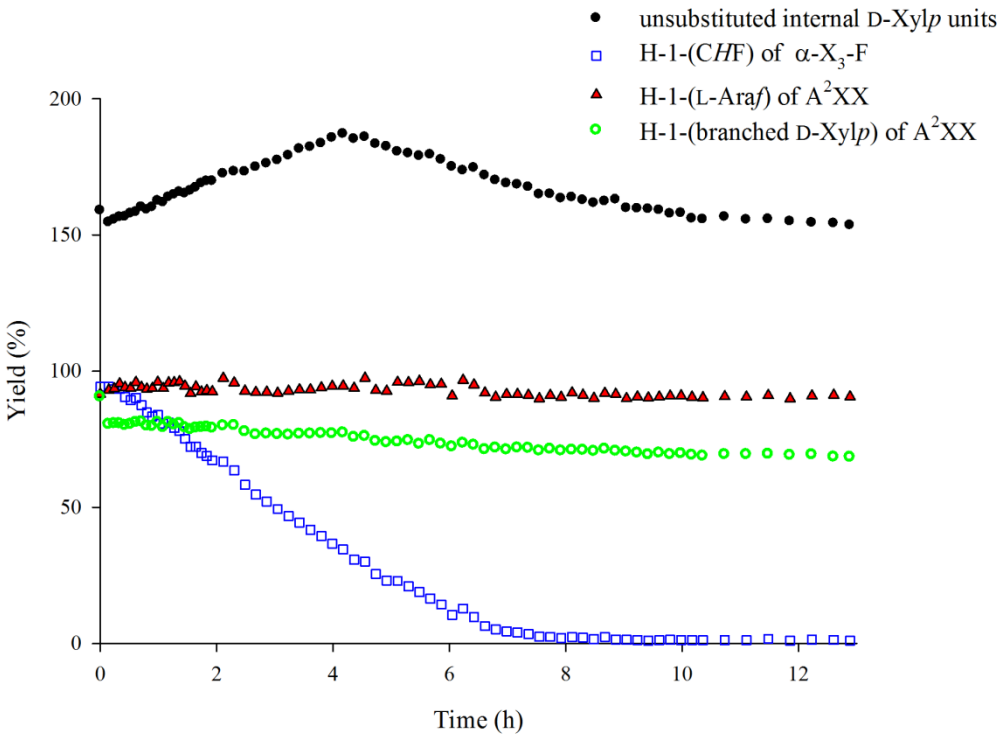


Fig. SI 11. Evaluation of the potential of A^{2}XX as an acceptor for XT6-E265G. The glycosynthase reaction was performed with α-X_3-F as donor (25 mM) and A^{2}XX as acceptor (25 mM), in deuterated sodium acetate buffer (50 mM, pD 5.8), at 30 °C. Values are expressed as % of 25 mM.

General Discussion

Conclusions,

and Perspectives

Global Discussion

The focal point of this doctoral work has been the creation of 'non-Leloir' transarabinofuranosylases. To achieve this, we have employed an iterative molecular design approach, using a hydrolytic GH51 α-L-arabinofuranosidase as the parental enzyme and relying on the progressive acquisition of detailed knowledge of the enzyme mechanism through the careful analysis of the different mutants that were generated. Gratifyingly, the findings of this study are generally consistent with an overarching hypothesis, which we have formulated using knowledge drawn from an extensive literature review on H/T modulation in GHs. Although the experimental pathway that led to the creation of a transarabinofuranosylase was not straightforward, some of the conclusions from this work are consistent with the ones described in another doctoral thesis written by David Tézé (Université de Nantes, 2012). In his thesis Dr. Tézé described the use of semi-rational enzyme engineering to increase the T/H ratio of a GH1 β-glycosidase (Teze et al., 2014).
Concerning the hypothesis that describes the key determinants of H/T modulation in retaining GHs, the main points are related to reaction conditions, regioselectivity and transfer capability. Each of these points merits discussion and clarification, thus this is the aim of the following sections.

On reaction conditions: competition between the donor substrate and the product of transglycosylation

During this study, we deliberately chose not to focus on the alteration of reaction conditions *per se* (e.g. large excess of acceptor or use of co-solvent). This is because our objective was to prove that the H/T ratio can be controlled by acting on the molecular determinants of this equilibrium. Nevertheless, in the course of our investigation we were obliged to acknowledge that higher donor substrate concentrations (from 5 to 15 mM) clearly delay the secondary hydrolysis of the transglycosylation product, leading to yields multiplied by a factor of 2 (Part II). Coupling this fact to the observation that transfer rates are not modified, demonstrates that competition occurs between the donor substrate, *p*NP-α-L-Ara*f*, and the products of transglycosylation, the latter being good substrates once a [*p*NP-α-L-Ara*f*] threshold has been reached. This critical level is around 80-90% of donor conversion, although in certain mutants this threshold was higher or slightly lower. Overall, the results obtained rather clearly demonstrate that secondary hydrolysis is governed by thermodynamic equilibrium constants, and that the K_d of the donor substrate can be tuned relative to that of the transglycosylation products (and *vice versa*) through mutation. Accordingly, it was proposed that when mutations favor the K_d of the donor substrate relative to that of the transglycosylation product(s), secondary hydrolysis could be delayed. This phenomenon was clearly

pinpointed by the *in silico* study that was performed using the *BindScan* protocol. In this study, we set out to improve the recognition of xylotriose, by introducing mutations at position G179, which is located in subsite + 1. When an aromatic residue was introduced at position 179, predictably xylotriose recognition was increased, but the affinity for the *p*NP moiety (in subsite + 1) of *p*NP-α-L-Ara*f* was also increased, which leads to better affinity of the donor substrate. In hindsight this collateral effect is predictable and thus might be an original strategy to increase the accumulation of transglycosylation products and thus improve the final yield of transglycosylation reactions.

On the creation of regiospecific reactions

From the point of view of glycosynthesis, the relatively low selectivity of *exo*-acting GHs can be seen as drawback, since it is well known that a relationship can be established between hydrolytic selectivity and transglycosylation regiospecificity (Malá et al., 1999). The control of this parameter in GHs is still a matter of progress and computational methods will no doubt be of value in this respect. Regarding *Tx*Abf, it has been described as an *exo*-acting hydrolytic enzyme, removing α-(1,2) or α-(1,3) linked L-arabinofuranosyl substituents from the xylan backbone of arabinoxylans (Debeche et al., 2000). In agreement with the previous remark about hydrolytic selectivity and transglycosylation regiospecificity, we have shown that when *Tx*Abf operates in transglycosylation mode, several products are obtained, including XA^3X, A^3XX, A^2XX and traces of XA^2X (when the wild-type enzyme is employed). This reveals a degree of plasticity within the acceptor subsites of *Tx*Abf, although some configurations are probably much more productive, with A^2XX being the major product. In the framework of this doctoral work, the acceptor subsite plasticity of *Tx*Abf was further probed by screening an acceptor library, revealing that D-glucose-based compounds can be accommodated in the acceptor subsites (Part V). Additionally, during our work, regioselectivity was modulated using mutagenesis. To achieve this, a mutant of *Tx*Abf (N216W) was created that can perform regiospecific (1→2) reactions, the molecular basis of this effect being the introduction of steric hindrances in subsite +2 (Part I and III). Compared to mutations in subsite -1, more distal modifications generate only weak impacts on catalysis (i.e. in the case of N216W the hydrolytic catalytic constant is only weakly affected), revealing that the main effect of such mutations is the modification of acceptor positioning, rather than the modification of access to water molecules. Unfortunately, in this study we failed to identify any mutations that procure higher selectivity for the formation of (1→3) linkages, while lowering selectivity for the formation of (1→2) linkages. In this respect, it is noteworthy that *BindScan* analyses failed to identify any "hot-spots" that would be specific for the nucleophilic attack of a bound β-L-Ara*f* anomeric carbon by the *O*-3 of the non-reducing or central

unit of xylotriose. Overall, on the basis of the results obtained with *Tx*Abf, regioselectivity appears to be essentially driven by steric hindrances and stacking effects with sugar rings.

On the improvement of transfer capability

An early, highly significant observation (Rémond et al., 2005) was that the transglycosylation reaction catalyzed by *Tx*Abf was significantly enhanced when *p*NP-β-D-Gal*f* (a 5-hydroxymethyl analog of *p*NP-α-L-Ara*f*) was employed as a donor substrate. The underlying reasons for this remained elusive though it was quite clear that the interactions at *O*-5 position were involved in the alteration of this activity, since hydrolysis is also affected (Euzen et al., 2005). The main hypothesis was that the extra hydroxymethyl would provoke a modification of the position of the anomeric carbon of the covalently bound α-D-galactofuranosyl moiety relative to the catalytic amino acids, thus altering the transfer of the latter onto acceptors in a manner that would penalize water and/or favor glycoside acceptors. Even though this explanation remained unsatisfactory, the experimental work provided some clues concerning how the T/H ratio might be increased. Indeed, the K_M value for *p*NP-β-D-Gal*f* is described as > 50 mM while the k_{cat}/K_M is around 0.56 $s^{-1}mM^{-1}$ (at 60 °C, pH 5.8) showing that the catalytic efficiency needs to be drastically decreased, an effect that is achieved through the simultaneous increase of the K_M value and a decrease of the value of k_{cat}. Whereas the decrease in k_{cat} reflects the fact that the rate of water-mediated deglycosylation is lowered, the increase in K_M is probably more the result of a general impact on the overall reaction (rather than a simple indication of affinity changes as often stated). In this respect, it is Interesting to note that the first randomly-generated mutation (L352M) that was found (c.f. SI A3) to significantly increase the transfer rate, actually provoked a 60-fold increase in the K_M value on *p*NP-α-L-Ara*f* with the value of k_{cat}/K_M being 1.3% that of the parental enzyme. Therefore, this result confirms the nature of the phenotype that should be searched for. Surprisingly, the amino acid side chain at position 352 is located at the bottom of pocket-like subsite -1, just in front of the *O*-5 of the L-Ara*f* unit (Fig. 2, Part III). Therefore, it is tempting to speculate that the impact of the mutation L352M in some ways reflects the impact of using *p*NP-β-D-Gal*f* as a donor in the place *p*NP-α-L-Ara*f*. Overall, it appears that alterations that are in the vicinity of the *O*-5 position of the donor sugars will affect the ratio H/T partition. Combining the effect of L352M with that of the mutation R69H procured an even higher T/H ratio, which was achieved at the expense of global catalytic efficiency. Interestingly, R69 is not only totally conserved among GH51 Abfs, but it is also one of the eight key residues that define the GH clan A (Durand et al., 1997), which is composed of 19 GH families[9]. Consequently, it is plausible

[9] On the 08th of July 2014, 19 GH families belongs to Clan A: GH 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113 and 128.

that the introduction of mutations at analogous positions in any clan GH-A member might procure a similar effect. If this is the case, this would constitute a new strategy to modulate the H/T partition in an important subgroup of GHs.

More generally, one important lesson that can be drawn from the molecular evolution work is that decreasing both the k_{cat} and k_{cat}/K_M values is a prerequisite to increase the ability of a GH to perform transglycosylation, since this implies that water-mediated deglycosylation is hampered. This point is clearly illustrated by beneficial (i.e. reinforced transglycosylation), acceptor subsite mutations (e.g., G179F or N216W) that increase acceptor recognition, and consequently the T/H ratio, while procuring only modest overall yields of transglycosylation products when employed as single mutations. When this type of mutation is combined with mutations that cripple catalytic activity (e.g. L352M or R69H-L352M) they reveal their full potential, since a fairer competition (from an energy standpoint) between water and sugars is established, and the resultant enzymes are fully-fledged transarabinofuranosylases. Briefly, considering the artificial evolution *Tx*Abf in a more global context of natural evolution of GHs, it is likely that *exo*-acting GHs have acquired the ability to establish an extensive network of enzyme-substrate interactions in the donor subsite, which provides strong transition state stabilization. The implications of this, for *exo*-acting GHs, are that they display:

(i) elevated catalytic efficiencies (k_{cat}/K_M values), reflecting the glycosylation step (Michaelis-Menten complex formation followed by LG departure and covalent intermediate formation) and

(ii) quite rapid, albeit slower than glycosylation, water-mediated deglycosylation of the enzyme, since the major part of the energy required to overcome the transition state of this step is provided by donor subsite interactions.

This latter observation is also consistent with the rate-limiting character of the deglycosylation step, which in the case of hydrolysis does not involve the establishment of interactions in the acceptor subsites. These remarks constitute the corner stone of the whole reasoning, as developed in the literature review presented at the beginning of the manuscript (Chapter I, Part I-3). Indeed, the ease with which water can deglycosylate the glycosyl-enzyme intermediate means that:

(1) hydrolytic GHs are efficient biocatalysts because they operate in conditions in which the concentration of water (55.5 M maximum) far outweighs that of any other acceptor (e.g. 30 mM of acceptor versus 55,500 mM of water molecules).

(2) transition state destabilization is a prerequisite to make the energetic pathway more difficult for water, leading to lower k_{cat}/K_M and k_{cat} (glycosylation and deglycosylation transition states are interdependent). In these conditions, the energy required to overcome TS2 is higher and thus

water no longer benefits from the advantage procured by its overwhelming concentration, especially when interactions in the acceptor region favor acceptor binding.

(3) GHs that mainly perform synthesis are so slow enzymes, as reflected by hydrolytic k_{cat}/K_M values being 85-fold (GH1 β-glycosidase) to 1165-fold (sucrose-acting GH13) lower than their hydrolytic counterparts. The access to TS energy level values is currently a matter of study, notably through computational approaches (Pierdominici-Sottile et al., 2011; Pierdominici-Sottile et al., 2013; Wang et al., 2011; Wang et al., 2013; Badieyan et al., 2012), but it is foreseeable that for GHs performing glycosynthesis these values should be much higher than for hydrolytic ones (ca. 20-25 kcal.mol^{-1}).

Conclusions

Conversion of GHs into transglycosylases

Apart from the development of glycosynthases and related engineering strategies, relatively few GHs (that do not possess natural transglycosylating counterpart) have so far been submitted to artificial molecular evolution with the aim of converting them into transglycosidases. Indeed, to the best of our knowledge, only GHs from families 1 (β-glycosidases,(Kim et al., 2004; Feng et al., 2005)), 29 (α-L-fucosidases, (Osanjo et al., 2007)) and 36 (β-galactosidase, (Placier et al., 2009)) have been engineered in this manner. Until now, no attempts have been made to engineer transglycosylase activity into an arabinofuranosidase (or any other furanose-acting enzyme), although much of the earlier work performed by our group on *Tx*Abf was done with this intention in mind. Therefore, at this point in time it is legitimate to consider our findings as watershed results that open the way towards the development of specific enzymes for the synthesis of glycofuranose-containing compounds.

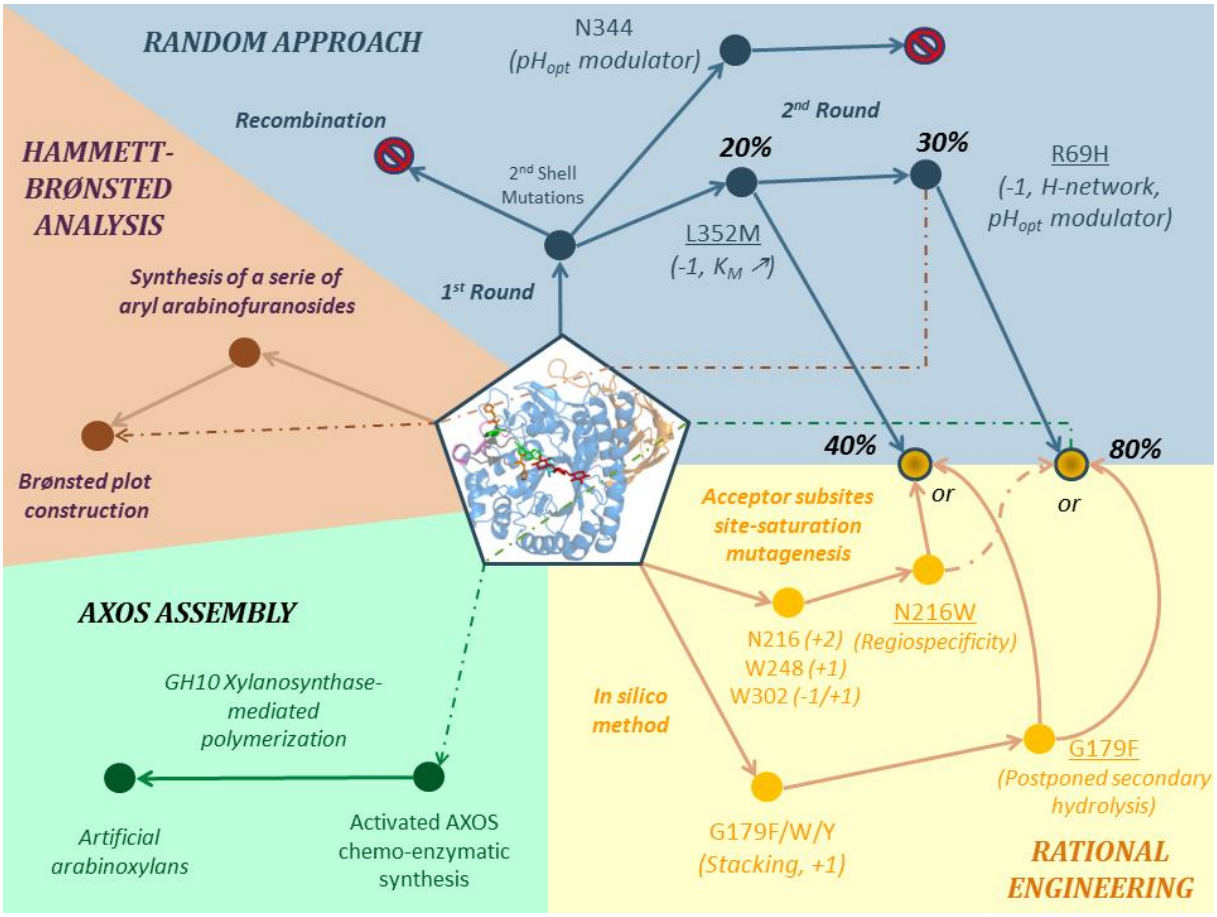


Fig. 1. *Tx*Abf Story

The evolutionary process that is reported herein was far from being smooth and took us rather off the proverbial beaten track. Indeed, the somewhat chaotic progression along the evolutionary

pathway might be better likened to a walk through a dense jungle, with some dead ends being encountered along the way. The major reason for this was a severe lack of initial knowledge pertaining to the molecular determinants of the H/T ratio. Consequently, it was necessary to simultaneously employ several approaches that together revealed interesting results and provided the possibility to combine beneficial mutations that together displayed highly synergistic effects (Fig. 1).

This iterative approach, which was developed over approximately 5 years, has provided us with a *modus operandi*, identifying key targets for mutation and the way in which to combine these, which might be generally applicable for the engineering of other GHs. Although this remains to be demonstrated, it seems reasonable to anticipate that the lessons learned in this work will greatly assist other enzyme engineers in their quest to modify GHs. The first of these lessons is the absolute necessity to decrease water's ability to deglycosylate the covalent glycosyl-enzyme, despite its overwhelming presence in aqueous solutions. This implies the destabilization of transition states (of glycosylation and deglycosylation steps), which can be achieved by the introduction of mutations in and/or around the donor subsite and nucleophile catalytic residue. In this respect, prudence is required, because some such mutations will definitively cripple GHs, thus the precise identification of amino acids that can be altered to achieve the desired effect remains a challenge. For these reasons, random mutagenesis remains a useful approach to tackle this phase of the process. However, to make library screening less time-consuming, in the future it might be pertinent to only mutate a sub-region of the target enzyme, around its catalytic core, assuming that sufficient 3D data is available to do this. Nevertheless, the identification of R69 is at least one significant step towards a more rational way to diminish water-mediated deglycosylation. This residue is highly conserved among clan A, and thus might represent an interesting target for mutagenesis. Regarding the improvement of acceptor-mediated deglycosylation step, *in silico* and/or semi-rational approaches will be attractive ways to improve acceptor recognition and guide acceptor positioning in the active site and/or control the regioselectivity.

Tailor-made furanosylated compounds design

Beyond the conversion of a FH into a 'non-Leloir' transarabinofuranosylase, a significant challenge in this research project was to develop methods to precisely synthesize components of natural pentose-based polysaccharides, for example arabinose-substituted xylooligosaccharides, or alternatively to design furanose/pentose-based glycomotifs that do not necessarily have natural counterparts. To achieve this, the first "glycobricks" that we have developed are AXOS, arabinoxylan motifs that are widespread in plant cell walls. The potential applications of these synthetic AXOS are numerous and include their use as prebiotic compounds, or as building blocks for the synthesis of

bio-inspired materials. Moreover, the synthetic AXOS could prove to be useful as research tools, allowing the raising of highly specific antibodies or for the characterization (e.g. structure-function and mechanistic studies) of enzymes such as Abfs, xylosidases, xylanases or even GTs involved in the biosynthesis of arabinoxylans. For the latter purpose, and also because the physiological or physicochemical features of AXOS depend on their degree of polymerization and on the exact nature and density of the substitution, the development of a whole range of chemo-enzymatic reactions is necessary, providing a complete toolbox for the synthesis of made-to-measure oligosaccharides. For this, it seems plausible that *Tx*Abf could be used as a robust protein scaffold for the preparation of several different transarabinofuranosylases displaying for example different regioselectivities. In this respect, our work already revealed that it is possible to create an enzyme that can synthesize arabinoglucooligosaccharides, but further work will no doubt open access to tailored galactofurano-oligosaccharides.

Like all good stories, it is always preferable to end on a happy note. In the final chapter of this doctoral study, we describe how the action of synthetic transarabinofuranosylases can be combined with that of a xylanosynthase to assemble small glycobricks into a larger polymeric structure. To our knowledge these results constitute the first description of the chemo-enzymatic synthesis of artificial arabinoxylans. Looking into the future, it is likely that it will be unnecessary to use xylanosynthases, since the knowledge gained in this work will open the way to the design of efficient transxylanosylases, thus obviating the need for fluoride-activated donors.

Perspectives

This investigation has raised many fundamental questions regarding the mode of action of *Tx*Abf and mutants thereof, but unfortunately this work did not provide all of the answers. Hereafter, some of the key questions are discussed. One of these concerns carbohydrate-protein interactions, while the other is all about the intrinsic source of the power of enzyme catalysis.

Deeper investigation of protein-ligand interactions

What is the role of TxAbf global dynamic into inhibition phenomena?

NMR is a powerful method to study proteins, but this methodology has not yet been deployed to study *Tx*Abf. This is mainly because the use of NMR to study protein structure was for a long time limited to smaller proteins (The Mw of *Tx*Abf is ~57 kDa). However, this limitation might be no longer valid, thanks to recent progresses in the field (Mainz et al., 2013; Frueh, 2014). If NMR protein studies were used on *Tx*Abf, these could bring considerable new insights into several fundamental questions. Among these, the study carried out on N344 position revealed that N344Y or P mutations introduced a pH-dependent inhibition by the leaving group (aromatic ring), which is actually beneficial because it prevents hydrolysis of the transglycosylation product. However, in spite of extensive studies, we were unable to unravel the underlying factors that are responsible for this phenomenon, except that the proton sharing system has been affected. In Bissaro *et al.* (2014), we postulated that the inhibition might involve some interference of pH-dependent enzyme dynamics, notably loop movements, since such as phenomenon has already been observed for certain α-amylases (Nahoum et al., 2000). Therefore, it would be really interesting to better characterize the β2α2 loop movement, a structural feature that was previously shown to be relevant for function (Paës et al., 2008; Arab-Jaziri et al., 2012). Using protein NMR analysis to monitor the loop movement, it would be possible to assess the impact of the N344 or R69 mutations and observe the changes on key residues, such as W99 and H98 (located at β2α2 loop end), upon substrate binding. Moreover, the use of NMR to monitor the protonation states of certain residues (e.g. $H^{\varepsilon 1}$ of H98) during pH titration should provide information on the effect of different mutations and the presence of aromatic rings (e.g. *p*NP identified as inhibitor). In this way, it might be possible to more accurately define the inhibition and embark upon new structure-function studies using site-directed mutagenesis. However, to achieve such goals it will first be necessary to establish that relevant proton signals can be well isolated and characterized.

What are the relative impacts of acceptor reactivity and affinity on H/T modulation?

In order to further our understanding of how the H/T modulation is established in GHs, it would be useful to delineate the exact contributions of acceptor reactivity and affinity to the outcome of

deglycosylation. For this, a first step would be to trap the covalent intermediate through the reaction of *Tx*Abf with 2,4-dNP-2-fluoro-2-deoxy-arabinofuranosyl substrate, for which the chemical synthesis is far from being trivial. Then, to build on the work that was initiated with the *BindScan* study, the experimental characterization of enzyme-acceptor interactions, such as "stacking" effects (from a substrate point of view using STD NMR methods for instance) might shed light on these questions. A comparative study could be then conducted, using modified acceptor substrates displaying for example thiol functions (instead of hydroxyls), which would increase their reactivity. The correlation of the results of this type of analysis with the ability to perform transglycosylation might provide information on the relative importance of acceptor reactivity and affinity. In this respect, it would be very interesting to measure the impact on the T/H ratio of mutations in subsite -1 coupled to the use of thiol-containing acceptors. Going further, these studies may also be the opportunity to get a more fundamental understanding of the control of regioselectivity in transfer reactions. This would provide the knowledge necessary to rationally engineer *Tx*Abf, creating transarabinofuranosylases capable of α-(1,3) linkage formation. In this respect, the R69H-L352M matrix constitutes a good protein background for semi-rational engineering of acceptor subsites, since this mutant allows an easy detection of acceptor-mediated activation for small libraries screening using the purpose-made automated kinetics method, which could be then coupled to a NMR-based secondary screening for linkage formation analysis.

How to further investigate the electron displacement system?

Both the work performed in this thesis and data from the literature (c.f. review article), reveal that crippling the overall enzyme activity is a necessary condition to significantly improve the T/H balance. Fine biochemical studies have shown that the origin of this was localized on the axis [nucleophile-anomeric carbon-leaving or attacking group]. Beyond the question regarding the general role of R69 as a H/T modulator in other GHs from clan A, deeper studies regarding the electron displacement system could be conducted. Indeed, we do believe that differences of selectivity between water and sugar acceptor, at the deglycosylation step, come from deep energetics alterations whose investigation may reveal a lot about the nature of enzyme catalysis.

Could QM/MM studies reveal differences of TS2 energy levels between water and xylotriose?

In spite of a few recent QM/MM studies conducted on retaining GHs (Brás et al., 2010; Wang et al., 2011; Wang et al., 2013; Pierdominici-Sottile et al., 2011; Pierdominici-Sottile et al., 2013) the *in silico* exploration of TS2 relative energy barriers between water and a carbohydrate acceptor, in *Tx*Abf, would be original for two reasons. First, because no such QM/MM studies on furanose-acting GHs

have been carried out to date and, second, because comparative studies of TS2 mediated by different acceptors (i.e. water or sugar) are actually never considered. Currently, crystallographic studies that will hopefully provide structural insight into how *Tx*Abf was evolved into transarabinofuranosylases are on-going and may constitute a strong basis for such a purpose.

Can the transition state structure be mimicked through inhibitor synthesis?

It is well established that it is possible to design potent inhibitors for GHs that mimic transition states (Schramm, 2011). Therefore, it is interesting to wonder whether *Tx*Abf inhibitors that mimic the TS of this enzyme (currently these are not available) would also inhibit its evolved transglycosylating counterparts. If not, this would provide information on the geometry and electronic states of TS in the 'non-Leloir' transarabinofuranosylases. Similarly, it is pertinent to wonder whether specific inhibitors could be designed that would only mimic the TS of transfuranosylases. In this respect, maybe the literature on inhibitors of enzymes involved in the biosynthesis of pathogen cell walls (containing furanoses moieties) should be investigated.

Can the protonation state of key residues be probed via NMR titration?

As a complementary approach to QM/MM and inhibitor studies, and assuming that it has been possible to perform protein NMR analyses, it would be very interesting to monitor the titration of the catalytic nucleophile residue and its surrounding residues (E298, R69, N175, E176) using NMR. Assuming that the resolution of the signals is sufficient, this would provide a reliable view of the proton sharing network in *Tx*Abf and mutants thereof. Such an overarching study would provide one of the most accurate pictures on H/T modulation within the GH class.

References

Arab-Jaziri, F., Bissaro, B., Barbe, S., Saurel, O., Débat, H., Dumon, C., Gervais, V., Milon, A., André, I., Fauré, R. & O'Donohue, M.J. (2012) Functional roles of H98 and W99 and β2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *FEBS J.*, 279 (19), pp.3598–3611.

Badieyan, S., Bevan, D.R. & Zhang, C. (2012) Probing the active site chemistry of β-glucosidases along the hydrolysis reaction pathway. *Biochemistry*, 51 (44), pp.8907–8918.

Brás, N.F., Fernandes, P.A. & Ramos, M.J. (2010) QM/MM studies on the β-galactosidase catalytic mechanism: hydrolysis and transglycosylation teactions. *J. Chem. Theory Comput.*, 6 (2), pp.421–433.

Debeche, T., Cummings, N., Connerton, I., Debeire, P. & O'Donohue, M.J. (2000) Genetic and biochemical characterization of a highly thermostable α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *Appl. Environ. Microbiol.*, 66 (4), pp.1734–1736.

Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B. & Mornon, J.P. (1997) Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology*, 7 (2), pp.277–284.

Euzen, R., Lopez, G., Nugier-, Chauvin, C., Ferrières, V., Plusquellec, D., Rémond, C. & O'Donohue, M. (2005) A chemoenzymatic approach for the synthesis of unnatural disaccharides containing D-galacto or D-fucofuranosides. *Eur. J. Org. Chem.*, 2005 (22), pp.4860–4869.

Feng, H.-Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C. & Dion, M. (2005) Converting a β-glycosidase into a β-transglycosidase by directed evolution. *J. Biol. Chem.*, 280 (44), pp.37088–37097.

Frueh, D.P. (2014) Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog. Nucl. Magn. Reson. Spectrosc.*, 78, pp.47–75.

Kim, Y.-W., Lee, S.S., Warren, R.A.J. & Withers, S.G. (2004) Directed evolution of a glycosynthase from *Agrobacterium* sp. increases its catalytic activity dramatically and expands its substrate repertoire. *J. Biol. Chem.*, 279 (41), pp.42787–42793.

Mainz, A., Religa, T.L., Sprangers, R., Linser, R., Kay, L.E. & Reif, B. (2013) NMR spectroscopy of soluble protein complexes at one mega-dalton and beyond. *Angew. Chem., Int. Ed.*, 52 (33), pp.8746–8751.

Malá, S., Dvoráková, H., Hrabal, R. & Králová, B. (1999) Towards regioselective synthesis of oligosaccharides by use of α-glucosidases with different substrate specificity. *Carbohydr. Res.*, 322 (3-4), pp.209–218.

Nahoum, V., Roux, G., Anton, V., Rougé, P., Puigserver, A., Bischoff, H., Henrissat, B. & Payan, F. (2000) Crystal structures of human pancreatic α-amylase in complex with carbohydrate and proteinaceous inhibitors. *Biochem. J.*, 346, pp.201–208.

Osanjo, G., Dion, M., Drone, J., Solleux, C., Tran, V., Rabiller, C. & Tellier, C. (2007) Directed evolution of the α-L-Fucosidase from *Thermotoga maritima* into an α-L-transfucosidase. *Biochemistry*, 46 (4), pp.1022–1033.

Paës, G., Skov, L.K., O'Donohue, M.J., Rémond, C., Kastrup, J.S., Gajhede, M. & Mirza, O. (2008) The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry*, 47 (28), pp.7441–7451.

Pierdominici-Sottile, G., Horenstein, N.A. & Roitberg, A.E. (2011) A free energy study of the catalytic mechanism of *Trypanosoma cruzi trans*-sialidase. From the Michaelis complex to the covalent intermediate. *Biochemistry*, 50 (46), pp.10150–10158.

Pierdominici-Sottile, G., Palma, J. & Roitberg, A.E. (2013) Free energy computations identify the mutations required to confer *trans*-sialidase activity into *Trypanosoma rangeli* sialidase. *Proteins*, 82 (3), pp.424–435.

Placier, G., Watzlawick, H., Rabiller, C. & Mattes, R. (2009) Evolved β-galactosidases from *Geobacillus stearothermophilus* with improved transgalactosylation yield for galacto-oligosaccharide production. *Appl. Environ. Microbiol.*, 75 (19), pp.6312–6321.

Rémond, C., Plantier-Royon, R., Aubry, N. & O'Donohue, M.J. (2005) An original chemoenzymatic route for the synthesis of β-D-galactofuranosides using an α-L-arabinofuranosidase. *Carbohydr. Res.*, 340 (4), pp.637–644.

Schramm, V.L. (2011) Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu. Rev. Biochem.*, 80, pp.703–732.

Tézé, D., Hendrickx, J., Czjzek, M., Ropartz, D., Sanejouand, Y.-H., Tran, V., Tellier, C. & Dion, M. (2014) Semi-rational approach for converting a GH1 β-glycosidase into a β-transglycosidase. *Protein Eng. Des. Sel.*, 27 (1), pp.13–19.

Wang, J., Hou, Q., Dong, L., Liu, Y. & Liu, C. (2011) QM/MM studies on the glycosylation mechanism of rice BGlu1 β-glucosidase. *J.Mol. Graph.*, 30, pp.148–152.

Wang, J., Hou, Q., Sheng, X., Gao, J., Liu, Y. & Liu, C. (2013) Theoretical study on the deglycosylation mechanism of rice BGlu1 β-glucosidase. *Int. J. Quantum Chem.*, 113 (8), pp.1071–1075.

RÉSUMÉ DES TRAVAUX DE THÈSE

Le présent manuscrit de thèse décrit la conception des premières transarabinofuranosylases de type « non-Leloir » obtenues par évolution moléculaire de l'α-L-arabinofuranosidase de *Thermobacillus xylanilyticus* (*Tx*Abf), hydrolysant les résidus L-arabinofuranosyles présents dans les poly- et oligosaccharides d'arabinoxylanes, hémicelluloses de parois végétale. Outre l'obtention de nouveaux outils de glycosynthèse spécifiques des sucres furanoses, cette étude a permis de manière plus fondamentale et générale de proposer une hypothèse de synthèse quant à la partition des activités de transglycosylation et d'hydrolyse, catalysées par les glycoside hydrolases (GHs).

Étude Bibliographique

En amont des travaux de recherche menés au cours de ces trois années de thèse, une analyse de synthèse des données présentes dans la littérature scientifique, au sujet de la partition hydrolyse/transglycosylation (H/T), a été réalisée. Un simple raisonnement de départ s'appuie sur l'observation principale que des enzymes extrêmement proches sur le plan évolutif (même famille de GH) et partageant des propriétés structurales (même clan) sont capables d'orienter de manière sélective leur activité soit vers l'hydrolyse soit vers la glycosynthèse. A titre d'exemple l'on pourra citer les couples amylosaccharases/sucrose hydrolase (GH13), xyloglucan endo-transglycosylases/hydrolases (GH16) ou bien encore des trans-sialidases/sialidases (GH33), en tant que transglycosylases/hydrolase naturelles. Etant donné que le mécanisme global de ces enzymes est similaire (en 2 étapes avec rétention de configuration du centre anomérique), des stratégies moléculaires communes, indépendamment de la famille d'appartenance, ont forcément dues être mises en place, profitant d'un principe physico-chimique ou thermodynamique sous-jacent à mettre à jour. Une 2ème observation, qui peut apparaître simpliste, est que l'eau est une molécule ubiquitaire, présente en grande abondance (55,5 M) dans les milieux réactionnels où ces enzymes sont mises en œuvre. Par conséquent, lors de l'étape de déglycosylation, la molécule d'eau doit être soumise à des « contraintes » physico-chimiques ou thermodynamiques telles que cette dernière ne puisse pas être une molécule acceptrice dans le cas de transglycosylases naturelles, pour ainsi favoriser la voie de glycosynthèse (en présence d'un sucre accepteur), au dépend de l'hydrolyse. Partant de ces deux constats, la revue analytique menée en continue au cours de cette thèse a permis de mettre à jour des caractéristiques communes développées par les GHs de familles différentes afin de catalyser des réactions de transglycosylation. Ce travail s'est d'une part appuyé sur la comparaison d'enzymes hydrolytiques et glycosynthétiques naturelles mais aussi sur l'analyse du comportement d'enzymes naturellement hydrolytiques converties par évolution moléculaire en transglycosylases artificielles. Une hypothèse de synthèse englobant le comportement des GHs en transglycosylation a été tirée de cette étude, et peut être résumée selon les 3 points clés suivants (Figure 1) :

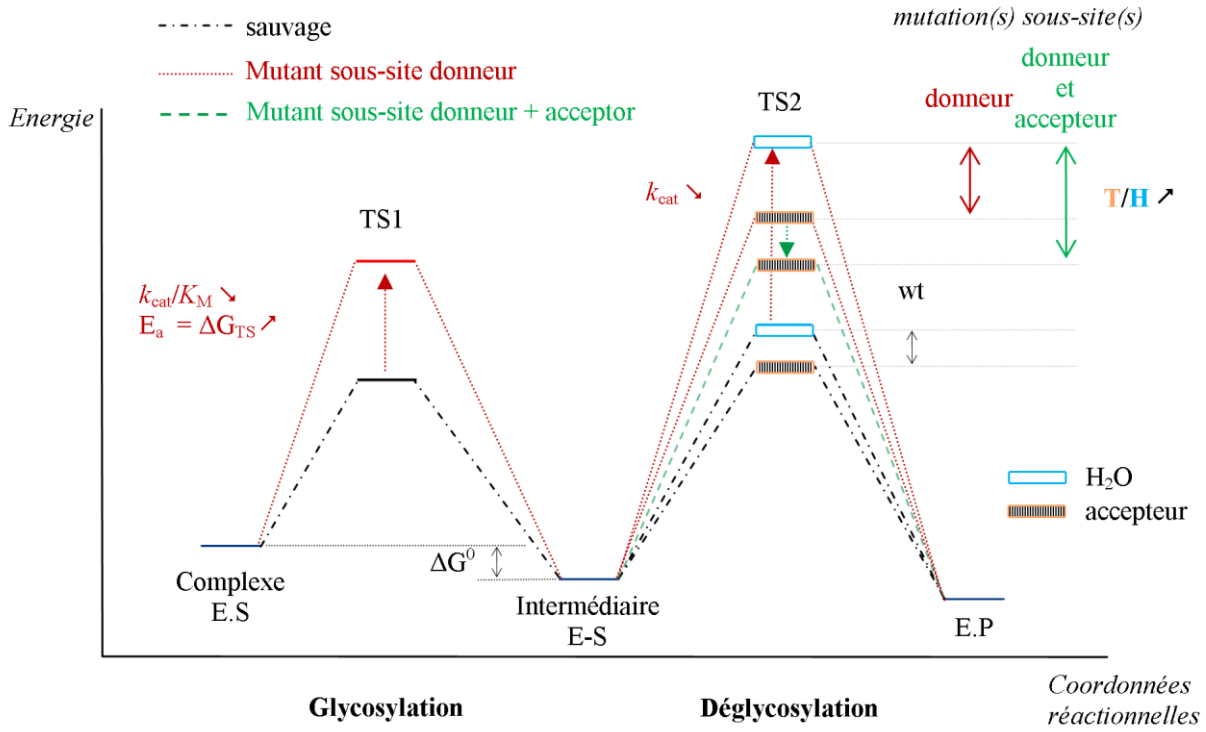


Fig. 1. Diagramme énergétique du mécanisme de rétention à 2 étapes (pointillés noirs) et chemin énergétique alternatif potentiel pour des transglycosylases évoluées. L'introduction de mutations dans le sous-site -1 subsite provoque la déstabilisation de l'état de transition TS2 (pointillés rouges) lié à l'eau (rectangle bleu), qui se reflète mécaniquement dans l'augmentation du niveau énergétique de TS1. Cette augmentation des niveaux d'énergie absolue de $TS2_{eau}$ (—) aussi bien que $TS2_{accepteur}$ (rectangle hachuré) s'accompagne nécessairement d'une plus grande différence d'énergie entre ces deux TS2, du fait de l'augmentation du ratio T/. Cet écart énergétique peut être élargi via l'introduction de mutations positives dans les sous-sites accepteurs. Par ailleurs, dès lors que le réseau d'interactions a été altéré, le niveau énergétique de l'intermédiaire E-S peut aussi être rehaussé (moins stable), favorisant une déglycosylation médiée par l'accepteur sucre. Une combinaison de ces deux phénomènes n'est pas à exclure pour expliquer le comportement des transglycosylases.

(1) La déstabilisation de l'étape de déglycosylation médiée par l'eau, via une augmentation du niveau énergétique de l'état de transition correspondant ($TS2_{eau}$), apparait comme une condition nécessaire pour éviter l'hydrolyse et contrebalancer l'avantage statistique qu'elle présente sur une molécule acceptrice de type sucre.

(2) Une telle déstabilisation trouve probablement son origine dans l'altération des interactions développées au sein des sous-sites donneurs, résultant en une augmentation des barrières énergétiques des étapes de déglycosylation et de glycosylation (TS2 et TS1, respectivement) dès lors qu'elles partagent des propriétés géométriques et électroniques équivalentes.

(3) Faire en sorte que des interactions compensatoires dans les sous-sites accepteurs soient requises pour franchir la barrière énergétique de l'état de transition de la déglycosylation pourrait être une des clés moléculaires complémentaires permettant aux transglycosylases de catalyser la glycosynthèse.

Sur un plan cinétique, la proposition (2) implique qu'une diminution de l'efficacité catalytique (k_{cat}/K_M) serait un prérequis pour que des GHs hydrolytiques soient converties en transglycosylases (dont les représentantes naturelles affichent en effet une très faible efficacité). En complément de cette stratégie, une augmentation de l'affinité dans les sous-sites accepteurs, via des interactions de stacking par exemple, doit pouvoir favoriser la conversion H→T. Certaines stratégies pour défavoriser l'accès de l'eau (augmentation de l'hydrophobicité du site actif, mutation des « canaux à eau ») ou sa réactivité (tyrosines servant « d'ancres à eau ») peuvent également être mises en place afin de diminuer l'activité hydrolytique.

Travaux antérieurs (M2R) et début de thèse

Mon implication dans la compréhension du comportement catalytique de *Tx*Abf débuta avant que je ne commence mes travaux de thèse, précisément durant mon stage de master 2 recherche (M2R) réalisé sous la supervision de Régis Fauré et Michael O'Donohue. Ces travaux de recherche consistèrent dans une première partie à étudier le rôle fonctionnel de la boucle mobile β2α2 de *Tx*Abf sur son activité hydrolytique et sur la reconnaissance du substrat donneur via des méthodes biophysiques telle que la STD-RMN. Cette étude s'est notamment focalisée sur deux résidus clés, H98 et W99, situés à l'extrémité de cette boucle et intervenant dans la formation dynamique des sous-sites +2' et -1, respectivement, lorsqu'une configuration fermée est adoptée (Figure 2, A). Le rôle de W99, passant d'un état exposé au solvant à une position enfouie dans le sous site -1 lors de la fermeture de la boucle, a démontré être prépondérant pour l'intégrité de l'activité hydrolytique. Les études de dynamique moléculaire ont montré une perte de liaisons hydrogènes avec le substrat, l'α-L-arabinofuranoside de *para*-nitrophényle (*p*NP-α-L-Ara*f*), lorsque la mutation W99A était insérée, corroborées par une diminution des interactions mesurées par STD-RMN et une augmentation des valeurs de K_M obtenues par approche biochimique. H98 serait davantage impliqué dans la formation dynamique du site actif et par conséquent la définition d'un site actif opérationnel sur le plan catalytique. Ces travaux ont donné lieu à une publication (Bissaro et al., 2012, *FEBS J.*, 279, 3598-3611)[10] disponible à la section des informations supplémentaires de ce manuscrit.

[10] Arab-Jaziri, F., Bissaro, B., Barbe, S., Saurel, O., Débat, H., Dumon, C., Gervais, V., Milon, A., André I., Fauré, R., O'Donohue, M. J. (2012). Functional roles of H98 and W99 and β2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *FEBS J.*, *279*(19), 3598–3611.

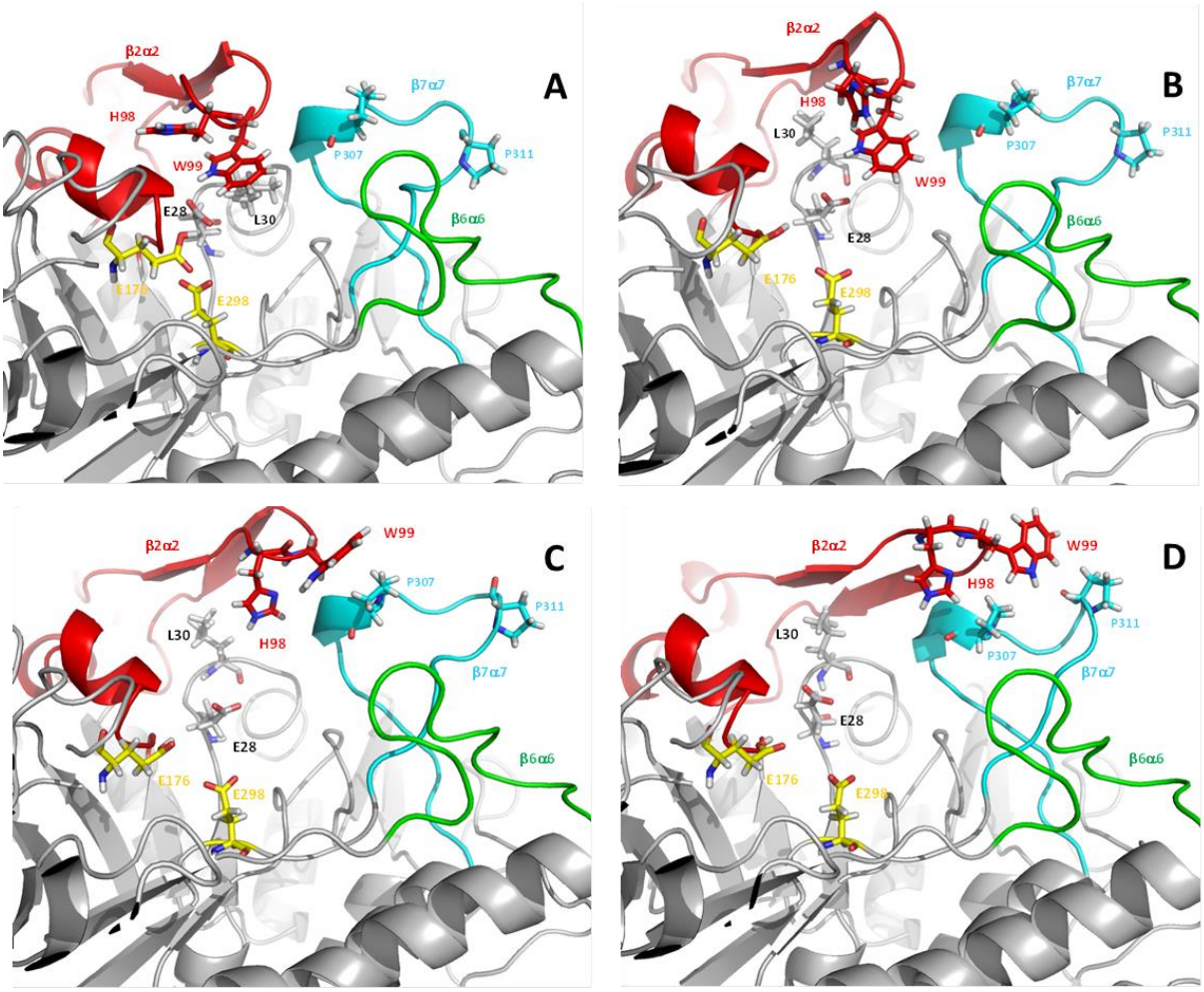


Fig. 2. Prises de vue au cours de la simulation de 40 ns faite par dynamique moléculaire sur *Tx*Abf. Le mouvement des boucles β2α2 (en rouge) et β7α7 (en bleu) au cours de la simulation peut être observé de A à D.

Suite à cette étude nous nous sommes intéressés à la modulation du potentiel de transglycosylation de la *Tx*Abf. Faten-Arab-Jaziri, la doctorante m'ayant précédé sur cette thématique de recherche, avait largement initié ces travaux en réalisant des banques aléatoires sur le gène codant pour la *Tx*Abf. Cette banque fut dans un premier temps criblée en sélectionnant les clones présentant une faible activité hydrolytique sur l'α-L-arabinofuranoside de 5-bromo-4-chloro-3-indolyle (X-α-L-Ara*f*), apparaissant ainsi comme bleu pâle selon le protocole développé par Koné et coll. en 2009[11] et adapté à la *Tx*Abf en collaboration avec l'équipe de Charles Tellier (UFIP, Nantes). Ces mutants furent ensuite évalués vis-à-vis de leur capacité à transglycosyler l'accepteur synthétique α-L-arabinofuranoside de benzyle (Bn-α-D-Xyl*p*), étant un des accepteurs de référence pour l'étude de la transglycosylation de l'enzyme sauvage (Rémond et al., 2004)[12]. Au cours de mon stage de M2R certains de ces mutants (notés mutants 6, 1364 et 1753) ont été plus finement

[11] Koné, F. M. T., Le Béchec, M., Sine, J.-P., Dion, M., & Tellier, C. (2009). Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng. Des. Sel.*, *22*(1), 37 –44.

[12] Rémond, C., Plantier-Royon, R., Aubry, N., Maes, E., Bliard, C., & O'Donohue, M. J. (2004). Synthesis of pentose-containing disaccharides using a thermostable α-L-arabinofuranosidase. *Carbohydr. Res.*, *339*(11), 2019–2025.

caractérisés notamment via la mise au point de suivi cinétique en temps réel par RMN des activités d'hydrolyse et de transglycosylation. Cette première étude de mutants de *Tx*Abf générés de manière aléatoire a permis de montrer que la partition H/T pouvait être modulée pour une furanosidase, comme cela avait été précédemment démontré pour des pyranosidases, et fut l'objet d'une publication (Arab-Jaziri et al. (2013), NBT, 30)[13]. Par ailleurs, certaines mutations localisées près du sous-site -1 peuvent être identifiées comme potentielles coupables (Figure 3), bien qu'elles n'aient pas encore été étudiées individuellement.

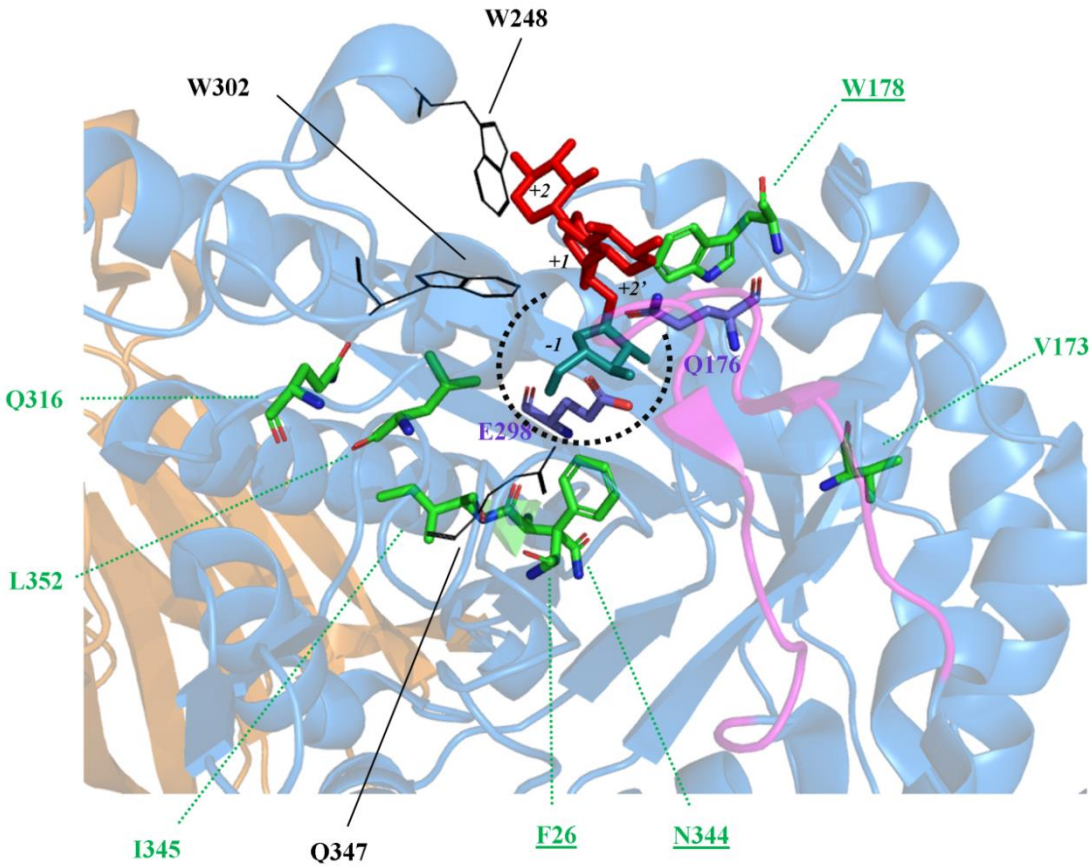


Fig. 3. Récapitulatif des mutations identifiées au cours des 2 premiers tours, indépendants, d'évolution dirigée. Ces mutations, situées dans la 1ère ou 2ème couronne sont représentées par des bâtonnets (vert) au sein de la structure de *Tx*Abf (PDB 2VRQ). Les positions mutées lors de la 1ère étude d'évolution dirigée sont soulignées, sachant que la position N344 a été retrouvée mutée au cours des deux études. Les acides aminés impliqués dans la topologie du site actif et susceptibles d'être impactés par les mutations identifiées sont représentés par des lignes noires alors que les résidus catalytiques nucléophile E298 et acide/base E176 (muté en Q) sont représentés par des bâtonnets violets. La ligne en pointillés démarque le sous-site -1 accueillant la partie L-arabinofuranosyle du substrat.

Dans un 2ème temps, toujours dans la perspective d'améliorer l'activité de transglycosylation de la *Tx*Abf, une 2ème banque aléatoire fut générée à partir du gène sauvage de *Tx*Abf par F. Arab-Jaziri. Cette fois-ci, une méthode de criblage en deux étapes fut appliquée en ayant pour ambition de

[13] Arab-Jaziri, F., Bissaro, B., Dion, M., Saurel, O., Harrison, D., Ferreira, F., Milon, A., Tellier, C., O'Donohue, M.J., Fauré, R. (2013). Engineering transglycosidase activity into a GH51 α-L-arabinofuranosidase. *N. Biotechnol.*, *30*(5), 536–544.

détecter des clones possédant, d'une part, une faible activité hydrolytique et, d'autre part, un potentiel de transfert amélioré. Lors de ce criblage, la cible à L-arabinofuranosyler ne fut plus un accepteur artificiel, comme lors de la 1ère étude ci-avant évoquée, mais un substrat naturel, à savoir des xylo-oligosaccharides (XOS). L'enzyme sauvage étant naturellement inhibée par les accepteurs XOS, l'identification de mutants positifs en est rendue d'autant plus compliquée. Suite au criblage *in vivo* réalisé en milieu solide, 199 mutants furent analysés *in vitro* sur extraits enzymatiques bruts pré-purifiés pour tester une éventuelle activation en présence de XOS (une activité supérieure à 100% serait alors synonyme d'une véritable activité de transglycosylation). Malheureusement, seulement des levées apparentes d'inhibition furent observées, avec des activités relatives avoisinant les 100% quand l'enzyme sauvage ne présente plus qu'environ 30% d'activité relative dans des conditions équivalentes. A la suite de ce criblage secondaire, des suivis de transglycosylation par CCM ont été réalisés afin d'obtenir une vision directe de l'activité de synthèse et 11 mutants furent alors retenus. Au début de mon projet de thèse l'activité de transglycosylation de ces 11 mutants a alors été caractérisée par RMN permettant de valider 5 d'entre eux, notés comme n° 1, 8, 169, 2032 et 2068. Presque toutes les mutations aléatoires portées par ces mutants ont été isolées au cours de la 1ère année de ce projet en utilisant 2 critères de sélection : (a) une analyse du taux de conservation des résidus au sein de la diversité des séquences de GH51 et (b) la localisation des dites mutations au sein de la structure de *Tx*Abf. En pratique, les mutations V173D, N344I, N344Y, I345L, L352M et Q316L ont ainsi été conservées car ces substitutions n'apparaissent pas dans la diversité des séquences de GH51 à des positions équivalentes et/ou sont proches du site actif (1ère ou 2ème couronne, i.e. à une distance maximale de 15 Å du résidu catalytique nucléophile) (Figure 3). Une caractérisation en transglycosylation des mono-mutants a permis d'identifier les mutants N344Y et L352M comme potentiellement intéressants pour la modulation de la partition H/T et méritant de plus amples investigations. Ce 2ème volet d'évolution dirigée ayant mené à l'identification de ces dernières mutations est l'objet d'une publication actuellement soumise (cf annexes). Néanmoins, le potentiel de synthèse des mutants retenus restant très modeste (< 20%), une stratégie d'évolution moléculaire plus poussée et plus élaborée était nécessaire et fait l'objet des travaux de thèse décrits ci-après.

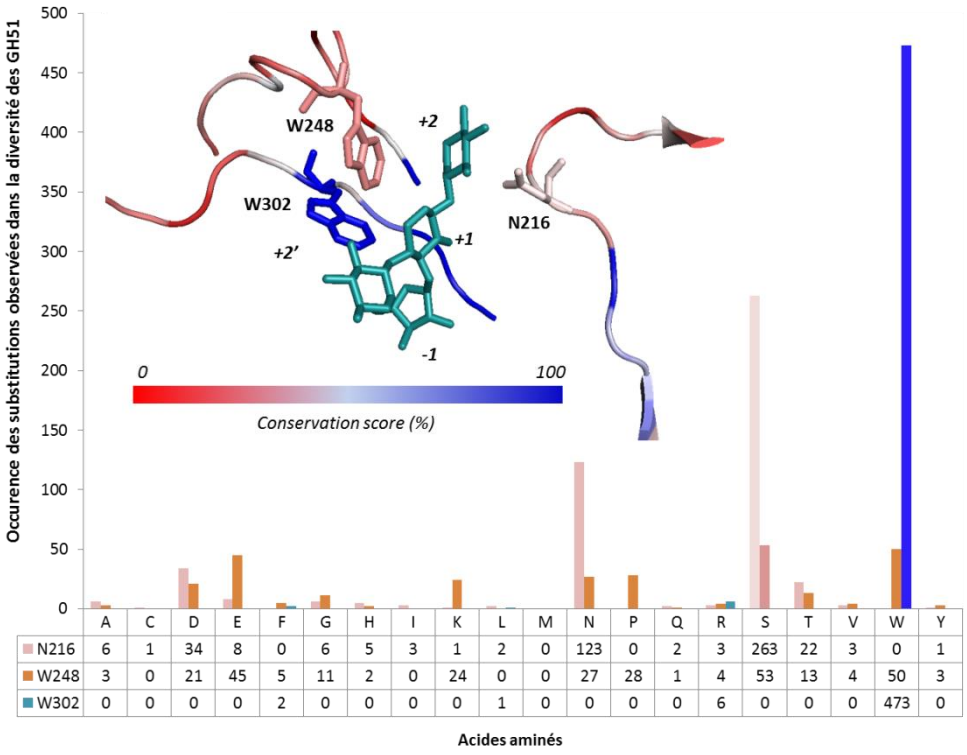
Au cœur du projet

Partie I : Étude des déterminants de la régiosélectivité chez TxAbf

La formation de liens glycosidiques via l'action d'enzymes permet un contrôle total de la stéréospécificité (i.e. formation de lien de configuration α ou β) mais souffre souvent d'un manque de régiospécificité, plusieurs régioisomères étant habituellement obtenus lors des réactions de transglycosylation. Dans l'optique de comprendre quels déterminants permettent de contrôler ou d'influer sur la régiosélectivité de *Tx*Abf, trois banques à saturation furent réalisées sur des positions a priori clés pour le positionnement de la molécule acceptrice. Ces positions, W302, W248 et N216 sont respectivement situées dans les sous-sites -1/-+1, +1 et +2 (Figure 4).Par ailleurs, cette étude a été couplée à la compréhension de l'impact de l'anomérie du sucre accepteur sur la régioselectivité. En effet, il avait été précédemment remarqué (Rémond et al, 2004)[14] pour le substrat accepteur artificiel Bn-α-D-Xyl*p* qu'un seul régioisomère avec une liaison (1,2) était obtenu et il a été montré au cours de cette étude que l'utilisation de son équivalent avec une anomérie β (Bn-β-D-Xyl*p*) menait à un mélange de régioisomères (1,2) et (1,3) en proportion 2:1, respectivement. Le même ratio a été mesuré pour des structures naturelles, telle qu'avec le xylobiose (X_2) ou xylotriose (X_3), présentant un lien β-(1,4). Il a alors été proposé, par modélisation moléculaire appuyée par des données biochimiques de mutagenèse dirigée-activité, qu'un positionnement alternatif était adopté par le groupement benzyle lié en α-D. L'aglycone se positionne alors dans le sous-site +2' en imposant une regiospecificité (1,2). Au contraire, les accepteurs avec une anomérie β-D (Bn-β-Xyl*p*, accepteurs naturels tels que X_2 et X_3) adoptent un positionnement plus « classique » du point de vue de l'enzyme, dans le sens +1→+2, mais sans contrôle drastique de la régiosélectivité.

L'étude des banques à saturation a notamment mis en lumière la mutation N216W permettant de supprimer la production du régioisomère (1,3) en présence d'accepteurs contenant des liens β-D et créant ainsi une enzyme régiospécifique pour le transfert de résidus L-arabinofuranosyles ou D-galactofuranosyles avec une liaison (1,2). Une publication relative à cette étude est actuellement cours de finalisation (cf partie I du chapitre résultats).

[14] Rémond, C., Plantier-Royon, R., Aubry, N., Maes, E., Bliard, C., & O'Donohue, M. J. (2004). Synthesis of pentose-containing disaccharides using a thermostable α-L-arabinofuranosidase. *Carbohydr. Res.*, *339*(11), 2019–2025.



	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
N216	6	1	34	8	0	6	5	3	1	2	0	123	0	2	3	263	22	3	0	1
W248	3	0	21	45	5	11	2	0	24	0	0	27	28	1	4	53	13	4	50	3
W302	0	0	0	0	2	0	0	0	0	1	0	0	0	0	6	0	0	0	473	0

Fig. 4. Localisation des résidus W302, W248 et N216 au sein du site actif de *Tx*Abf et analyse du taux de conservation aux positions équivalentes au sein de la diversité des GH51 (488 séquences).

Partie II : N344, un modulateur de pH optimum permettant de supprimer l'hydrolyse secondaire

Dans la perspective de comprendre la modulation de rapport H/T, la position N344, retrouvée mutée à plusieurs reprises en différents acides aminés (N344I, K, S et Y) dans deux banques aléatoires différentes, fut considéré comme un « point chaud », notamment de par sa position sur le brin β8 en dessous du résidu catalytique nucléophile E298, situé sur le brin parallèle β7 (Figure 5A). Une banque à saturation fut créée à cette position et le potentiel de transglycosylation des 19 variants fut évalué par CCM puis par RMN. Deux types de comportements furent observés, l'un permettant une augmentation légère du rendement de synthèse et l'autre introduisant une suppression de l'hydrolyse secondaire du produit de synthèse. En effet, ce dernier étant généralement un substrat de l'enzyme, son hydrolyse secondaire mène souvent à des rendements quasi nuls après des temps prolongés de réaction. La compréhension des mécanismes sous-jacents à la suppression apparente de cette hydrolyse secondaire nous a alors apparu essentielle. Une caractérisation biochimique et biophysique des mutants N344P et N344Y mena aux conclusions principales suivantes :

(1) Les mutations N344Y et P handicapent de manière générale l'activité hydrolytique (réduction du k_{cat}).

(2) Ces mutations impactent l'état d'ionisation du résidu nucléophile E298 comme le suggère l'augmentation du pK_a de ce dernier alors que l'acide/base E176 n'est pas affecté (Figure 5B).

(3) La mutation de N344 impacte très probablement le réseau d'interactions hydrogène établi avec E298, peut-être via le résidu R69 (à distance d'interaction ce dernier) vers lequel la chaine latérale de N344 pointe.

(4) Cette modification se traduit par une inhibition pH-dépendante de l'enzyme par le groupement partant. Cette dernière permet alors de stopper la réaction avant que l'hydrolyse secondaire ne se déclenche en réalisant la réaction à pH 5,8. À pH 7,0 les activités (hydrolytiques et transglycosylation) sont restaurées. Néanmoins, la compréhension fondamentale des liens entre modification des réseaux d'interactions hydrogène et cette inhibition pH-dépendante reste à établir.

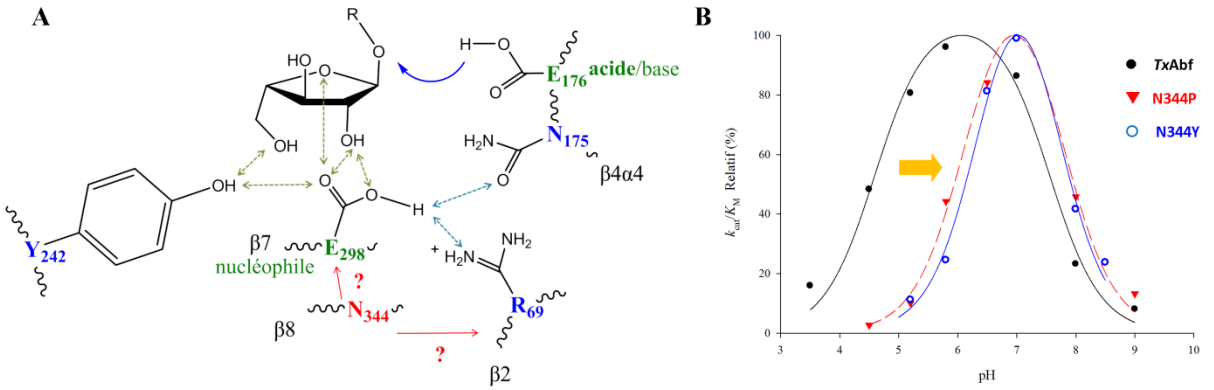


Fig. 5. Impact des mutations N344Y et P. (A) Schéma du réseau d'interactions au sein du sous-site -1 de *Tx*Abf et (B) impact des mutations N344P et Y sur le pK_a du résidu catalytique nucléophile E298.

Partie III : Analyse in silico via BindScan pour une meilleure reconnaissance de l'accepteur

Bien que l'approche semi-rationnelle ait permis d'identifier la mutation N216W comme étant intéressante pour l'introduction d'une régiospécificité (1,2), le développement d'outils plus rationnels pour la prédiction de mutations permettant l'introduction de nouvelles spécificités d'accepteur ou le contrôle de la régioselectivité de transfert constitue un des challenges actuels et à venir en ingénierie enzymatique. A cet égard, une collaboration avec le Dr. Xevi Biarnés et le Prof. Antoni Planas de l'IQS de Barcelone a été engagée afin de tirer profit d'un nouvel algorithme, nommé « BindScan », développé dans leur laboratoire et ayant besoin d'être éprouvé. Brièvement, l'objectif de ce programme consiste à prédire des mutations (« hotspot ») au sein de la séquence protéique permettant la reconnaissance d'un substrat non naturel pour l'enzyme (création de nouvelle spécificité régio- ou de substrat) ou, à la marge, d'améliorer la reconnaissance d'un substrat de l'enzyme (Figure 6). C'est dans ce dernier cas de figure que se situe notre problématique. La recherche de mutations permettant un gain d'affinité significatif pour le xylotriose (envers le complexe *Tx*Abf-β-L-Ara*f*, construit *in silico* à partir de *Tx*Abf libre et de l'intermédiaire covalent

*Gs*Abf-β-L-Ara*f*) [15] a permis l'identification d'un « point chaud », la position G179, située dans le sous-site +1. Les mutations de G179 en Y, F et W ont alors été réalisées. Par ailleurs ces mutations furent également couplées à la mutation la plus efficace identifiée par voie aléatoire, à savoir L352M, située dans le sous-site -1.

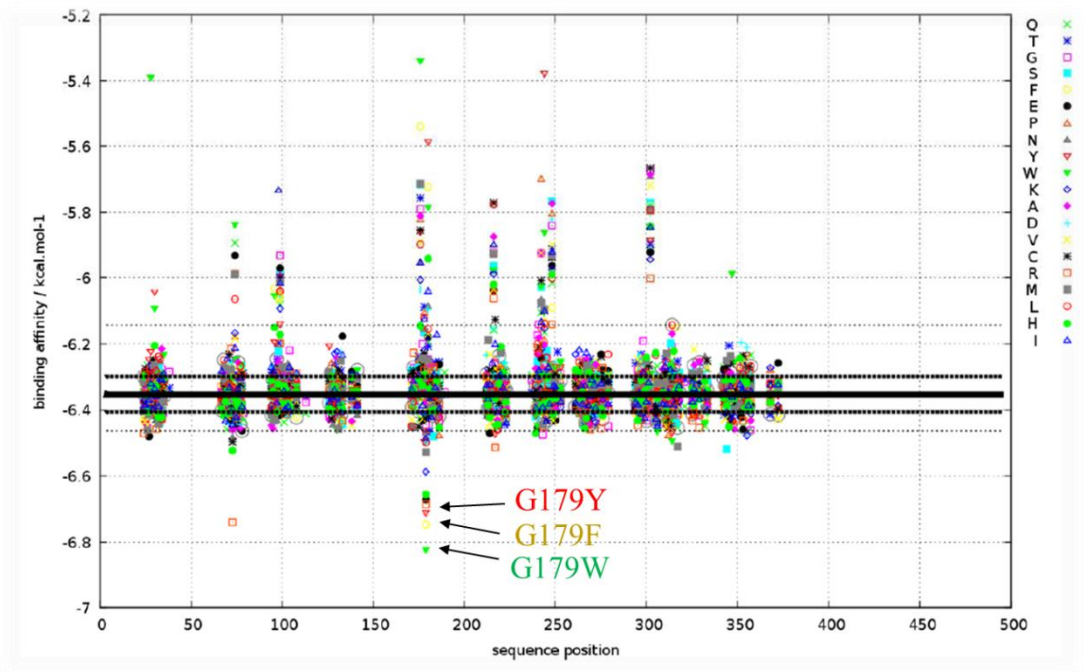


Fig. 6. Résultat de BindScan pour la *Tx*Abf visant à former une liaison α-(1,3) entre l'L-arabinofuranose et le xylotriose en position centrale (XA^3X). Sur le plan pratique, les 19 variants sont générés à chaque position de la séquence protéique (0 à 496) et pour chaque mono-mutant une expérience de docking est réalisée afin d'évaluer l'affinité pour l'accepteur. Ainsi, des points chauds de mutation, comme la position G179, sont identifiés.

L'analyse des paramètres cinétiques en hydrolyse a permis de montrer que ces substitutions en acide aminés aromatiques permettent d'augmenter l'affinité pour le substrat donneur (diminution du K_M, 12-42% relative values for G179F, Y or W), probablement via des interactions de stacking avec le *p*NP se positionnant dans le +1. Cet impact se vérifie que l'on parte d'une matrice vierge *Tx*Abf (K_M = 0.27 mM) ou d'une matrice mutée L352M (K_M = 9 mM). Concernant l'activité de transglycosylation, en présence de *p*NP-α-L-Ara*f* et de xylotriose, une augmentation du taux de transfert (jusqu'à 2 fois) a été mesurée. Plus intéressant encore, l'hydrolyse secondaire est repoussée en toute fin de réaction comparativement aux enzymes sans les mutations G179F, Y ou W. Ceci amena à la conclusion qu'une différence d'affinité a été introduite entre le substrat donneur (*p*NP-α-L-Ara*f*) et le produit de synthèse (A^2XX par exemple) permettant à l'enzyme d'hydrolyser dans un premier temps, de manière hautement préférentielle, le donneur reportant ainsi la dégradation du produit d'intérêt. Cette hypothèse a pu être en partie corroborée par une étude calorimétrique (ITC) des interactions entre

[15] *Gs*Abf : α-L-arabinofuranosidase de *Geobacillus stearothermophilus*

des mutants inactivés (E176Q, E176Q-G179F et E176Q-N216W) et deux ligands, le donneur *p*NP-α-L-Ara*f* et le produit de transglycosylation majoritaire A^2XX. En effet, ces expériences ont permis de montrer que la mutation G179F prédite par approche computationnelle (sous-site +1) permet effectivement d'augmenter l'énergie d'interaction avec le A^2XX. En revanche, le différentiel d'affinité entre donneur et produit de synthèse, relativement à l'enzyme sauvage, n'a pas pu être mesuré via cette méthode.

Au global, cette étude se trouve être originale sur plusieurs plans : d'une part, elle permet de confronter les prédictions de l'algorithme *BindScan* à des données expérimentales et démontre ainsi la pertinence de cet outil pour l'enzyme modèle *Tx*Abf. D'autre part, un effet synergique entre mutations issues de différentes approches (*in silico* et aléatoire) a pu être mis en évidence, appuyant le fait que L352M pourrait constituer une enzyme de départ adéquate pour un nouveau tour d'évolution. Enfin, cette stratégie a permis de mettre à jour une nouvelle approche pour sauvegarder le produit de synthèse.

Partie IV : Conception par évolution moléculaire des 1ère transarabinofuranosylases de type « non Leloir »

Suite au succès de recombinaison avec L352M observé via l'étude du *BindScan*, la séquence codante pour ce mutant fut soumise à une mutagenèse aléatoire et le criblage de cette nouvelle banque (40 000 clones), sur milieu solide en présence de X-α-L-Ara*f*, permis l'identification de 200 clones ayant *a priori* une activité hydrolytique plus faible. Après avoir développé un protocole de cinétiques automatisées (robot TECAN de la plateforme Ingénierie et Criblage d'enzyme Originale - ICEO- du LISBP), chacun des extraits bruts enzymatiques de ces clones fut utilisé pour la réalisation de deux cinétiques, l'une en présence de donneur (*p*NP-α-L-Ara*f*) et l'autre en présence de donneur et d'accepteur (xylotriose). Le calcul des ratios d'activité mit à jour, pour la première fois, des mutants très activés (> 200% d'activité), tous présentant une mutation commune, R69H, dont le rôle individuel fut ensuite validé. Cette position, suspectée lors de l'étude sur les mutants N344, s'avère être une position totalement conservée au sein du clan A, rassemblant actuellement 19 familles de GHs. Sur le plan mécanistique, la mutation R69H supprime une liaison hydrogène avec le résidu catalytique E298 et provoque une augmentation du pK_a de ce dernier, comme observé avec les mutants N344P et Y. Ceci est reflété par des diminutions drastiques du k_{cat} (19 and 4% for N344P and Y resp.) et du k_{cat}/K_M (4.8 and 1.7% resp.), imputées à une altération de l'efficacité de création/rupture de liaison covalente entre le résidu nucléophile E298 et le carbone anomérique de l'arabinofuranoside. Ceci a été démontré par une analyse de Hammett-Bronsted, nécessitant une synthèse d'un panel d'L-arabinofuranosides portant différents groupements partants, couplée à une mesure d'effets cinétiques isotopiques du solvant. Cet impact sur les états de transition, en total

accord avec la proposition de synthèse faite lors de l'étude bibliographique, se transcrit par l'introduction d'un comportement initial (première phase de la réaction) de quasi « pure » transglycosylase lorsque les réactions de transglycosylation sont analysées par RMN, en utilisant le mono-mutant R69H comme biocatalyseur. Cependant, une hydrolyse secondaire majeure, diminuant le rendement final, est toujours catalysée par ce mutant. Pour remédier à cela un effet synergique a alors été démontré via la combinaison de R69H avec L352M, permettant de diminuer significativement l'hydrolyse secondaire, bien que les rendements maximaux ne dépassent toujours pas 30%. Après avoir altéré profondément la partition H/T en faveur de la transglycosylation grâce à ces mutations du sous-site -1, il s'est agi de tirer profit des mutations identifiées dans les sous-sites accepteurs via les approches semi-rationnelles (N216W, sous-site +2) ou computationnelles (e.g. G179F, sous-site +1). Il s'est alors avéré que les mutants R69H-L352M-N216W et R69H-L352M-G179F affichèrent des comportements de transglycosylases, avec des rendements supérieurs à 80%. Ces dernières peuvent être décrites comme les toutes premières transarabinofuranosylases de type 'non-Leloir' (Figure 7). La connaissance acquise sur le contrôle pH-dépendant de l'inhibition a pu être réinjectée dans ces dernières expériences permettant alors de sauvegarder les produits de transarabinofuranosylation, avec des rendements d'environ 70%.

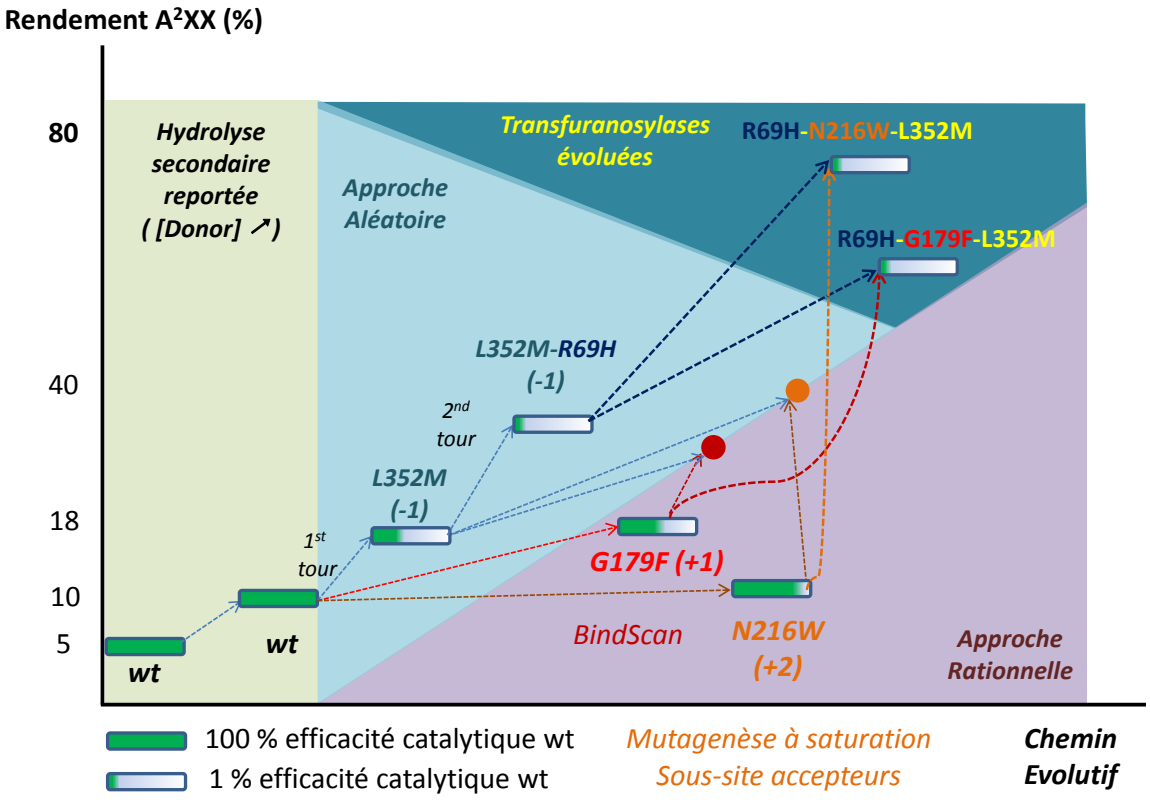


Fig. 7. Chemin evolutif de la *Tx*Abf vers les 1ères transarabinofuranosylases de type 'non-Leloir'.

Partie V: En route vers la synthèse chimio-enzymatique d'arabinoxylanes artificiels

Cette partie de mon projet de thèse, développée en toute de fin de parcours, constitue un travail prospectif ayant pour objectif de mettre en œuvre les biocatalyseurs développés et décrits ci-avant. L'idée centrale est de recréer un « négatif » en synthèse de l'arsenal hydrolytique développé par la Nature pour hydrolyser les parois végétales et plus précisément les arabinoxylanes (AXs). Ceci dans l'objectif d'avoir des outils de synthèse chimio-enzymatique pour la conception, à-façon, de motifs (oligo- et polysaccharides) bio-inspirés aux propriétés physiologiques ou physico-chimiques avérés ou restant à démontrer. Dans cette perspective, une des premières étapes consiste à créer de courts motifs, arabinoxylo-oligosaccharides (AXOS), chose rendue possible grâce à l'utilisation des transarabinofuranosylases artificielles. Ensuite, ces courts motifs doivent être assemblés (Figure 8).

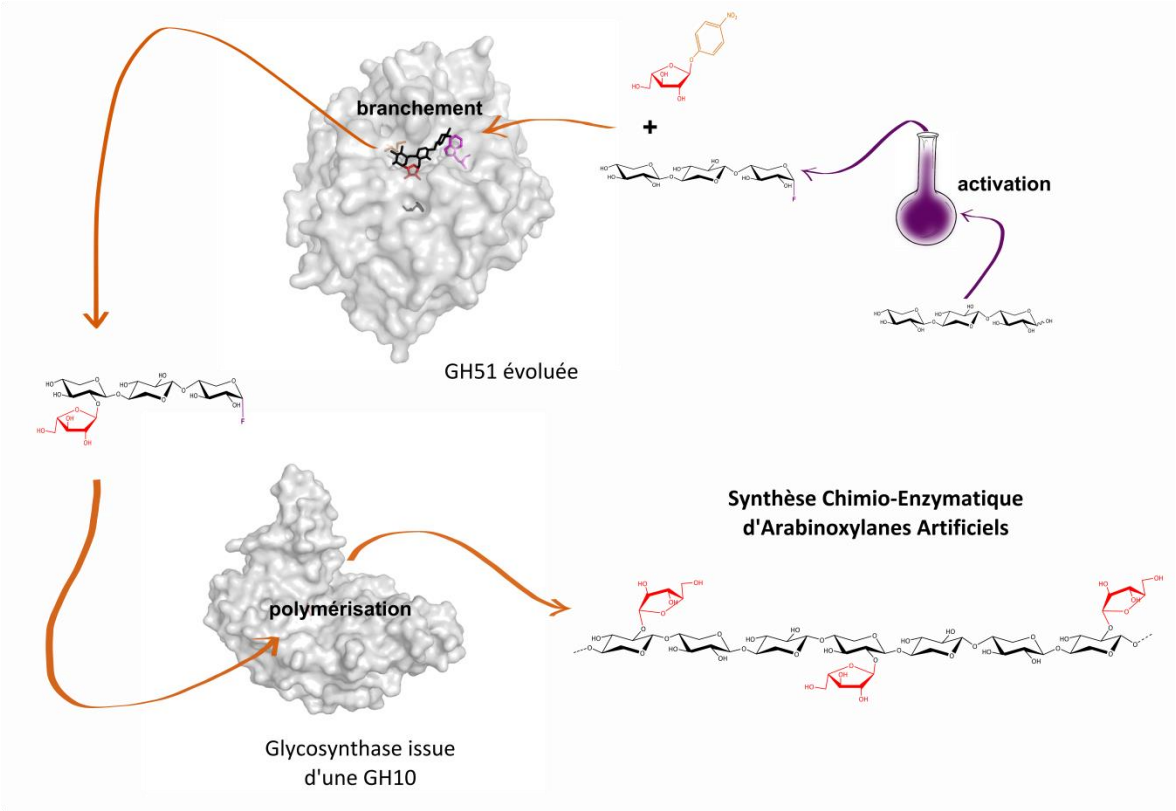


Fig. 8. Schéma de principe de la synthèse chimio-enzymatique d'arabinoxylanes artificiels.

L'option choisie consiste à utiliser une enzyme qui hydrolyse naturellement les liens entre ces motifs présents dans les AXs, à savoir une xylanase de la famille GH10. Cette enzyme avait précédemment été convertie en glycosynthase (XT6-E265G) au sein de l'équipe de Yuval Shoham (Technion, Israël) qui accepta de nous envoyer le plasmide contenant le gène de cette dernière. Les glycosynthases nécessitant un substrat fluoré pour fonctionner, il fallut d'abord modifier par voie chimique les AXOS. Deux stratégies ont été mises en place pour une telle activation, soit post- soit pré-transglycosylation (à savoir une activation sur le produit de transglycosylation synthétisé par le mutant de *Tx*Abf ou bien une activation directe de l'accepteur (xylotriose) utilisé par la *Tx*Abf menant

à un oligosaccharide « branché » alors activé et prêt à être utilisé par la glycosynthase). La voie post-transglycosylation s'est révélée être relativement compliquée notamment du fait que la fluoration, en milieu très acide, provoque l'hydrolyse partielle du lien glycosidique entre le résidu L-arabinofuranosyle et le xylotriose, menant en un mélange difficile à purifier à l'échelle de produit considérée (<20 mg). Les analyses HPAEC-PAD des polymères synthétisés et des produits issus de leur hydrolyse acide et enzymatique ont néanmoins montré que la xylanosynthase était capable d'introduire des oligosaccharides branchés (i.e. AXOS) dans le polymère. Ceci constituant une première preuve encourageante quant à l'utilisation de XT6-E265G pour un tel objectif. Concernant la voie d'activation, celle de « pré-transglycosylation » étant possible (l'accepteur α-xylotriose-F est bien reconnu par *Tx*Abf), cette dernière fut privilégiée. L'analyse en spectrométrie de masse des polymères est en cours lors du dépôt du présent manuscrit.

Conclusions

Le domaine de la glycosynthèse est un champ d'investigation relativement ancien, abordé par les sciences chimiques depuis la fin du 19ème siècle. Malgré d'importants progrès, le contrôle des paramètres influant la glycosylation reste encore relativement obscur. L'approche enzymatique offre une alternative de premier choix pour catalyser la formation d'une liaison entre deux (ou plusieurs) unités glycosidiques tout en contrôlant la stéréochimie des centres anomériques ainsi que la régiosélectivité. Les glycosides hydrolases (GHs) représentent une source riche et abondante de biocatalyseurs de par notamment la diversité des substrats cibles et donc des activités correspondantes. L'identification des facteurs clés contrôlant cette partition constitue une première étape majeure vers la conversion systématique de GHs en transglycosylases, sans utiliser de substrats activés (e.g. substrats fluorés pour les glycosynthases et jusqu'alors impossible à synthétiser en conformation furanose). Bien que la façon d'influer sur ces paramètres de manière purement rationnelle représente un champ d'investigation encore à explorer, cette connaissance permettra d'enrichir de manière incroyable le panel des outils de synthèse en Glycosciences. La présente étude a permis d'aborder cet enjeu en se focalisant sur un sous-domaine de la chimie des sucres relativement délaissé, celui des furanoses. Différentes questions concernant la partition H/T chez les GHs sont à soulever :

- Comment augmenter la spécificité de transfert envers un sucre tout en diminuant celle vers l'eau ?
- Comment influer sur l'hydrolyse secondaire du produit de synthèse, substrat de l'enzyme, pour conserver un rendement final acceptable ?
- Comment contrôler la régioselectivité de transfert ?

Ces différentes questions furent abordées sans ordre prédéfini et de manière complémentaire au

cours de ces travaux de thèse. L'analyse critique des données de la littérature a permis de montrer que des règles générales ont probablement été mises en place par la Nature pour permettre à certaines GHs de catalyser la transglycosylation en milieu aqueux. Ces règles s'appliquent au cas de la *Tx*Abf pour laquelle nous avons réussi à recréer un chemin évolutif artificiel combinant des stratégies d'ingénierie enzymatique à la fois semi-rationnelle, aléatoire et computationnelle, menant aux premières transarabinofuranosylases de type 'non-Leloir'. De nombreuses questions restent néanmoins sans réponse, promettant de nombreuses études à venir.

Table of Illustrations

Figures

General Discussion, Conclusions and Perspectives

Résumé des Travaux de Thèse

Tables

Supporting Information

the FEBS Journal

Functional roles of H98 and W99 and β2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*

Faten Arab-Jaziri[1,2,3,4,*], Bastien Bissaro[1,2,3,*], Sophie Barbe[1,2,3], Olivier Saurel[1,5], Hélène Débat[6], Claire Dumon[1,2,3], Virginie Gervais[1,5], Alain Milon[1,5], Isabelle André[1,2,3], Régis Fauré[1,2,3] and Michael J. O'Donohue[1,2,3]

1 Universite´ de Toulouse, INSA, UPS, INP, Toulouse, France

2 INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France 3

3 CNRS, LISBP UMR5504, Toulouse, France

4 Agence de l'environnement et de la Maîtrise de l'Energie, Angers, France

5 CNRS, IPBS UMR5089, Institut de Pharmacologie et de Biologie Structurale, Toulouse, France

6 Université de Versailles St Quentin-en-Yvelines, Institut de Génétique et Microbiologie, Faculté des Sciences Paris-Sud, Orsay, France

Keywords
glycoside hydrolase; molecular dynamics; STD-NMR; a-L-arabinofuranosidase; βα loop

Correspondence
M. J. O'Donohue, INRA, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, 135 avenue de Rangueil, 31077 Toulouse Cedex 4, France
Fax: +33 5 6155 9400
Tel: +33 5 6155 9428
E-mail : michael.odonohue@insa-toulouse.fr

***These authors contributed equally to this work**

(Received 6 April 2012, revised 20 June 2012, accepted 10 July 2012)

doi:10.1111/j.1742-4658.2012.08720.x

This study is focused on the elucidation of the functional role of the mobile β2α2 loop in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*, and particularly on the roles of loop residues H98 and W99. Using site-directed mutagenesis, coupled to characterization methods including isothermal titration calorimetry (ITC) and saturation transfer difference nuclear magnetic resonance (STD-NMR) spectroscopy, and molecular dynamics simulations, it has been possible to provide a molecular level view of interactions and the consequences of mutations. Binding of *para*-nitrophenyl α-L-arabinofuranoside (*p*NP-α-L-Ara*f*) to the wild-type arabinofuranosidase was characterized by K_d values (0.32 and 0.16 mM, from ITC and STD-NMR respectively) that highly resembled that of the arabinoxylo-oligosaccharide XA^3XX (0.21 mM), and determination of the thermodynamic parameters of enzyme : *p*NP-α-L-Ara*f* binding revealed that this process is driven by favourable entropy, which is linked to the movement of the β2α2 loop. Loop closure relocates the solvent-exposed W99 into a buried location, allowing its involvement in substrate binding and in the formation of a functional active site. Similarly, the data underline the role of H98 in the 'dynamic' formation and definition of a catalytically operational active site, which may be a specific feature of a subset of GH51 arabinofuranosidases. Substitution of H98 and W99 by alanine or phenylalanine revealed that mutations affected K_M and/or k_{cat}. Molecular dynamics performed on W99A implied that this mutation causes the loss of a hydrogen bond and leads to an alternative binding mode that is detrimental for catalysis. STD-NMR experiments revealed altered binding of the aglycon motif in the active site, combined with reduced STD intensities of the α-L-arabinofuranosyl moiety for W99 substitutions.

Abbreviations
Abf, α-L-arabinofuranosidase; ITC, isothermal titration calorimetry; L-Ara*f*, L-arabinofuranosyl; MD, molecular dynamics; *p*NP, *para*-nitrophenol; *p*NP-α-L-Ara*f*, *para*-nitrophenyl α-L-arabinofuranoside; *p*NP-β-D-Gal*f*, *para*-nitrophenyl β-D-galactofuranoside; STD-NMR, saturation transfer difference nuclear magnetic resonance spectroscopy; *Tx*Abf, α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*; *Tx*Abf†, inactivated *Tx*Abf E176A mutant; vdW, van der Waals.

Introduction

α-L-Arabinofuranosidases (Abfs) are *exo*-acting glycoside hydrolases (GHs) that catalyse the hydrolysis of terminal non-reducing α-L-arabinofuranosyl (α-L-Ara*f*) moieties in oligo- and polysaccharides, through the cleavage of (1→2) and/or (1→3) osidic bonds. In the CAZy classification, Abfs (**EC 3.2.1.55**) that act not only on arabinoxylans, but also on a wide range of L-arabinofuranose-containing polymers, including arabinogalactans and arabinans, and synthetic

substrates such as *p*-nitrophenyl α-L-arabinofuranoside (*p*NP-α-L-Ara*f*), are present in the GH families 3, 43, 51, 54, and 62 [1]. While some of these families display multiple enzyme specificities, family GH51 is mostly composed of Abfs, although four GH51 *endo*-glucanases have been reported.

Currently, family GH51 is composed of 699 members, all of which are in principle enzymes that perform hydrolysis via a retaining mechanism. So far, only six structures of GH51 members have been resolved, for the Abfs from *Bifidobacterium longum* (*Bl*Abf, Uniprot **Q841V6**), *Clostridium thermocellum* (*Ct*Abf, **A3DIH0**) [2], *Geobacillus stearothermophilus* (*Gs*Abf, **B3EYN**4) [3], *Thermotoga maritima* (*Tm*Abf, **Q9WYB7**) [4], *Thermotoga petrophila* (*Tp*Abf, **A51KC8**) [5], and *Thermobacillus xylanilyticus* (*Tx*Abf, **O69262**) [6]. These structures reveal that all six enzymes contain a catalytic domain displaying a $(\beta/\alpha)_8$ fold, whose general structural organization is characteristic of the GH-A clan [7,8], with the catalytic acid base and nucleophile being localized on β4α4 loop and on β7 strand, respectively. In addition, each of these GH51 members also bears a C-terminal domain that adopts a β-sandwich fold, the functional role of which is thus far unknown. Closer examination of the six representative structures has revealed that the active site of these GH51 Abfs is formed by βα loops that are constitutive parts of the $(\beta/\alpha)_8$ barrel.

Although the six known structures of GH51 Abfs are globally similar, comparative analyses have revealed some local differences (Fig. 1). The 3D structures of *Gs*Abf and *Ct*Abf are most similar and their actives sites can be described as a funnel-shaped canyon, with the subsite -1 being localized in the narrower part [9,10]. The wider 'mouth of the funnel' constitutes a subsite +1 that can accommodate a range of aglycon groups [2,3]. Although narrower and more hydrophobic, the subsite -1 of *Tm*Abf resembles that of *Gs*Abf [4]. Crystal structures of *Gs*Abf and *Ct*Abf, trapped in their Michaelis complex with either *p*NP-α-L-Ara*f*, α-(1→5)-linked L-arabinotriose, or α-(1→3)-L-arabinoxylo-oligosaccharides (denoted as A^3X and A^3 according to Fauré *et al.* [11]) have been obtained [2,3].

These revealed that the L-Ara*f* moiety in the subsite -1 is bound by a large number of hydrogen bonds, which is in agreement with the results of the detailed analysis of the complex formed between *Tx*Abf and the pentasaccharide XA^3XX, *i.e.* the β-D-xylopyranosyl-(1→4)-[α-L-arabinofuranosyl-(1→3)]-β-D-xylopyranosyl-(1→4)-β-D-xylopyranosyl-(1→4)-D-xylopyranose [6].

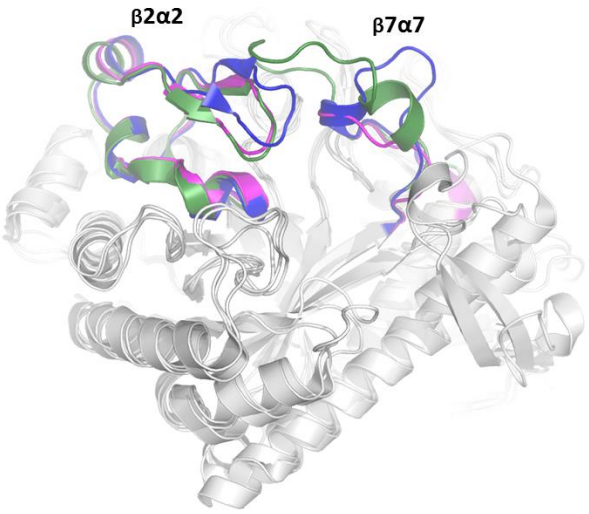


Fig. 1. Superimposition of the structures of *Tx*Abf (PDB code **2VRQ**), *Gs*Abf (**1QW8**), and *Tm*Abf (**3UG5**). The β2α2 and β7α7 loops of *Tx*Abf (blue), *Gs*Abf (green), and *Tm*Abf (magenta) are highlighted.

Examination of the 3D structure of *Tx*Abf has shown that unlike *Gs*Abf and *Ct*Abf, its active site can be described as an extended narrow groove, whose main feature is a deep pocket that constitutes subsite -1. This difference in topology is due to the presence in *Tx*Abf of a longer β6α6 loop, which is located above the deep subsite -1. Apart from subsite -1, the *Tx*Abf:XA^3XX complex reveals three other putative subsites, which have been designated as +2, +1 and +2′ [6]. Beyond the static description of the active site of *Tx*Abf, an indication of 'dynamic' active site formation was revealed by two alternative substrate-bound crystal states, which displayed open and closed conformations for the β2α2 loop (Fig. 2). In the *apo*-*Tx*Abf structure, the β2α2 loop was exclusively observed in a conformation that contributes to a rather open or accessible active site, whereas analysis of the *Tx*Abf:XA^3XX complex revealed two conformations: one identical to that seen in the *apo*-*Tx*Abf structure, while the other was characterized by a profound rearrangement of the β2α2 loop and the formation of well-defined subsites -1 and +2' [6]. Using site-directed mutagenesis and fluorescence titration spectroscopy, Paës *et al.* further probed this observation and provided data suggesting that the substrate-induced movement of the 39-residue long β2α2 loop (P71-G109) allows the formation of a catalytically-operational active site structure. Importantly, the existence of

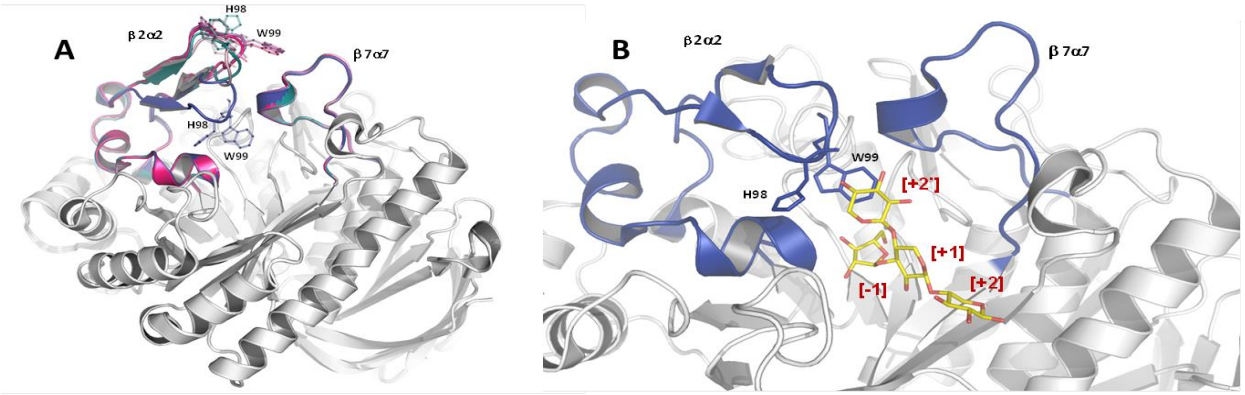


Fig. 2. (A) Superimposition of the open and closed conformations of β2α2 loop of *Tx*Abf observed in crystallographic structures (PDB codes **2VRQ** - chains A, B, and C of the *Tx*Abf:XA^{3}XX complex structure in shaded blue - and **2VRK** - chains A, B, and C of the *apo-Tx*Abf-Se structure in shaded magenta). (B) Enlarged view of *Tx*Abf active site bound to XA^{3}XX (PDB code **2VRQ**). Binding subsites and residues 98 and 99 are shown for reference.

such a loop movement would have major consequences on the active site structure, because examination of the *Tx*Abf:XA^{3}XX complex reveals that H98 and W99 amino acid residues make direct contacts with the bound substrate. More specifically, W99 establishes van der Waals (vdW) interactions with both the D-xylopyranosyl unit at subsite +2' and the L-Ara*f* moiety at subsite -1, while H98 binds the D-xylopyranosyl residue by creating a hydrogen bond in the subsite +2'. Interestingly, in other Abfs the amino acid at position 98 is not always a His, but is often replaced by an alanine and indeed, this is the case for the other five Abfs for which a structure is available. Also, it is particularly noteworthy that the superimposition of the structures of the open conformation of *Tx*Abf with the structures of *Gs*Abf, *Ct*Abf and *Bl*Abf reveals that the β2α2 loops of these latter are almost certainly unable to adopt an open conformation, because their β7α7 loops are longer than that of *Tx*Abf (Fig. 1 and 3A) [5], and consequently their spatial positioning would prevent this. In contrast, superposition of the β7α7 loops of *Tm*Abf and *Tp*Abf on that of *Tx*Abf reveals that the former are slightly smaller than the latter. However, no β2α2 loop movements have yet been described for these enzymes. Intramolecular mobility and conformational changes of flexible loops is quite common in GHs, such as amylases [12–14], amylomaltase [15], β-glucosidases [16], and xylanases [17,18]. Moreover, loop motions are often associated with substrate binding and product release, and also for the correct positioning of amino acids that are important for catalysis [19,20]. Therefore, it is likely that the putative movement of the β2α2 loop in *Tx*Abf is a crucial feature of the catalytic mechanism in this enzyme.

In this work, we have employed site directed mutagenesis to create mutants, altered at positions 98 and/or 99, which have been closely examined with regard to their ability to bind and hydrolyse *p*NP-α-L-Ara*f* and *p*-nitrophenyl β-D-galactofuranoside (*p*NP-β-D-Gal*f*). Moreover, in order to better characterize the dynamic behaviour of *Tx*Abf and the conformational changes of the β2α2 loop during substrate binding, we have carried out molecular dynamics simulations.

Results

Analysis of sequence conservation in loops β2α2 and β7α7

According to statistics drawn from the CAZy database, GH51 consists of 699 enzyme members that are produced by both microorganisms (bacteria, archaea and microbial eukaryotes) and plants. The vast majority of GH51 enzymes appear to be Abfs, although to date only 53 GH51 enzymes have been biochemically characterized. Using the protein sequence of *Tx*Abf as a query for a blast search generated an alignment consisting of 488 complete sequences, all of which were GH51 Abfs from bacteria, archaea and fungi. Further analysis of this large GH51 sequence subset revealed that in the β2α2 loop region the amino acid occupying position 99 (*Tx*Abf numbering) is highly conserved as tryptophan among the sequences and is only very infrequently replaced by phenylalanine in certain enzymes from *Bifidobacterium* sp. (Uniprot **C0BSM2**, **A1A2Q3** and **B6XT40**), *Oenococcus* sp. (**D9J2Q1** and **D9J2Q2**) and *Leconostoc mesenteroides* (**Q03XS45** and **G7VMK1**). Similarly, in very exceptional cases, W99 is also replaced by glycine (sequences from *Koribacter versatilis*, **Q1IUW7**, and *Ignisphaera aggregans*, **E0SNU3**) or threonine (*Acidobacterium* sp., **E8WVY7**). However, the amino acid at position 98 (*Tx*Abf numbering) is more variable, with alanine, histidine and asparagine being most frequent occurrences (Fig. 3B). Furthermore, using the *Tx*Abf sequence and its secondary structure assignments as a

reference, the position of the β7α7 loop on the various sequences was located and the minimum and maximum lengths of the loops were determined. In *Tx*Abf, the β7α7 loop is delimited at its N-terminal extremity by W302, which is very highly (but not totally) conserved among all of the 488 sequences examined, and is identifiable at its C-terminal extremity by a motif that is QQN (residues 316-318) in *Tx*Abf and related sequences, and E[E/D]*x*Y in sequences more related to that of *Gs*Abf (Fig. 3A). The shortest β7α7 loops, typical of *Tx*Abf and its closest homologs, were 15-17 amino acids long, whereas the longest loop was that of the Abf from *Clostridium ljungdahlii* (**D8GKA0**), measuring 38 amino acids long. Importantly, a shorter β7α7 loop appears to coincide with the presence of both the QQN motif and, frequently, the occurrence of histidine at position 98. However, while the presence of a histidine residue at position 98 appears to be incompatible with longer β7α7 loops (> 20 amino acids), it

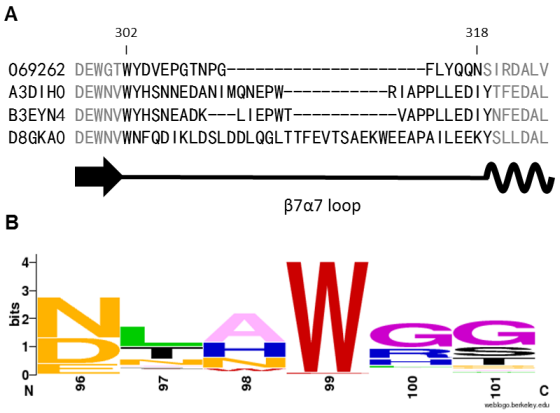


Fig. 3. (A) The β7α7 loop of *Tx*Abf (Uniprot **O69262**, *Tx*Abf numbering) compared to those of *Gs*Abf (**B3YN4**), *Ct*Abf (**A3DIH0**) and that of the the Abf from *Clostridium ljungdahlii* (**D8GKA0**). **(B)** A consensus sequence logo containing residues 98 and 99 (*Tx*Abf numbering) generated using Weblogo [30] and alignment of 488 family GH51Abf sequences.

was observed that the presence of a short β7α7 loop does not automatically imply the occurrence of histidine at position 98, and indeed other amino acids, including alanine, are found at this position.

Production and characterization of enzyme variants mutated at positions H98 and W99

Site-directed mutagenesis was employed to create four single mutants (H98A or F and W99A or F) and two double mutants (H98A/W99A and H98F/W99F), which were all successfully expressed in *E. coli* and purified using methods developed for the parental enzyme [21]. Investigation of pH dependency (Fig. S1) and thermostability (data not shown) of the different mutants revealed that all exhibited profiles similar to those of *Tx*Abf, implying that mutagenesis had not radically compromised key enzyme properties. Moreover, determination of the kinetic parameters for the hydrolysis of *p*NP-α-L-Ara*f* revealed that, compared to *Tx*Abf, all mutants displayed reduced specific activities and significantly altered K_M values, except in the case of H98F (Table 1). Likewise, all mutants displayed lowered (3-7 fold) k_{cat} values compared with *Tx*Abf. The greatest change in K_M was observed for the double mutant H98A/W99A (approximately a 20-fold increase), while W99F displayed the lowest k_{cat} value (an approximate 7-fold decrease). Overall, the impact of mutations at position 99 caused drastic effects on hydrolytic activity, with both K_M and k_{cat} values being affected. Comparing the performance constant (k_{cat}/K_M) of *Tx*Abf acting on *p*NP-α-L-Ara*f* (795 ± 67 $s^{-1}.mM^{-1}$) and on *p*NP-β-D-Gal*f* (k_{cat}/K_M = 0.557 ± 0.010 $s^{-1}.mM^{-1}$) has previously revealed that *Tx*Abf is over 1400-fold more active on *p*NP-α-L-Ara*f* [22]. However, a similar comparison of the values of k_{cat}/K_M of each mutant described in this study on the two substrates showed that all mutants were less specific towards *p*NP-α-L-Ara*f*, although the implications of this observation are unclear, because the values of k_{cat}/K_M

Table. 1. Kinetic parameters for the hydrolysis of *p*NP-α-L-Ara*f* and *p*NP-β-D-Gal*f* by *Tx*Abf and β2α2 loop mutants thereof. SA, specific activity at pH 5.8 and 60 °C.

Enzyme	*p*NP-α-L-Ara*f*				*p*NP-β-D-Gal*f*
	SA ($IU.mg^{-1}$)	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}.mM^{-1}$)	k_{cat}/K_M ($s^{-1}.mM^{-1}$)
Wild-type	615 ± 51	575 ± 48	0.72 ± 0.04	795 ± 67	0.557 ± 0.010
H98A	205.7 ± 5.9	192 ± 6	1.52 ± 0.06	126 ± 4	0.296 ± 0.020
H98F	212.8 ± 9.5	199 ± 9	0.87 ± 0.01	228 ± 10	0.262 ± 0.004
W99A	146.0 ± 3.8	137 ± 4	9.1 ± 0.4	15.1 ± 0.4	0.220 ± 0.012
W99F	81.9 ± 1.6	77 ± 2	4.2 ± 0.3	18 ± 4	0.277 ± 0.009
H98A/W99A	165.0 ± 24.7	154 ± 23	15 ± 5	9.8 ± 1.5	0.159 ± 0.002
H98F/W99F	103.1 ± 0.8	97 ± 1	3.4 ± 0.2	28.2 ± 0.2	0.168 ± 0.004

for all mutants catalysing the hydrolysis of *p*NP-β-D-Gal*f* were very low, in the range 0.2 to 0.3 $s^{-1}.mM^{-1}$ (Table 1).

Measuring enzyme-substrate interactions

To further characterize the different mutants and their interactions with *p*NP-α-L-Ara*f* and *p*NP-β-D-Gal*f*, saturation transfer difference (STD)-NMR spectroscopy [23–26] and isothermal titration calorimetry (ITC) were employed. To ensure that hydrolysis of the substrates would not interfere with these experiments, the different enzymes were inactivated by mutating the catalytic acid/base residue (E176) to alanine, thus generating *Tx*Abf E176A hereafter denoted *Tx*Abf†. After mutagenesis, reactions using *p*NP-α-L-Ara*f* and the different mutant enzymes revealed very low (*e.g.* 0.06 ± 0.01 $IU.mg^{-1}$ for *Tx*Abf†) or unmeasurable activity (for other mutants) in the conditions employed for STD-NMR and ITC, meaning that these experiments could be conducted without the risk of artefacts due to hydrolysis.

STD-NMR is a technique in which the saturation of protein proton resonances is transferred to neighbouring substrate protons, thus reflecting protein-substrate interactions and substrate affinity. Likewise, STD-NMR was used in this study to map the interactions that characterize substrate binding to the impotent *Tx*Abf† mutant and the variants thereof. First, using *p*NP-α-L-Ara*f* and the reference enzyme, *Tx*Abf†, it was shown that the protons H-2, H-3, and H-4 of the L-Ara*f* unit displayed relative STD intensities between 82 and 100% (the STD effects, expressed as percentages, are relative to the highest proton H-4 signal), while the anomeric proton, H-1, the H-5 protons and the aromatic protons of the *p*NP moiety were characterized by weaker STD intensities, ranging from 30 to 59% (Fig. 4). This trend is globally consistent with the previously reported strong binding of the L-Ara*f* unit in the subsite -1, which is mediated by seven hydrogen bonds and several vdW interactions that are expected to influence neighbouring protons [6]. Moreover, using STD-NMR, it was possible to estimate a value of 0.16 ± 0.06 mM for the dissociation constant (K_d).

The STD-NMR investigation of interactions of all the mutants with *p*NP-α-L-Ara*f* revealed that the STD intensities of the aromatic protons were generally lower, displaying intensities that were 11 to 44% of those of the reference enzyme (Fig. 5). Nevertheless, the STD effects for the various protons observed in each of the mutants was very well preserved and correlated well with the relative binding energies. For the single mutants *Tx*Abf† H98A and *Tx*Abf† H98F, STD intensities corresponding to the protons of the L-Ara*f* moiety appeared to be unaltered when compared to *Tx*Abf†, whereas the mutants *Tx*Abf† W99A and W99F displayed decreased STD intensities for all of the sugar-associated protons. This was also the case for double mutants *Tx*Abf† H98A/W99A and *Tx*Abf† H98F/W99F. This correlates well with the modifications to K_M for the various mutations, the largest effects being observed with the mutations at position W99 (Table 1). Finally, when performing STD-NMR experiments using *p*NP-β-D-Gal*f* as the substrate for *Tx*Abf†, and variants thereof, no significant STD effects were measured, which is consistent with previous results that have shown that *p*NP-β-D-Gal*f* is a very poor substrate for *Tx*Abf [22].

Using ITC it was also possible to investigate binding of *p*NP-α-L-Ara*f* to *Tx*Abf† (Fig. S2), revealing a K_d value of 0.32 ± 0.09 mM, quite similar to the one found by NMR. However, another calorimetric experiment using *p*NP-β-D-Gal*f* as substrate failed to reveal any significant interactions with *Tx*Abf†, consistent with the findings of STD-NMR. Data were fitted for a single-site binding model (N = 1.12 ± 0.02 site), in good agreement with the STD-NMR experiments, which only revealed interactions within the active site of *Tx*Abf. The positive enthalpy (3.73 ± 1.16 $kcal.mol^{-1}$) and entropy (29.2 ± 4.7 $cal.mol^{-1}$) changes indicate that the binding at 283K is driven by entropy. The comparison of the free energy associated with the formation of the *Tx*Abf:*p*NP-α-L-Ara*f* complex showed that the data obtained using STD-NMR or ITC experiments (ΔG = -4.92 ± 0.15 and -4.53 ± 0.17 $kcal.mol^{-1}$, respectively) performed at 283K were in good agreement.

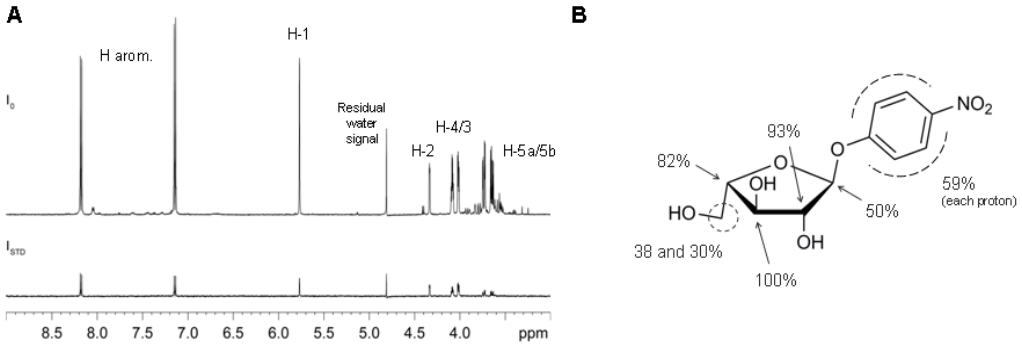


Fig. 4. (A) Reference (I_0) and STD-NMR (I_{STD}) spectra of *p*NP-α-L-Ara*f* binding to *Tx*Abf† mutant. (B) STD-NMR ligand fingerprinting of *p*NP-α-L-Ara*f* bounded to *Tx*Abf† mutant. The values of the STD effects were normalized relative to the highest proton signal (H-4) of the bound sugar moiety.

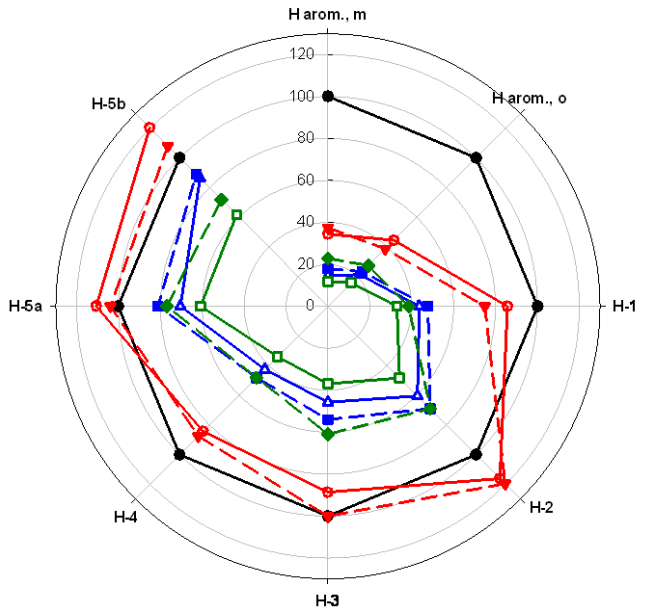


Fig. 5. Standardized STD effects (based on inactivated *Tx*Abf† as reference) of *p*NP-α-L-Ara*f* protons with inactivated *Tx*Abf† (●, black) and β2α2 loop mutants thereof H98A† (○, solid red), H98F† (▼, dashed red), W99A†, (△, solid blue), W99F† (■, dashed blue), H98AW99A† (□, solid green) and H98FW99F† (◆, dashed green).

In silico investigation of the impact of β2α2 mutations on dynamic behaviour and substrate binding

To investigate at the atomic level the impact of mutations on the dynamics of the β2α2 loop of *Tx*Abf and on substrate specificity, large scale molecular dynamics (MD) simulations (40 ns) in explicit water were performed both on *Tx*Abf and on the mutant *Tx*Abf W99A, in the absence or presence of *p*NP-α-L-Ara*f*. These MD simulations enabled us to acquire for the first time detailed insight into the conformational rearrangements that occur in *Tx*Abf, and in its mutant derivative *Tx*Abf W99A, upon binding to the substrate *p*NP-α-L-Ara*f*.

To monitor the stability of the protein structures during the MD simulations, the time-dependent evolution of the RMSD of backbone atoms was calculated after least squares fitting (Fig. S3A). Accordingly, it was clear that the variations of backbone RMSD values were very similar for both the wild-type and mutant *Tx*Abfs in their *apo* form, with both showing an increase of ~ 0.7 Å within the first 8 ns of the simulation, before becoming stable. The calculated B-factors indicated a similar trend for both enzymes, with two highly flexible regions corresponding to β2α2 (residues 71-109) and β7α7 (residues 302-318) loops respectively, which is consistent with current knowledge concerning these loops (Fig. 6A). In both *Tx*Abf and *Tx*Abf W99A, conformational rearrangements generally led to greater exposure of their active sites. However, in spite of these global similarities, closer inspection of the simulations revealed some significant differences in the dynamic behaviour of the β2α2 and β7α7 loops in these enzymes, as revealed by inspection of the RMSD profiles (Fig. S3B, C). At the beginning of the MD simulation, *Tx*Abf adopted a so-called 'closed' state in which the β2α2 and β7α7 covered the active site, with W99 located at the tip of the β2α2 loop (Fig. 2), apparently playing a key role in the stabilization of the closed conformation by establishing a hydrogen bond with E28 and forming a hydrophobic core with L30 (subsite -1) and two residues (W302 and L314) of the β7α7 loop, thus forming a pocket that is narrower than the one observed in the 'open' conformation. Simultaneously, H98 made a hydrogen bond with the backbone of the residue C74. As the simulation proceeded, substantial conformational rearrangements occurred within *Tx*Abf. Notably, these alterations involved the opening of the β2α2 loop, reaching a conformational state that highly resembled the *apo*-*Tx*Abf x-ray structure. The variability shown in the conformation of the β2α2 loop is in agreement with the high crystallographic B-factor values observed for this loop.

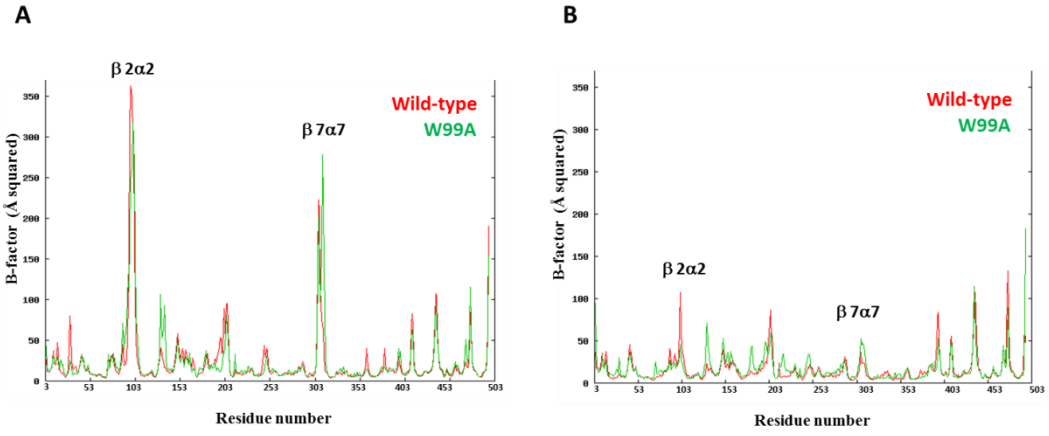


Fig. 6. Calculated B-factors of wild-type *Tx*Abf (red) and W99A mutant (green) from MD simulations. (A) *Apo* enzymes. (B) Complexes with *p*NP-α-L-Ara*f*.

During this movement, a series of events occurred that involved *(i)* the rupture of hydrogen bonds between E28 and W99, and between H98 and C74, accompanied by *(ii)* the disruption of the hydrophobic core formed by W99 (β2α2 loop), L30 (β1α1 loop), L314 and W302 (both from the β7α7 loop), *(iii)* the movement of the β1α1 loop toward β2α2 loop, which in turn brought L30 close to F75 residue, and *(iv)* the formation of a novel stacking interaction between W99 and H98 aromatic side chains.In the next step of the simulation, the outward trajectory of bulky W99 and H98 residues was facilitated by a series of conformational jumps that occurred within the β7α7 loop (Fig S4 and Supplementary movie 1), leading to the disruption of subsite +2' and the opening of the active site cleft.Despite these modifications, the motion of β7α7 loop was generally more constrained during the simulation (RMSD increase of less than 1Å between starting and end of simulation) than that of the β2α2 loop (RMSD increase of ~ 2 Å between starting and end of simulation), mainly because of the presence of two prolines, P308 and P311, which rigidify the loop backbone conformation. Regarding the mutant *Tx*Abf W99A, the absence of the tryptophanyl side chain greatly reduced the steric difficulties encountered during the closing ↔ opening movement of the β2α2 loop, as shown by the comparison of the RMSD profiles of *Tx*Abf and *Tx*Abf W99A (Fig. S3B). Basically, this is because the original strong interaction between W99 and E28 is abolished in *Tx*Abf W99A, which in turn leads to a rather poor definition of the subsite +2' (Fig. S5 and Supplementary movie 2). Moreover, the conformational rearrangements that occurred in *Tx*Abf W99A led the β2α2 loop to adopt a much more compact conformation when compared to the β2α2 loop of *Tx*Abf.

When a bound *p*NP-α-L-Ara*f* was introduced into the active site of *Tx*Abf and *Tx*Abf W99A, the mobility of the β2α2 and β7α7 loops was drastically reduced along MD simulations (Fig. 6B), with *p*NP-α-L-Ara*f* being tightly bound to the enzymes through a network of hydrogen bonds involving residues such as E28, C74, N175, E176, Y242, Q347 and stacking interactions with W99, W248 and W302 (Fig. 7), in agreement with the findings of Paës *et al.* [6]. Importantly, the MD simulation revealed that E28, which hitherto (in *apo*-*Tx*Abf) made a hydrogen bond with W99, was also hydrogen bonded to the substrate, thus allowing the W99 indolyl group to closely interact with the L-Ara*f* moiety. The distances between the anomeric carbon and the glycosidic oxygen of the *p*NP-α-L-Ara*f* with the catalytic dyad E298 (nucleophile) and E176 (general acid/base) were measured as 3.1 Å and 2.7 Å for the C-1 (L-Ara*f*)…O(E298) and O-1 (L-Ara*f*)…O(E176) distance respectively, for both *Tx*Abf and *Tx*Abf W99A. However, detailed analysis of the first 2 ns of the MD simulation revealed a critical difference in the behaviour of *p*NP-α-L-Ara*f* within the active site of the two enzymes. In wild-type enzyme, MD simulation induced the L-Ara*f* moiety to explore conformations that generally maintained distances from E176 and E298 residues that are compatible with catalytic competency (Fig. S6A, B), although during the course of the MD simulation

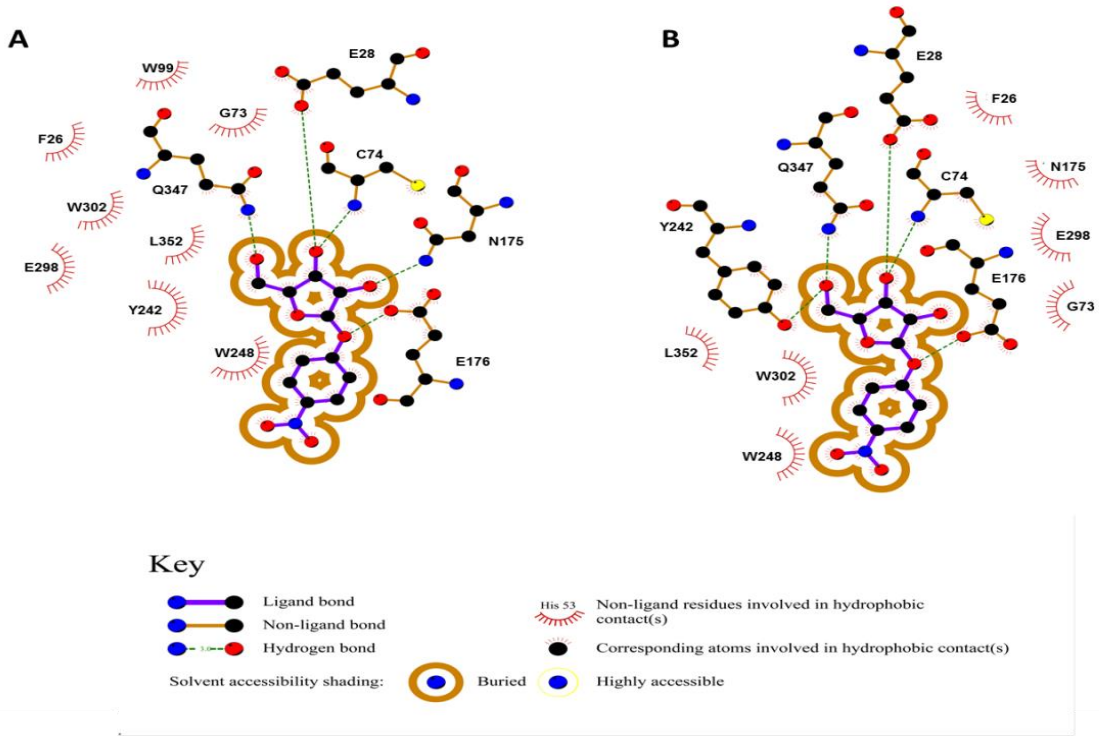


Fig. 7. Schematic representation of *Tx*Abf interactions with *p*NP-α-L-Ara*f*. (A) wild-type *Tx*Abf. (B) W99A mutant.

the bonding network that fixes *p*NP-α-L-Ara*f* in the active site was significantly altered, in particular in the case of *Tx*Abf W99A. While the *p*NP moiety was stabilized in the subsite +1 of *Tx*Abf through vdW interactions with W248, Y242 and W302, it moved away from the subsite +1 in *Tx*Abf W99A, establishing a hydrogen bond interaction with backbone amide of C74. As already observed for *apo-Tx*Abf, during the simulation H98 (which forms part of subsite +2') was hydrogen bonded to the backbone amide of C74, and thus never interacted with *p*NP-α-L-Ara*f*. However, in *Tx*Abf W99A this hydrogen bond was absent, because the H98 imidazole group shifted towards the bound substrate, stacking with the *p*NP moiety and establishing a hydrogen bond with the glycosidic oxygen O-1.

Along the MD trajectory, the L-Ara*f* ring was highly flexible, interconverting between two main conformations of the furanose ring, $_1$E and ^{4}E (Fig. S6C). When bound to *Tx*Abf, the five-membered ring mainly adopted a $_1$E conformation, which was strongly stabilized through multiple hydrogen bonds that were formed during the course of simulation (OH-2…N175, OH-3…E28, OH-3…W99, OH-3…C74, OH-5…W99, OH-5…Q347, OH-5…E298, OH-5…Y242, O-5…W99 and O-1…E176). At the same time, the *p*NP moiety stacked with W248 and W302 residues. Conversely, when bound to *Tx*Abf W99A the L-Ara*f* moiety more frequently adopted a ^{4}E conformation, which was stabilized through an alternative network of hydrogen bonds that were formed over time OH-2…N175, OH-2…E176, OH-3…E28, OH-5…E298, OH-5…Y242, OH-5…Q347, O-5… H98, O-1…H98 and O-1…E176.

To assess the binding free energies of selected complexes, a Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) protocol was used. The theoretical free energy of binding, which includes the entropic contribution for *Tx*Abf:*p*NP-α-L-Ara*f* complex (-6.5 kcal.mol^{-1}) was quite consistent with experimental data. To identify the amino acids from the β2α2 and β7α7 loops that are most involved in the binding of *p*NP-α-L-Ara*f*, we estimated their individual contribution to the binding free energy. This revealed that the major contributors were R69, C74, Q347, G73, Y242, W302, W99, W178, E298, D297 and E28 residues.

Discussion

A previous study by Paës *et al.* [6], clearly indicated that, in the *Thermobacillus xylanilyticus* GH51 Abf, W99 is a functionally-important residue that is localized on a mobile β2α2 loop. Moreover, this study implied that an open-close movement of the β2α2 loop is an intrinsic feature of catalysis, which relocates residue W99 into subsite -1, thus creating a catalytically-competent active site. Therefore, in the present study, we set out to provide new insight into substrate binding, notably to acquire a clearer understanding of the roles played by H98 and W99 and to better define dynamics involved in the open-close movement of the β2α2 loop.

Analysis of protein sequences of GH51 family members revealed that the β2α2 loop contains a characteristic motif in which position 99 (*Tx*Abf numbering) is almost exclusively occupied by a tryptophan residue (Fig 3B), which is consistent with the fact that this residue appears to be an indispensable element of the subsite -1. Likewise, sequence analysis showed that position 98 was less conserved, but indicated a possible correlation with the length of the β7α7 loop. This is noteworthy because previously Paës *et al.* postulated that the presence of long β7α7 loops in enzymes such as *Gs*Abf and *Ct*Abf would prevent the movement of the β2α2 loop, locking the latter in a closed conformation, which in turn might well have important implications for substrate binding, with the open-close motion possibly providing the enzyme with the means to better adapt to its substrate through an induced fit mechanism. Similarly, structural analysis of *Tx*Abf bound to XA^{3}XX indicated that H98 is a major determinant of a subsite +2' that appears to be absent or less well-defined in enzymes such as *Gs*Abf that are characterized by an alanine at position 98. In this study through sequence alignment of 488 GH51 Abfs, we have shown that there is a coincidental, but non-exclusive relationship between the presence of histidine at position 98 and a short β7α7 loop. This implies that the presence of a fully defined subsite +2' could be somehow linked to the open-close mechanism, although the meaning of this remains unclear. Moreover, reduced or abolished binding in subsite +2' (*i.e.* possibly when a residue such as alanine is present at position 98) does not mean that the open-close movement is impossible, since many GH51 enzymes characterized by an alanine at position 98 also display short β7α7 loops.

From a catalytic point of view, replacement of H98 by alanine or phenylalanine had less impact on catalysis, with k_{cat} being most affected, although H98A also displayed an approximate 2-fold higher value for K_M. This is unsurprising, because the substrate used to assay activity, *p*NP-α-L-Ara*f*, is not expected to bind in subsite +2'. The results of STD-NMR further confirm this interpretation, indicating that H98 does not in any way intervene in subsite -1. Nevertheless, it is interesting to note that in the case of H98A, STD-NMR revealed alterations in the binding of the aglycon moiety in subsite +1 binding, even though the MD simulation did not reveal any direct interactions between H98 and *p*NP. However, MD simulation did suggest a hydrogen bond between H98 and C74, a residue that appears to be at the interface between subsites +1 and -1, and thus could be in close contact with both the glycon and aglycon moieties. Therefore, although H98 does not play a direct role in the binding of small substrates such as *p*NP-α-L-Ara*f*, the dynamics of the β2α2 loop means that it is a determinant of the topology of the active form of the catalytic site. Therefore, in enzymes such as *Gs*Abf, *Ct*Abf, *Bl*Abf, *Tm*Abf and *Tp*Abf, it is expected that the absence of histidine at position 98 might imply alternative interactions, in subsite

+1 binding. Indeed, this is probably the case, because there is a good correlation between H98 and the presence of cysteinyl side-chains at positions 74 and 180. In Abfs displaying alanine at position 98, position 74 is mostly occupied by an asparagine residue and the equivalent of residue 180 is absent.

Regarding W99, mutation to alanine or phenylalanine led to significant alterations in the values of both K_M and k_{cat}. While this is unsurprising for the mutant W99A, which almost certainly displays a highly impaired subsite -1, it is more surprising for W99F, since one can argue that in this mutant the hydrophobicity and the steric occupation of the side-chain are at least partially conserved. However, the MD simulations show that in substrate-bound *Tx*Abf, W99 forms a hydrogen bond with E28, which in turn is hydrogen bonded to the sugar moiety bound in subsite -1, which would explain why the substitution by phenylalanine at position 99 leads to poorer substrate binding. Nevertheless, the mutant W99F also displays the lowest value for k_{cat}, which is rather surprising. However, the results of MD simulation revealed that in the case of W99A, an alternative network of hydrogen bonding involving H98 is put into place to stabilize sugar binding in subsite -1. In the case of W99F, presumably the steric occupation of phenylalanine does not permit this, and thus both sugar binding and ring distortion are diminished. Interestingly, in some other GH51 Abfs, a phenylalanine residue is present at position 99. Therefore, one might assume that in these enzymes, a different hydrogen bonding network is involved in subsite -1 binding.

Overall, the data revealed by the MD simulations strongly support the existence of a functionally-important open-close movement of the β2α2 loop and fit well with the previously proposed induced-fit mechanism. Concerning the thermodynamic data, the experimental work presented here clearly indicated that substrate binding is driven by favorable entropy changes, which is often indicative of a process involving release of water molecules and the burial of hydrophobic groups. The MD simulation provides further insight into this phenomenon, because it reveals that the closing movement of the β2α2 loop displaces W99 from a highly solvent exposed position to a buried one. Moreover, it is highly probable that substrate binding is also accompanied by the expulsion of bulk water molecules from subsite -1.

A striking feature of the STD experiments is that all the mutations led to a strong decrease in the STD signals for the aromatic protons (Fig. 5), even for mutations at H98 that did not significantly alter sugar binding in subsite -1. Manifestly, this observation cannot be explained by decreased substrate affinity, but rather by a slight relocalisation of the ligand in the subsite -1, which leads to diminution of the exposure of the aromatic protons to those of the protein residues. By contrast, mutations affecting W99 led to reduced binding affinity and this probably the main mechanism for the observed reduced STD effect.

A previous STD-NMR study performed on *Ct*Abf only revealed weak STD intensities for H-2, H-3, and H-4 of the L-Ara*f* moiety bound in subsite -1 [27]. This is surprising and is in contradiction with a structural model of the Michaelis complex of *Ct*Abf with A^3X, which indicates that the L-Ara*f* moiety was strongly stabilized in the subsite -1 through multiple hydrogen bonds: OH-2 of the L-Ara*f* moiety interacts both with the oxygen atom (2.6 Å) of the nucleophile E292 (E298 in *Tx*Abf) and with N172 (N175 in *Tx*Abf), OH-3 interacts both with the side chain of E27 (E28 in *Tx*Abf) and with the N72 backbone amide (instead of the C74 in *Tx*Abf), while OH-5 interacts both with the side chains of Y244 and Q352 (Y242 and Q347 in *Tx*Abf, respectively). The STD-NMR data obtained in our study were much more compliant with the structural data, with strong STD effects being measured for H-2, H-3, and H-4 of the *Tx*Abf-bound L-Ara*f* moiety. Moreover, the experimental values for K_d obtained in this work were similar (0.16 ± 0.06 and 0.32 ± 0.09 mM) and consistent with both the values of K_M (0.72 ± 0.04 mM) and the previously measured K_d value (0.21 ± 0.02 mM) for the binding of XA^3XX [6]. Nevertheless, it is noteworthy that the STD intensity attributed to the H-5 was weaker, suggesting either that this proton is less involved in substrate binding, or that the hydroxymethyl moiety remains mobile in the bound state (internal motion of substrate part being a known cause for the reduction of STD effects), or simply that *p*NP-α-L-Ara*f* is not bound in exactly the same way as XA^3XX. Interestingly, a mobile hydroxymethyl moiety would be quite consistent with the observation that the β-D-fucofuranosyl moiety of *p*NP-β-D-fucofuranoside can be accommodated in subsite -1 without inducing drastic changes in K_M [28]. However, the failure of STD-NMR and ITC analysis to reveal any STD effects or binding when *p*NP-β-D-Gal*f* was employed as the substrate further confirms that the presence of the extra hydroxymethyl group (C-6) is extremely detrimental for sugar binding in subsite -1.

In conclusion, this study has provided new insight into the role of two residues in *Tx*Abf that are localized on a mobile loop whose movement constitutes a key feature of substrate binding and catalysis. Nevertheless, this work further highlights an enigma, in as much that certain GH51 Abfs probably cannot adopt the open form of the β2α2 loop, because of the presence of a longer β7α7 loop that obstructs and thus radically limits β2α2 loop movement. Taking into account the fact that the Michaelis complexes of both *Tx*Abf and *Gs*Abf (representative of GH51 Abfs that possess longer β7α7 loops) with L-Ara*f* are highly similar, what is the benefit of β2α2 loop movement? Rather than focussing on the subsite -1, the answer to this question might be linked to the formation of a well-defined subsite +2', which may help certain Abfs to better recognize longer branched oligosaccharides. However, to establish this, it will be necessary to carefully compare the hydrolytic activity of several GH51Abfs in standardized conditions, using an appropriate set of substrates.

Materials and methods

Bioinformatics analysis of β2α2 and β7α7 loops

To perform bioinformatics analyses on sequences encoding family GH51 members, the sequence encoding *Tx*Abf (Uniprot **O69262**) was used a query to perform a blast analysis (http://www.uniprot.org/blast/) using mostly default settings, but requiring a report of the top 500 hits. This analysis generated a group of 500 sequences encoding GH51 Abfs. After eliminating incomplete sequences, 488 sequences were retained for alignment, which was performed using ClustalW embedded in the BioEdit v7.0.9.0 sequence alignment analysis software [29]. To localize the approximate positions of β2α2 and β7α7 in each of the 488 sequences, the *Tx*Abf protein sequence was used as reference, with secondary structure elements being visualized on the primary sequence using the sequence display function on the PDB database website (http://www.rcsb.org/pdb/) and author-defined secondary structure for the PDB file **2VRK**. To define a consensus sequence logo, the web-based service (http://weblogo.berkeley.edu/) was used [30].

Mutagenesis, protein expression, and purification

The plasmids containing the complete coding sequence of *Tx*Abf (GenBank accession number **CAA76421.2**), pET21- or pET24-*Tx*Abf (original pET vectors from Novagen), were used as a template for *in vitro* mutagenesis using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Mutants were obtained with the following primers (bold lettering indicates mismatch positions):

H98A	5'-ATGGTGAACACG**GCG**TGGGGCGGCGTC-3'
H98F	5'-GGATGGTGAACACG**TTC**TGGGGCGGCGTCATCG-3'
W99A	5'-GTGAACACGCAC**GCG**GGCGGCGTCATC-3'
W99F	5'-GGATGGTGAACACGCAC**TTT**GGCGGCGTCATCG-3'
H98A/W99A	5'-ATGGTCAACACG**GCGGCG**GGCGGCGTCATC-3'
H98F/W99F	5'-GGATGGTGAACACG**TTCTTT**GGCGGCGTCATCG-3'
E176A	5'-GGCGTCGGCAACG**CC**AACTGGGGCTGC-3'

The successful introduction of mutations was confirmed by DNA sequencing (GATC Biotech). Expression and subsequent purification of wild-type and mutated recombinant *Tx*Abf, produced in *E. coli* BL21(DE3), were performed as previously described [21].

Substrates

The substrates, *p*NP-α-L-Ara*f* and *p*NP-β-D-Gal*f*, were synthesized as previously described [22,31].

Kinetic Studies

The activity of *Tx*Abf and mutants thereof was determined using a discontinuous enzyme assay. Reactions were performed in triplicate at 60 °C in buffered conditions (50 mM sodium acetate buffer, pH 5.8) using a *p*NP-α-L-Ara*f* or *p*NP-β-L-Gal*f* (5 mM) and 1 $mg.mL^{-1}$ BSA. The total reaction volume was 400 µL and reactions were performed over a 10-20 min period, removing samples at regular intervals. Hydrolysis reactions were terminated by addition of 250 µL of sodium carbonate (1 M) to 50 µL of reaction mixture. After, absorbance at 405 nm was measured using a VersaMax microplate spectrophotometer and the quantity of *p*NP released was calculate using an appropriate standard curve prepared using pure *p*NP. Negative controls containing all of the reactants except the enzyme were used to correct for spontaneous hydrolysis of the substrates. One unit (IU) of enzyme specific activity corresponds to the amount of enzyme (in mg) releasing one µmol of *p*NP per minute. To determine Michaelis-Menten kinetic parameters K_M and k_{cat}, or directly k_{cat}/K_M, appropriate enzyme quantities (*i.e.* nM for *p*NP-α-L-Ara*f* and µM for *p*NP-β-L-Gal*f*) were used in combination with suitable substrate concentrations (*i.e.* 0.1-18 and 0.5-2 mM for *p*NP-α-L-Ara*f* and *p*NP-β-D-Gal*f*, respectively), thus allowing the accurate determination of the initial rate of reactions. All data were processed using using SigmaPlot 10.0 software.

STD-NMR experiments

All buffered enzymatic solutions were exchanged against D_2O (99.90%) using two cycles of 10-fold dilution, followed by concentration on Amicon Ultra (regerenated cellulose 10 kDa, Millipore) at 4240 x g during 10 min. Samples were prepared in 600 µL of D_2O containing 100 µM of *p*NP-α-L-Ara*f* or 500 µM of *p*NP-β-D-Gal*f* with a molar ratio (protein:ligand) of (1:100) for *p*NP-α-L-Ara*f* and (1:500) for *p*NP-β-D-Gal*f*. STD-NMR experiments were performed at 283 K with a Bruker Avance 600 spectrometer equipped with a cryoprobe TCI as follows [32]. The proteins were saturated on resonance at -0.4 ppm on methyl signals, and off resonance at 30 ppm, using selective 50 ms Gaussian-shaped pulses, at a radiofrequency field amplitude of 86 Hz, with a 100 ms delay between each pulse. The total duration of the saturation time was 2 s. A WATERGATE sequence was used to suppress the residual water signal [33]. A similar experiment was performed without enzyme in order to verify the selectivity of saturation and the efficiency of the signal subtraction that was used to obtain the difference spectrum conditions. Intensities of all STD effects (I_{STD}) were calculated by integrating the respective ^1H NMR signals, followed by standardization with the reference signal I_0. The ratio I_{STD}/I_0 was normalized using the largest STD effect as the reference (*i.e.* all STD effects were reported as a percentage of the STD effect of the H-4 proton of the L-Ara*f* unit, which was arbitrarily set to 100%). In order to compare the different mutants, STD effects of protons exhibiting identical chemical shifts were standardized for each mutant, using the STD data obtained with *Tx*Abf†. All data were acquired and processed using Topspin v2.1 software (Bruker).

Determination of K_d

The mutant *Tx*Abf[†] was titrated with *p*NP-α-L-Ara*f* and the general setup for STD-NMR experiments was followed with 1 µM of enzyme and a ligand excess (ligand:enzyme molar ratio) ranging from 50 to 1500 µM. For each ratio STD effects with (I_{STD}) and without enzyme ($I_{STD,we}$) were measured and A_{STD} was calculated as follow: $A_{STD} = [(I_{STD} - I_{STD,we}).\text{ligand excess}] / (I_0 - I_{STD,we})]$ is the STD effect measured when there is no enzyme. For each proton, A_{STD} values were plotted against substrate concentrations and the resulting curves were fitted with a Michaelis-Menten like model: $A_{STD} = [\alpha_{STD}.[S] / (K_d + [S])].\text{ligand excess}$, with $\alpha_{STD} = A_{STD,max}$ and K_d the dissociation constant [34].

ITC experiments

ITC experiments were performed on an ITC200 instrument (Microcal, GE Healthcare) and data were acquired and analyzed using fully automatized features in Microcal Origin Software. To ensure minimal buffer mismatch, samples of protein and ligand were prepared in the same physiologically relevant buffer (Hepes 50 mM, pH 7). Experiments were conducted at 10 °C with protein concentration of 100-200 µM in the thermostatic cell. Experiments consisted of 20 x 2 µL injections of ligand (*p*NP-α-L-Ara*f* or *p*NP-β-D-Gal*f*, 1-2 mM) from the syringe (initial delay of 60 s, duration of 4 s, spacing of 180 s) into the protein containing thermostatic cell. The corrected binding isotherms were fitted using the least-squares regression method to obtain association constant (K_a), stoichiometry (N), the enthalpy (ΔH), and entropy (ΔS) changes associated with the ligand binding.

Molecular modelling procedures

To investigate at the atomic level the impact of mutations on the dynamics of the β2α2 loop of *Tx*Abf and on substrate specificity, large scale molecular dynamics (MD) simulations (40 ns) in explicit water were performed both on *Tx*Abf and its *Tx*Abf W99A mutant, in the absence or presence of *p*NP-α-L-Ara*f* using as the starting model the high-resolution crystal structures of *Tx*Abf (PDB code: 2VRQ, *i.e.* chain A in closed conformation) [6]. The missing hydrogen atoms of proteins were added using the *tleap* program in AMBER11 software package [35]. *p*NP-α-L-Ara*f* was manually docked in the active site of *Tx*Abf and its W99A mutant using as reference the XA^3XX pentaccharide (PDB code: **2VRQ**). In the starting conformation, the L-Ara*f* moiety was placed in the subsite -1 while the *p*NP ring occupied the subsite +1. The *p*NP-α-L-Ara*f* was pre-processed through the *Gaussian 03* program and the restrained electrostatic potential (RESP) fitting technique to determine it partial atomic charges. The molecular all-atom ff03 [36,37], the carbohydrate GLYCAM06 [38] and the general AMBER *gaff* force fields were used to establish the potential of the wild-type *Tx*Abf, its W99A mutant and the ligand, respectively. To obtain a neutral charge of the simulated systems, a number of counter-ions were included. Each enzyme or enzyme:ligand complex together with the counter-ions was solvated with TIP3P water molecules, using the rectangular parallelepiped box with a minimum distance of 0.12 nm between the solute and the simulation box edge. All molecular systems were then subjected to several energy minimization cycles (steepest descent + conjugated gradient) using the AMBER11 suite of programs [35]. Following the minimization steps, MD simulations were carried out upon a slow heating to 300 K under constant volume over a period of 80 ps. At the final required temperature (300 K), the system was equilibrated under constant volume condition over 10 ps and then it was turned on constant pressure (1 bar) condition over 60 ps. Atomic positions of the protein backbone were first restrained using a harmonic potential during the minimization schedule. The force constant was then progressively diminished until a final unrestrained minimization step. Harmonic constraints applied on the ligand were maintained until the end of the heating phase and afterwards, they were gradually removed along the equilibration in the isothermal-isobaric ensemble. The final production phase of simulations was then carried out for a total of 40 ns at constant temperature (300 K) and pressure (1 bar) conditions. The temperature and pressure were controlled using Langevin thermostat [39] and Berendsen barostat [40] with a collision frequency of 2 ps^{-1} and pressure relaxation time (2 ps). Long-range electrostatic forces were handled by using the particle-mesh Ewald method [41]. The time step of the simulations was 2.0 fs and the SHAKE algorithm was used to constrain the lengths of all chemical bonds involving hydrogen atoms to their equilibrium values [42]. The resulting trajectories were analyzed using the Ptraj module of the AMBER 11 package. The RMSD was calculated for the protein backbone atoms using least squares fitting. Atomic positional fluctuations (Δr_i^2) of protein backbone were calculated. A mass-weighted average value was then calculated for each residue. These parameters are related to the B-factors through the following relationship: $B_i = \frac{8\pi^2}{3} < \Delta r_i^2 >$.

The simulated B-factors were calculated using the coordinates of the 40 ns trajectories.

Puckering parameters (Q, *Φ*) were calculated using the definition given by Altona *et al.* [43] for the first 2ns of MD simulations. MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) and entropy calculations were performed to estimate the binding free energy ΔG_{bind} between *p*NP-α-L-Ara*f* and *Tx*Abf as well as its W99A mutant, as follows:

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{solv} - T\Delta S$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdW}$$

$$\Delta G_{solv} = \Delta G_{PB} + \Delta G_{SA}$$

where ΔE_{MM}, ΔG_{solv}, $-T\Delta S$ are the changes of the gas phase MM energy, the solvation free energy and the conformational entropy upon binding, respectively. ΔE_{MM} includes $\Delta E_{internal}$ (bond, angle and dihedral energies), electrostatic $\Delta E_{electrostatic}$ and vdW energies ΔE_{vdW}. ΔG_{solv} is the sum of electrostatic

solvation energy (polar contribution), ΔG_{PB} and non-electrostatic solvation component (non-polar contribution), ΔG_{SA}. The electrostatic contribution to the solvation free energy (ΔG_{PB}) was estimated by solving the linear Poisson-Boltzmann equation for each solute configuration using the *pbsa* program included in AmberTools [35]. The non-polar contribution was determined on the basis of solvent accessible surface area (SASA) using the LCPO method [44] implemented within *sander*. For the calculations of ΔE_{MM}, ΔG_{PB} and ΔG_{SA} 6250 snapshots evenly extracted from the MD trajectory of complex from 0 to 5 ns were used. The conformational entropy change -$T\Delta S$ was calculated by normal-mode analysis (NMA) of the harmonic vibrational frequencies of 32 configurations considering the complete protein using the AMBER11 *mmpbsa_py_nabnmode* program [35]. To this purpose, each configuration was energy minimized with a Generalized-Born solvent model using of 500 000 steps and a target root-mean-square (rms) gradient of 10^{-4} kcal.mol^{-1}.Å^{-1}. All these MM-PBSA and entropy calculations were performed using the MMPBSA.py.MPI script implemented within AMBER11. The energy decomposition per amino acid residue was estimated using the MM-PBSA approach implemented in AMBER 11. Graphics were prepared using PyMOL Molecular Graphics System, Version 1.5.0.1 (Schrödinger, LLC) and Ligplot [45].

Acknowledgements

To perform this work, F. Arab received a doctoral fellowship from the French National Institute for Agricultural Research (INRA) and the French Environment and Energy Management Agency (ADEME). We thank Prof. C. Rémond (University of Reims Champagne Ardenne, UMR INRA 614, Fractionnement des AgroRessources et Environnement, F-51100 Reims, France) for supplying *Tx*Abf[†]. Equipment of microcalorimetry was financed with IBiSA grant. NMR equipment was financed by the French Research Ministry, CNRS, Université Paul Sabatier, the Région Midi-Pyrénées and European structural funds. NMR and ITC experiments were performed at the 'Integrated Screening Platform of Toulouse' (PICT, IPBS, CNRS - Université de Toulouse). This work was granted access to the HPC resources of CALMIP under the allocation P0836. The authors also thank the Center for Computing Resources (CRI) of INSA-Toulouse for providing computing resources and support.

References

[1] Dumon C, Song L, Bozonnet S, Fauré R & O'Donohue MJ (2012) Progress and future prospects for pentose- specific biocatalysts in biorefining. *Process Biochem* 47, 346–357.

[2] Taylor EJ, Smith NL, Turkenburg JP, D'Souza S, Gilbert HJ & Davies GJ (2006) Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from *Clostridium thermocellum*. *Biochem J* 395, 31–37.

[3] Hövel K, Shallom D, Niefind K, Belakhov V, Shoham G, Baasov T, Shoham Y & Schomburg D (2003) Crystal structure and snapshots along the reaction pathway of a family 51 α-L-arabinofuranosidase. *EMBO J* 22, 4922–4932.

[4] Im D-H, Kimura K, Hayasaka F, Tanaka T, Noguchi M, Kobayashi A, Shoda S, Miyazaki K, Wakagi T & Fushinobu S (2012) Crystal structures of glycoside hydrolase family 51 α-L-arabinofuranosidase from *Thermotoga maritima*. *Biosci Biotechnol Biochem* 76, 423–428.

[5] Souza TACB, Santos CR, Souza AR, Oldiges DP, Ruller R, Prade RA, Squina FM & Murakami MT (2011) Structure of a novel thermostable GH51 α-l- arabinofuranosidase from *Thermotoga petrophila* RKU- 1. *Protein Sci* 20, 1632–1637.

[6] Paës G, Skov LK, O'Donohue MJ, Rémond C, Kastrup JS, Gajhede M & Mirza O (2008) The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry* 47, 7441–7451.

[7] Jenkins J, Leggio LL, Harris G & Pickersgill R (1995) b-Glucosidase, b-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold b/a architecture and with two conserved glutamates near the carboxy- terminal ends of b-strands four and seven. *FEBS Lett* 362, 281–285.

[8] Henrissat B & Bairoch A (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* 316, 695–696.

[9] Davies G & Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859.

[10] Davies GJ, Wilson KS & Henrissat B (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* 321, 557–559.

[11] Fauré R, Courtin CM, Delcour JA, Dumon C, Faulds CB, Fincher GB, Fort S, Fry SC, Halila S, Kabel MA *et al.* (2009) A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls. *Aust J Chem* 62, 533–537.

[12] Ramasubbu N, Ragunath C & Mishra PJ (2003) Probing the role of a mobile loop in substrate binding and enzyme activity of human salivary amylase. *J Mol Biol* 325, 1061–1076.

[13] Ramasubbu N, Ragunath C, Mishra PJ, Thomas LM, Gyémánt G & Kandra L (2004) Human salivary a-amylase Trp58 situated at subsite-2 is critical for enzyme activity. *Eur J Biochem* 271, 2517–2529.

[14] Tranier S, Deville K, Robert X, Bozonnet S, Haser R, Svensson B & Aghajari N (2005) Insights into the 'pair of sugar tongs' surface binding site in barley a-amylase isozymes and crystallization of appropriate sugar tongs mutants. *Biologia* 60, 37–46.

[15] Jung J-H, Jung T-Y, Seo D-H, Yoon S-M, Choi H-C, Park BC, Park C-S & Woo E-J (2011) Structural and functional analysis of substrate recognition by the 250s loop in amylomaltase from *Thermus brockianus*. *Proteins: Struct, Funct, Bioinf* 79, 633–644.

[16] Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Oonanant W, Bevan DR, Esen A, Chen C-J, Opassiri R, Svasti J *et al.* (2008) Structural insights into rice BGlu1 b-glucosidase oligosaccharide hydrolysis and transglycosylation. *J Mol Biol* 377, 1200–1215.

[17] Pollet A, Vandermarliere E, Lammertyn J, Strelkov SV, Delcour JA & Courtin CM (2009) Crystallographic and activity-based evidence for thumb flexibility and its relevance in glycoside hydrolase family 11 xylanases. *Proteins: Struct, Funct, Bioinf* 77, 395–403.

[18] Paës G, Tran V, Takahashi M, Boukari I & O'Donohue MJ (2007) New insights into the role of the thumb-like loop in GH-11 xylanases. *Protein Eng Des Sel* 20, 15–23.

[19] Allison TM, Hutton RD, Jiao W, Gloyne BJ, Nimmo EB, Jameson GB & Parker EJ (2011) An extended β7α7 substrate-binding loop is essential for efficient catalysis by 3-deoxy-d-*manno*-octulosonate 8-phosphate synthase. *Biochemistry* 50, 9318–9327.

[20] Pang Y, Buck M & Zuiderweg ERP (2002) Backbone dynamics of the ribonuclease binase active site area using multinuclear (^{15}N and ^{13}CO) NMR relaxation and computational molecular dynamics. *Biochemistry* 41, 2655–2666.

[21] Debeche T, Cummings N, Connerton I, Debeire P & O'Donohue MJ (2000) Genetic and biochemical characterization of a highly thermostable α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *Appl Environ Microbiol* 66, 1734–1736.

[22] Rémond C, Plantier-Royon R, Aubry N & O'Donohue MJ (2005) An original chemoenzymatic route for the synthesis of β-D-galactofuranosides using an α-L-arabinofuranosidase. *Carbohydr Res* 340, 637–644.

[23] Mayer M & Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem Int Ed* 38, 1784–1788.

[24] Meyer B & Peters T (2003) NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed* 42, 864–890.

[25] Fielding L (2007) NMR methods for the determination of protein–ligand dissociation constants. *Prog Nucl Magn Reson Spectrosc* 51, 219–242.

[26] Angulo J, Rademacher C, Biet T, Benie AJ, Blume A, Peters H, Palcic M, Parra F & Peters T (2006) NMR analysis of carbohydrate–protein interactions. *Methods Enzymol* 416, 12–30.

[27] Chlubnová I, Filipp D, Spiwok V, Dvoř á ková H, Daniellou R, Nugier-Chauvin C, Krá lová B & Ferriè res V (2010) Enzymatic synthesis of oligo-d-galactofuranosides and l-arabinofuranosides: from molecular dynamics to immunological assays. *Org Biomol Chem* 8, 2092–2102.

[28] Euzen R, Lopez G, Nugier-Chauvin C, Ferriè res V, Plusquellec D, Ré mond C & O'Donohue M (2005) A chemoenzymatic approach for the synthesis of unnatural disaccharides containing d-galacto- or d-fucofuranosides. Eur J Org Chem 2005, 4860–4869.

[29] Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp Ser* 41, 95–98.

[30] Crooks GE, Hon G, Chandonia J-M & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14, 1188–1190.

[31] Rémond C, Ferchichi M, Aubry N, Plantier-Royon R, Portella C & O'Donohue MJ (2002) Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α-L-arabinofuranosidase. *Tetrahedron Lett* 43, 9653–9655.

[32] Mayer M & Meyer B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc* 123, 6108–6117.

[33] Piotto M, Saudek V & Sklená r V (1992) Gradient-tailored excitation for single-quantum NMR spectros- copy of aqueous solutions. *J Biomol NMR* 2, 661–665.

[34] Angulo J, Enrı´ quez-Navas PM & Nieto PM (2010) Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chem Eur J* 16, 7803–7812.

[35] Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM *et al.* (2010) *AMBER11*. University of California, San Francisco, CA.

[36] Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed- phase quantum mechanical calculations. *J Comput Chem* 24, 1999–2003.

[37] Lee MC & Duan Y (2004) Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins: Struct, Funct, Bioinf* 55, 620–634.

[38] Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL & Woods RJ (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrate. *J Comput Chem* 29, 622–655.

[39] Pastor RW, Brooks BR & Szabo A (1988) An analysis of the accuracy of Langevin and molecular dynamics algorithm. *Mol Phys* 65, 1409–1719.

[40] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A & Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81, 3684–3690.

[45] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H & Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103, 8577–8593.

[46] Ryckaert JP, Ciccotti G & Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J Comput Chem* 23, 327–341.

[47] Altona C & Sundaralingam M (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J Am Chem Soc* 94, 8205–8212.

[48] Weiser J, Shenkin PS & Still WC (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem* 20, 217–230.

[49] Wallace AC, Laskowski RA & Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng Des Sel* 8, 127–134.

Supporting information

The following supplementary material is available:

Fig. S1. pH-Dependence of *Tx*Abf and β2α2 loop mutants thereof;

Fig. S2. Binding interactions of inactivated *Tx*Abf† with *p*NP-α-L-Ara*f* by ITC;

Fig. S3 Root mean square deviation of backbone atoms as a function of time of wild-type *Tx*Abf (red colored) and W99A mutant (green colored) in *apo* form: **(A)** Overall enzymes. **(B)** β2α2 loop. **(C)** β7α7 loop;
Fig. S4. Snapshots taken along the 40 ns MD simulation of wild-type *Tx*Abf. From A to D, figures show the conformational rearrangements of the β2α2 and β7α7; **Fig. S5.** Snapshots taken along the 40 ns MD simulation of W99A mutant of *Tx*Abf. From A to D, figures show the conformational rearrangements of the β2α2 and β7α7;
Fig. S6. Time variation of the distance **(A)** O-1 (L-Ara*f*)…O(E176). **(B)** C-1 (L-Ara*f*)…O(E298). **(C)** Time variation of the furanose ring puckering along the first 2ns of MD simulations for *p*NP-α-L-Ara*f* in complex with wild-type *Tx*Abf (red) and W99A mutant (green);
Supplementary movie 1. Conformational rearrangements of the β2α2 and β7α7 loops of wild-type *Tx*Abf occurring along the 40 ns of MD simulation;

Supplementary movie 2. Conformational rearrangements of the β2α2 and β7α7 loops of the W99A mutant of *Tx*Abf occurring along the 40 ns of MD simulation.

This supplementary material can be found in the online version of this article:

http://onlinelibrary.wiley.com/doi/10.1111/j.1742-4658.2012.08720.x/suppinfo

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

RESEARCH PAPER

New Biotechnology Volume 30, Number 5 June 2013

Research Paper

Engineering transglycosidase activity into a GH51 α-L-arabinofuranosidase

Faten Arab-Jaziri[1,2,3,4,7], Bastien Bissaro[1,2,3,7], Michel Dion[5], Olivier Saurel[1,6], David Harrison[1,2,3], Fernando Ferreira[1,2,3], Alain Milon[1,6], Charles Tellier[5], Régis Fauré[1,2,3] and Michael J. O'Donohue[1,2,3]

[1] Université de Toulouse; INSA, UPS, INP; 135 Avenue de Rangueil, F-31077 Toulouse, France
[2] INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France
[3] CNRS, LISBP UMR5504, F-31400 Toulouse, France
[4] Agence de l'environnement et de la Maîtrise de l'Energie, 20 Avenue du Grésillé, BP 90406, 49004 Angers Cedex 01, France
[5] UFIP FRE CNRS 3478, Faculté des Sciences et Techniques, Université de Nantes, 2 rue de la Houssinière, 44322 Nantes cedex 03, France
[6] CNRS, IPBS UMR 5089, Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, BP 64182 Toulouse, France

Directed evolution was applied to the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus* to confer better transglycosylation ability, particularly for the synthesis of benzyl α-L-arabinofuranosyl- (1,2)-α-D-xylopyranoside, starting from *p*-nitrophenyl α-L-arabinofuranoside (donor) and benzyl α-D- xylopyranoside (acceptor). The aim was to obtain mutants displaying both lower hydrolytic and greater transglycosylation activities to favour the stable production of the target disaccharide. The implementation of a simple chromogenic screen ultimately provided three mutant enzymes whose properties correspond to those sought after. These all displayed lowered hydrolytic activity and conserved or slightly improved transfer activity, while one of them also displayed lowered secondary hydrolysis of the transglycosylation product. DNA sequence analysis of the mutants revealed between three and seven point mutations and biochemical analysis combined with STD-NMR experiments indicated that distinct molecular mechanisms were active among the three mutants.

Introduction

Despite considerable progresses, glycochemistry is an extremely challenging and labour-intensive process that usually leads to limited overall yields. Over the past decade, biocatalytic methods have been increasingly developed and adopted for the syntheses of sugar-based compounds. This is because enzymes present several advantages to form glycosidic linkage when compared to traditional catalysts: stereospecificity (i.e. anomeric specificity), regioselectivity, mild reaction conditions, and no protection/deprotection steps.

In nature, glycosynthesis is mainly performed by glycosyltransferases, but glycoside hydrolases (GHs), and glycosynthase mutants thereof, as well as 'glycan phosphorylases' are often used to synthesize tailor-made carbohydrates [1–4]. While the availability of glycosyltransferases and 'glycan phosphorylases' is still rather limited, GHs displaying a large spectrum of specificities are widespread and highly available. Moreover, even if most GHs preferentially catalyse hydrolysis, many GHs display good ability to run in transglycosylation mode and thus the potential to perform glycosynthesis. Chemo-enzymatic strategies, combining sequential enzymatic hydrolysis of plant polysaccharides, followed by the chemical modification of the products and enzyme-catalysed coupling, have already proved useful to synthe- size well-defined complex oligosaccharides, which can be used to provide carbohydrate-based compounds, or for studies aimed at acquiring new understanding of the biological roles of carbohy- drate structures [5–7].

Corresponding author: O'Donohue, M.J. (michael.odonohue@insa-toulouse.fr)

[7] **These authors contributed equally to this work.**

Arabinoxylans, polymers composed of a main chain containing β-(1,4)-linked D-xylopyranosyl (D-Xyl*p*) residues that can be substituted at *C*-2 and/or *C*-3 with *α*-L-arabinofuranosyl (L-Ara*f*) units, are the most abundant pentose sugar-based polymers, being widespread in the plant kingdom. These polysaccharides, and oligosaccharide derivatives thereof, are known to display a wide range of physicochemical or physiological properties [8–11] and so their easy preparation, using *in vitro* synthetic methods, would be useful, not only for fundamental studies of the *in planta* roles of arabinoxylans, but also for the production of a variety of value-added commodities, including biobased films, prebiotics and health-related derivatives and biosurfactants [12–16]. However, as mentioned previously, chemical methods are inappropriate to produce such molecules in large quantities, because even the preparation of arabinoxylan-oligosaccharides involves multiple step synthesis pathways [17–20].

Most developments in the area of enzyme-mediated glycosynthesis applied to pentose sugars have focused on the use of D-xylopyranose, while sugars in the furanose form, in particular L-Ara*f* have been rather neglected, even if these are key elements of arabinoxylans. Nevertheless, new pentose sugar-specific chemoenzymatic strategies are emerging [21,22] and the glycosynthetic abilities of L-Ara*f*-acting GHs have been recently reported [23–28]. However, despite the fact that the enzymes used in these studies display interesting innate transglycosylation properties, the synthesis yields are generally modest.

α-L-Arabinofuranosidases (Abfs, EC 3.2.1.55) are mainly *exo*-acting GHs that release L-Ara*f* side-chain decorations from polysaccharides such as arabinoxylans. Regarding glycosynthesis, it is noteworthy that the Abf from *Thermobacillus xylanilyticus* (*Tx*Abf) is able to catalyse the synthesis of benzyl *α*-L-arabinofuranosyl-(1,2)-*α*-D-xylopyranoside (Bn α-L-Ara*f*-(1,2)-α-D-Xyl*p*), a benzyl derivative of a key arabinoxylan motif [24]. This particular reaction, and more generally transglycosylation, is possible because *Tx*Abf operates via a retaining mechanism, which means that the glycosyl-enzyme intermediate can be deglycosylated using appropriate sugar moieties, thus creating an alternative to hydrolysis (Fig. 1). Nevertheless, although *Tx*Abf displays strong global activity on *p*-nitrophenyl *α*-L-arabinofuranoside, *p*NP-*α*-L-Ara*f* (k_{cat} = 575 48 s^{-1}, K_M = 0.72 0.04 mM, and k_{cat}/K_M = 795 67 $s^{-1}.mM^{-1}$), the outcome is thermodynamically largely in favour of hydrolysis, with Bn *α*-L-Ara*f*-(1,2)-*α*-D-Xyl*p* being obtained in low yield (7%) and being subject to secondary hydrolysis. Therefore, despite being a potentially useful tool for glycosynthesis, *Tx*Abf only provides low yields (<10%) of L-Ara*f*-based products and thus in its present state is not useful for the glycochemist. From a mechanistic point of view, the low yield of L-Ara*f*-based products displayed by *Tx*Abf is almost certainly due to both an unfavourable transglycosylation/hydrolysis (T/H) ratio and *Tx*Abf-mediated secondary hydrolysis of nascent oligosaccharide products. Therefore, protein engineering strategies aimed at creating better glycosynthetic ability must address both the alteration of the partition between transglycosylation and primary hydrolysis, and the reduction in the level of secondary hydrolysis. In this work, we have deployed a random engineering strategy, coupled to screening, which has succeeded in addressing both of these aspects. Three mutants displaying a lower propensity for hydrolysis, but an intact ability to perform transarabinofuranosylation reactions have been isolated and characterized, notably to evaluate the extent of modification involved into the partition between hydrolysis and transglycosylation.

Materials and methods

General and specific reagents

*p*NP-α-L-Ara*f*, benzyl α-D-xylopyranoside (Bn-α-D-Xyl*p*), and 5-bromo-4-chloro-3-indolyl α-L-arabinofuranoside (X-α-L-Ara*f*) are synthetic substrates, prepared in-house using published protocols [24,29]. NMR spectra of 5,5′-dibromo-4,4′-dichloroindigo **1** were recorded on a Bruker Ascend 800 spectrometer at 313 K. Chemical shifts are herein expressed in ppm, coupling constants are reported in Hz, and multiplicity is denoted by 'd' (i.e. meaning doublet). High-resolution mass spectra (HRMS) analysis was performed at the CRMPO (Rennes University, France) in negative ionisation mode (ES) on a Bruker MicrOTOF-Q II spectrometer.

Preparation of 5,5′-dibromo-4,4′-dichloroindigo (1)

5-Bromo-4-chloro-3-indolyl 1,3-diacetate (100 mg, 0.30 mmol, 1 equiv.) in dry methanol (3 mL) was treated with sodium methoxide (1 M in methanol, 300 μL, 0.30 mmol, 1 equiv.) for 2 h at room temperature. The mixture was neutralized with acetic acid, concentrated under reduced pressure and co-evaporated with toluene. The crude product was washed with methanol and the expected indigo dimer **1** was quantitatively isolated by filtration (73 mg, 0.15 mmol) as a blue colour powder. ^{1}H NMR (800 MHz,

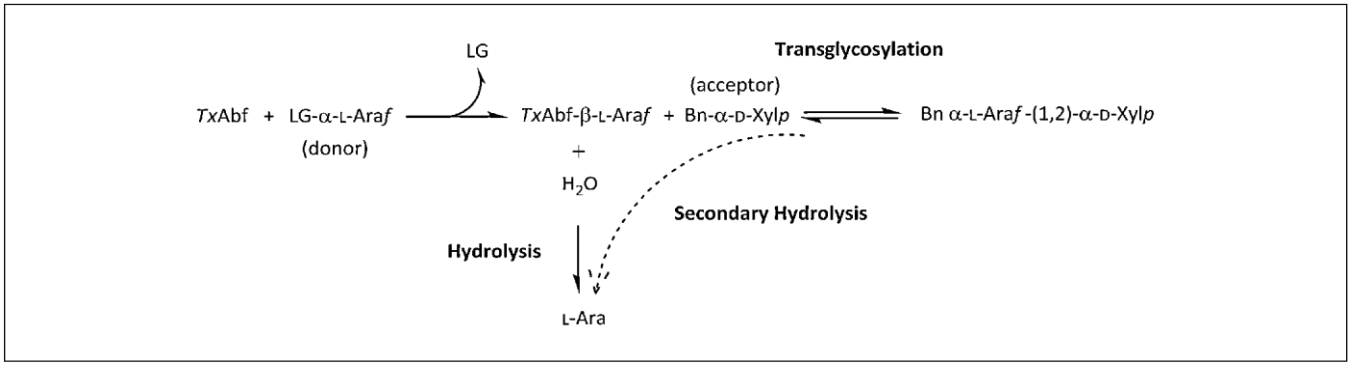


FIGURE 1. Reactions catalysed by *Tx*Abf (leaving group: LG = *p*NP or X).

$(CD_3)_2SO$): d 7.81 (2H, d, J = 8.6), 7.26 (2H, d, J = 8.6); ^{13}C NMR (200 MHz, $(CD_3)_2SO$): d 183.9, 152.9, 139.3, 129.9, 120.9, 116.9, 113.9, 112.9; ESI-HRMS: m/z calcd for $C_{16}H_5Br_2Cl_2N_2O_2Na$: [M H] 484.81003; found: 484.8101 (0 ppm).

Random mutagenesis

The plasmid containing the complete coding sequence of *Tx*Abf (GenBank accession no. CAA76421.2), pTXABF, was used as template DNA for *in vitro* random mutagenesis, performed using epPCR with forward (5′-TAATACGACTCACTATAGGG) and reverse (5′-GCTAGTTATTGCTCAGCGG) oligonucleotide primers for PCR-mediated DNA amplification. Specifically, 0.2 μM of each primer was mixed with 5 ng of plasmid pTXABF in a 25 μL reaction and epPCR was performed using the reaction conditions described by Cadwell and Joyce [30]: 1 *Taq* polymerase buffer, 0.2 mM dATP and dGTP, 1.2 mM dTTP and dCTP, 7 mM $MgCl_2$, 1.25 mM $MnCl_2$, and 1.25 unit of *Taq* polymerase. DNA amplification was then performed according to the following sequence: 1 cycle at 95 °C for 2 min; 30 cycles of 95 °C for 45 s, 55 °C for 1 min, and 72 °C for 2 min; then 1 cycle at 72 °C for 4 min. After amplification, the amplicon DNA was digested using *Nde* I and *Hind* III, purified using a QIAquick® gel extraction kit (Qiagen) and then ligated into linearized pET21 (original pET21 vectors from Novagen also digested by the same restriction enzymes), before being used to transform chemically competent *E. coli* TOP 10 cells (Novagen). The resultant mutant plasmid library was recovered by scraping all the transformed cells followed thereafter by extraction using QIAprep® spin Miniprep Kit (Qiagen) and stored at 20 °C.

Screening of randomly generated mutants

The epPCR library was used to transform chemically competent *E. coli* BL21 DE3 star (Invitrogen) and then submitted to a previously described screening protocol [31–33], with modifications accounting mainly for the specific characteristics of *Tx*Abf (i.e. choice of substrate and its concentration, reaction time, etc.). Briefly, transformants were grown for 12 h at 37 °C on a nitrocellulose membrane (0.45 *m*m, 20 cm 20 cm), placed on the surface of LB agar medium containing ampicillin (100 μg mL ^{1}), thus providing small bacterial colonies (Ø <0.5 mm). Colonies were transferred onto minimal medium (sodium phosphate buffer 100 mM, pH 7, agar 15 g L ^{1}) containing IPTG (0.4 mM), X-α-L-Ara*f* solubilized in DMSO (0.3 mM) and ampicillin (100 μg mL^{-1}) and incubated at 378C. After the development of colour within colonies (1 h), an image of the membrane was captured using a scanner (Epson 4490, 16 bits grey levels, 600 dpi) and digital image analysis was performed using Visilog 6.3 image analysis software (Noesis, France). Colour intensity was calculated using 65,000 grey scales and individual colony and mean intensity areas were determined after applying a threshold value to eliminate white colonies. The latter values were used to classify colonies in an ascending order of coloration, and thus to select paler colonies for further analysis. Crude extracts of selected mutants were prepared as follows: 5 mL of an overnight LB medium culture incubated at 37 °C of each clone was induced with IPTG (0.4 mM) and growth was pursued for a further 1 h 30 at 37 °C, before recovering bacterial cells using centrifugation (2057 *g*, 20 min, 4 °C). Cell pellets were suspended in Tris–HCl buffer (20 mM, pH 8, 500 μL) and subjected to brief sonication (4 cycles of 30 s, interspersed by 30 s pauses), with Bioblock Scientific Vibracell 72434 using a small probe operating at 30% of maximum power. Then, the lysates were recovered by centrifugation (15,871 *g*, 10 min), purified by a heat treatment step (75 °C, 30 min) and then submitted to a second identical centrifugation step. Hydrolytic activities were determined in the supernatant and the ability of the different enzymes was investigated to perform transglycosylation. Enzyme assays were performed introducing each normalized crude extract (0.055 IU) in a solution of sodium acetate buffer (50 mM, pH 5.8) containing *p*NP-α-L-Ara*f* (5 mM) and Bn-α-D-Xyl*p* (5 mM). The mixture was allowed to react at 60 °C (synthesis operating temperature for the wild-type *Tx*Abf) and transglycosylation was followed by TLC (precoated Silica Gel 60 sheets; Merck F254), using the eluent (ethyl acetate/acetic acid/water, 7:2:2, v/v/v), until the donor was fully consumed. After, spots on the TLC plates were visualized using ultraviolet light (254 nm) and then by soaking the plates in a 10% w/v orcinol solution containing sulphuric acid/ethanol/water solution (3:72.5:22.5, v/v/v) followed by charring.

Mutagenesis, protein expression and purification

The plasmids containing the complete coding sequence of mutants 6, 1364 and 1753 (original pET21 vectors from Novagen) were used as a template for *in vitro* mutagenesis using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Mutants for STD-NMR studies were inactivated by substitution of the acid/base catalytic residue into alanine (E176A) using the following primer (bold lettering indicates mismatch positions): 5′-GGCGTCGGCAAC**GCC** AACTGGGGCTGC-3′ as previously described [34]. The resultant inactive derivatives were denoted *Tx*Abf†, mutant 6†, 1364† and 1753†. The successful introduction of mutations was confirmed by DNA sequencing (GATC Biotech). Expression and subsequent purification of wild-type and all mutated recombinant *Tx*Abf, produced in *E. coli* BL21(DE3), were performed as previously described [35].

Hydrolysis kinetic parameters

All protein concentrations were determined spectrophotometrically (at 280 nm) with NanoDrop ND-1000 spectrophotometer and using theoretical absorption coefficients and molar masses provided by the ProtParam software (http://www.expasy.org/). Routinely, the hydrolytic activity was determined using a discontinuous enzyme assay as previously described [34]. The activity of wild-type *Tx*Abf was assayed at 37 °C toward X-α-L-Ara*f* using a substrate solution prepared in DMSO (the final DMSO concentration in reactions did not exceed 20% without damage for the *Tx*Abf hydrolytic activity). Optimal temperatures for mutants were determined by measuring hydrolytic activity at least at two temperature values lower and higher than the optimal one. Thereby, kinetic parameters for *p*NP-α-L-Ara*f* hydrolysis with mutants 6, 1364 and 1753 were performed at 60, 45, and 30 °C, respectively. Enzymes were assayed using X-α-L-Ara*f* (only for the wild-type enzyme) or *p*NP-α-L-Ara*f* as the substrate in the concentration range 0.5–15 mM and 0.25–18 mM, respectively. The release of indigoid derivative 1 or *p*NP was monitored using a VersaMax Microplate Reader spectrophotometer (at 660 and 401 nm, respectively) and quantified by comparison with a standard curve. The Michaelis–Menten kinetic parameters, k_{cat} and K_M, were obtained from plots of v (initial rate) versus [S] (substrate concentration)

TABLE 1

Kinetic parameters for hydrolysis of *p*NP-α-L-Ara*f* and transarabinofuranosylation yields onto Bn-α-D-Xyl*p* by wild-type *Tx*Abf and mutants thereof

Enzyme	*Tx*Abf	6	1364	1753
Operating temperature (°C)	60	60	45	30
k_{cat} (s^{-1})	575 ± *48*	0.065 ± *0.001*	0.158 ± *0.003*	22.4 ± *1.5*
K_M (mM)	0.72 ± *0.04*	1.33 ± *0.10*	1.25 ± *0.10*	7.49 ± *1.14*
k_{cat}/K_M ($s^{-1}.mM^{-1}$)	795 ± *67*	0.049 ± *0.003*	0.13 ± *0.02*	3.0 ± *0.5*
Maximal yield (%)	16	17	11	21
Time (min)	70	180	40	45

using SigmaPlot 10.0 software. One unit (IU) of enzyme specific activity represents the quantity of enzyme (expressed in mg) converting 1 μmol of *p*NP per minute.

Monitoring transglycosylation by NMR spectroscopy

Reactions were monitored by collecting ^{1}H NMR spectra using a Bruker Avance 700 spectrometer. The enzymatic solutions were transferred in D_2O buffer by dilution in D_2O (100 x) followed by concentration using an Amicon[1] Ultra 10 kDa device (Millipore). The chemical shift referencing was based on the residual water peak, calibrated at d = 4.72, 4.55 or 4.40 ppm for 30, 45 and 608C temperatures, respectively [36]. Transglycosylation reactions were performed at operating temperatures that were arbitrarily chosen to reflect a compromise between activity and stability (Table 1). Reactions were prepared in 600 μL of D_2O (final volume) with donor/acceptor ratio of 1:1 (5 mM each) and an amount of enzymes corresponding to 15, 15 or 0.25 μM for mutants 6, 1364 and 1753, respectively or 1 nM for *Tx*Abf wild-type (0.028, 0.068, 0.081 and 0.011 IU, respectively). ^{1}H NMR spectra were accumulated continuously in 4.75 min (32 scans with a repetition delay of 6 s) during 4–15 hours depending on enzymes. Each NMR spectrum was acquired using an excitation flip angle of 308 at a radiofrequency field of 29.7 kHz, and the residual water signal was pre-saturated during the repetition delay (with a radiofrequency field of 21 Hz). To quantify the time-dependent evolution of the different reactions species, the anomeric proton signals were integrated. The (temperature-dependent) chemicals shifts of these species were in the range: donor, 5.87–5.81 ppm; acceptor, 4.99–4.94 ppm; transglycosylation products, 5.08–5.05 and 5.09–5.04 ppm for H-1 of α-L-Ara*f* and α-D-Xyl*p*, respectively. Moreover, the signals associated with *p*NP were also integrated for quantification, and the correlation between acceptor consumption and transglycosylation product formation was verified. To compare the wild-type and mutant enzymes, NMR kinetic data were normalized by plotting transglycosylation yield versus donor consumption.

STD-NMR experiments

Samples were prepared in 600 μL of D_2O containing 100 μM of *p*NP-α-L-Ara*f* with a molar ratio (protein:ligand) of (1:100). STD-NMR experiments were performed at 10 °C with a Bruker Avance 600 spectrometer equipped with a cryoprobe TCi as follows [37]: the proteins were saturated on resonance at 0.4 ppm on methyl signals, and off resonance at 30 ppm, using selective 50 ms Gaussian-shaped pulses, at a radiofrequency field amplitude of 86 Hz, with a 100 ms delay between each pulse. The total duration of the saturation time was 2 s. A WATERGATE sequence was used to suppress the residual water signal [38]. A similar experiment was performed without enzyme to verify the selectivity of saturation and the efficiency of the signal subtraction that was used to obtain the different spectrum conditions. Intensities of all STD effects (I_{STD}) were calculated by integrating the respective ^{1}H NMR signals, followed by comparison with the reference signal I_0. To compare the STD-NMR bound-ligand fingerprinting, the ratio I_{STD}/I_0 was normalized using the largest STD effect as the reference (i.e. all STD effects were reported as a percentage of the STD effect of the H-2 proton of the L-Ara*f* unit for *Tx*Abf[†] and mutants thereof, which was arbitrarily set to 100%). To compare interaction levels for the different mutants, STD effects of protons exhibiting identical chemical shifts were standardized for each mutant, using the STD data obtained with *Tx*Abf[†] [34]. All data were acquired and processed using Topspin v2.1 software (Bruker).

Results

Library characteristics and screening

To improve the transglycosylation/hydrolysis ratio of the *Tx*Abf, an error-prone PCR (epPCR) library was created and screened, using a two-step procedure consisting of colony selection via screening on solid medium and then TLC analysis of transarabinofuranosylation reactions catalysed by selected mutants. Regarding the epPCR library, this contained approximately 10,000 clones and displayed an average mutation rate of 2.9 point mutations/kb, which provided a mean average of 3.8 amino acid substitutions per *Tx*Abf-encoding sequence.

Briefly, the logic of the previously established screening procedure [31–33] is to detect mutant enzymes that are able to perform the first glycosylation step of the double-displacement mechanism (Fig. 1) [39,40], but are deficient with respect to the water-mediated deglycosylation step, and thus are characterized by a prolonged lifetime for the glycosyl-enzyme intermediate. Such mutants can release the aglycon moiety of X-α-L-Ara*f*, allowing the formation of blue colour, but are then slow to be recycled for a second round of catalysis. In the absence of a suitable acceptor (other than water), the overall result is a much lower level of aglycon release and thus the formation of a light blue colour.

Library screening was performed using X-α-L-Ara*f* at a concentra- tion 10-fold below the K_M for this substrate, which was determined (from data collected from reactions operated at 37 °C) to be 2.8 mM. Likewise, 57% of all colonies displayed a blue coloration after 1 h, although differences could be distinguished (Fig. 2). Dark blue colonies, which are characteristic of 'normal' or unmodified hydrolysis of X-α-L-Ara*f*, represented 47% of all colonies, while light blue

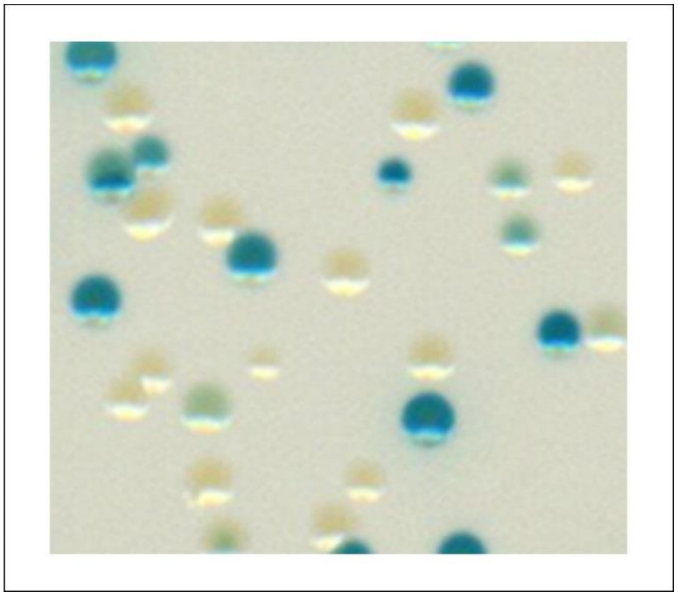


FIGURE 2

Digital image recognition of hydrolytically active (blue and pale blue) and inactive library mutants (white), using X-α-L-Ara*f* as the substrate.

colonies, which apparently expressed mutant enzymes that have reduced hydrolytic ability, represented 10% of the colonies. White colonies, apparently producing inactive mutants, accounted for the remaining 43%. DNA sequence analysis of five to fifteen clones from each of these phenotypic groups revealed that the mutational load was highest in plasmids extracted from white colonies (7.0 amino acid substitutions per coding sequence), whereas plasmids from dark and light blue colonies exhibited mean average amino acid substitutions per coding sequence of 0.8 and 3.8 respectively. Similarly, the localization of the point mutations within the protein sequence of *Tx*Abf revealed a differential distribution of point mutations, because an average of only 0.6 amino acid substitutions were located in the $(\beta/\alpha)_8$ catalytic domain of sequences extracted from dark blue colonies, whereas inactive mutants (white colonies) contained an average of 4.0 amino acid substitutions located within this domain. The light blue phenotype correlated with 2.8 amino acid substitutions located within the $(\beta/\alpha)_8$ catalytic domain. After digital imaging monitoring, 50 colonies displaying the palest blue colour were selected for further analysis.

As pale blue colonies could also come from a lower mutant expression level or a decrease of the overall (hydrolytic and transglycosylation) enzyme activity upon mutation, a second screening step was applied to the selected mutants. Crude cellular extracts of individual mutants were prepared and used as the enzyme component in transglycosylation reactions using *p*NP-*α*-L-Ara*f* (donor) and Bn-*α*-D-Xyl*p* (acceptor) as substrates. Analysis of the products of transglycosylation using TLC revealed that eight mutants were able to synthesize a product that remained relatively stable, compared to the product of transglycosylation synthesized by wild-type *Tx*Abf, which was almost immediately hydrolysed (data not shown). Moreover, among the eight mutants, three (from clones 6, 1364, and 1753) produced a product that persisted even after consumption of the donor, although reaction temperature might not be equally optimal for each mutant. Nevertheless, these three mutants were selected for further biochemical analysis.

Kinetic analysis of mutant hydrolytic activity

Mutants 6, 1364 and 1753 were successfully expressed in *E. coli* and obtained in pure soluble form using previously established methods [35]. Nevertheless, the analysis of the temperature dependency of the activity of the different mutant enzymes revealed that the optimal temperature for activity was generally lower (60, 45 and 308C for the mutants 6, 1364 and 1753, respectively) compared to wild-type *Tx*Abf (Table 1). Concerning the kinetic parameters, determined for the hydrolysis of *p*NP-*α*-L-Ara*f*, values of k_{cat} and k_{cat}/K_M were severely reduced for mutants by factors of up to 8800-fold for k_{cat} and over 16,000-fold for k_{cat}/K_M (in the case of mutant 6) and specific activity was <5% of that of wild-type *Tx*Abf efficiency, which is coherent with the pale blue phenotype. Concerning the K_M values, the impact of mutations was variable. Regarding mutants 6 and 1364, the value of K_M was approximately doubled, whereas that of mutant 1753 was increased nearly 10-fold compared to wild-type *Tx*Abf.

Glycosynthesis and secondary hydrolysis performed by mutants

In a previous study, it has been reported that *Tx*Abf can catalyse the synthesis of Bn *α*-L-Ara*f*-(1,2)-*α*-D-Xyl*p*. In the study conditions (i.e. using 10 nM of *Tx*Abf), a maximal product yield of 7% is reached after only 5 min and over longer incubation periods the product is consumed [24]. Although *p*NP-*α*-L-Ara*f* competes with Bn-*α*-D-Xyl*p* for the acceptor subsite (i.e. self-condensation versus condensation), the self-condensation reaction appears to be negligible, with only trace amounts of the arabinobioside product being detectable. Moreover, in previous work the *Tx*Abf-catalysed transarabinofuranosylation appeared to be highly regioselective toward the *O*-2 position of the acceptor substrate [24,25]. To investigate the transglycosylation ability of the mutants prepared in this work, reactions were quantitatively analysed using ^{1}H NMR spectroscopy, until quasi-complete consumption of donor substrate had occurred (Fig. 3). This quick and convenient analytic method allows the detection, in a single tube, of all the species

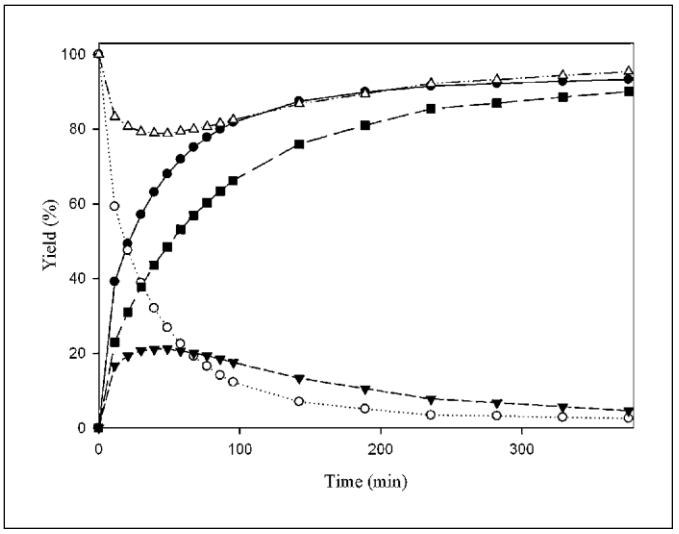


FIGURE 3

Monitoring the reaction kinetics catalysed by the mutant 1753 using ^{1}H NMR (500 MHz, D_2O, 30°C). Substrates were *p*NP-α-L-Ara*f* (5 mM, ○) and Bn-α-D-Xyl*p* (5 mM, △), while the products are L-Ara (■), *p*NP (●) and Bn-α-L-Ara*f*-(1,2)-α-D-Xyl*p* (▼).

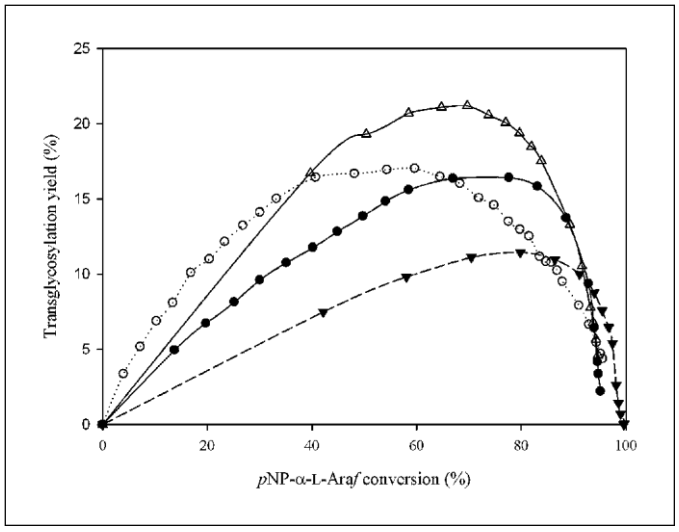


FIGURE 4

Transglycosylation product yield (%) as a function of donor substrate consumption (%). ^{1}H NMR provided the data describing the transarabinofuranosylation mediated by wild-type *Tx*Abf (●) and mutants 6 (○), 1364 (▼) and 1753 (△).

encountered in the reaction, except the covalent glycosyl-enzyme intermediate. Indeed, the initial donor and acceptor substrates as well as the thermodynamic and kinetic products evolutions (i.e. L-arabinose and transglycosylation product, respectively) can be monitored, providing an accurate quantification of the T/H partition. Consistent with previous data, the benzyl arabinoxyloside product formed in the reaction catalysed by wild-type *Tx*Abf was consumed in less than 20 min when using 10 nM *Tx*Abf (data not shown). However, when using 10-fold less enzyme (1 nM or 0.011 IU) and equimolar concentrations of substrates (5:5, mM/mM), the yield of benzyl arabinoxyloside became maximal (16%) after 70 min (Fig. 4 and Table 1). Similarly, the use of mutants 6, 1364 (both 15 μM) and 1753 (0.25 μM), that is, 0.028, 0.068 and 0.081 IU respectively, afforded yields from 11 to 21% over time periods in the range 40 min to 3 h. However, unlike the reaction catalysed by wild-type *Tx*Abf, the product synthesized by each of the mutants displayed extended lifetime (persistence even after almost complete consumption of the donor). Concerning the nature of the product, analysis of the ^{1}H NMR data, notably by comparing the chemical shifts of anomeric protons, 5.05 ppm (d, 1.7 Hz, H-1 of *a*-L-Ara*f*) and 5.04 ppm (d, 3.7 Hz, H-1 of α-D-Xyl*p*), at 60 °C of the products produced by wild-type *Tx*Abf and mutant 6 clearly indicated that this latter synthesized an α-L-(1,2)-linked benzyl arabinoxyloside [24]. In the case of mutants 1364 and 1753, a temperature-dependent coalescence of the anomeric proton signals was observed: 5.06 (H-1 of α-L-Ara*f*) and 5.06 ppm (H-1 of α-D-Xyl*p*) at 45 °C, and 5.09 (H-1 of α-D-Xyl*p*) and 5.08 ppm (H-1 of *a*-L-Ara*f*) at 308C. The ^{1}H NMR analyses of the synthesized Bn *a*-L-Ara*f*-(1,2)-α-D-Xyl*p* at different temperatures reinforced the highly regioselectivity of wild-type *Tx*Abf and mutants 6, 1364 and 1753 thereof toward the *a*-L-(1,2) linkage formation when using synthetic acceptor [25]. In addition to the major product, minor traces (<1%) of self-condensation product were observed. Analysis of the kinetics of the reactions monitored by ^{1}H NMR spectroscopy revealed that compared to wild-type *Tx*Abf the reaction catalysed by mutant 6 was particularly marked by slower secondary hydrolysis, despite the presence of 2.5-fold more enzyme (IU). To a lesser degree, this was also the case for mutant 1753, a reaction in which enzyme activity was 7-fold higher. However, compared to the reaction catalysed by wild-type *Tx*Abf, the reaction catalysed by mutant 1364 (containing 6-fold more activity) displayed lower product yield and almost unchanged secondary hydrolysis (Fig. 4).

Enzyme–substrate interactions by STD-NMR

To further characterize the different mutants, especially with respect to their interactions with the donor compound, *p*NP-*a*-L-Ara*f*, saturation transfer difference (STD)-NMR spectroscopy was

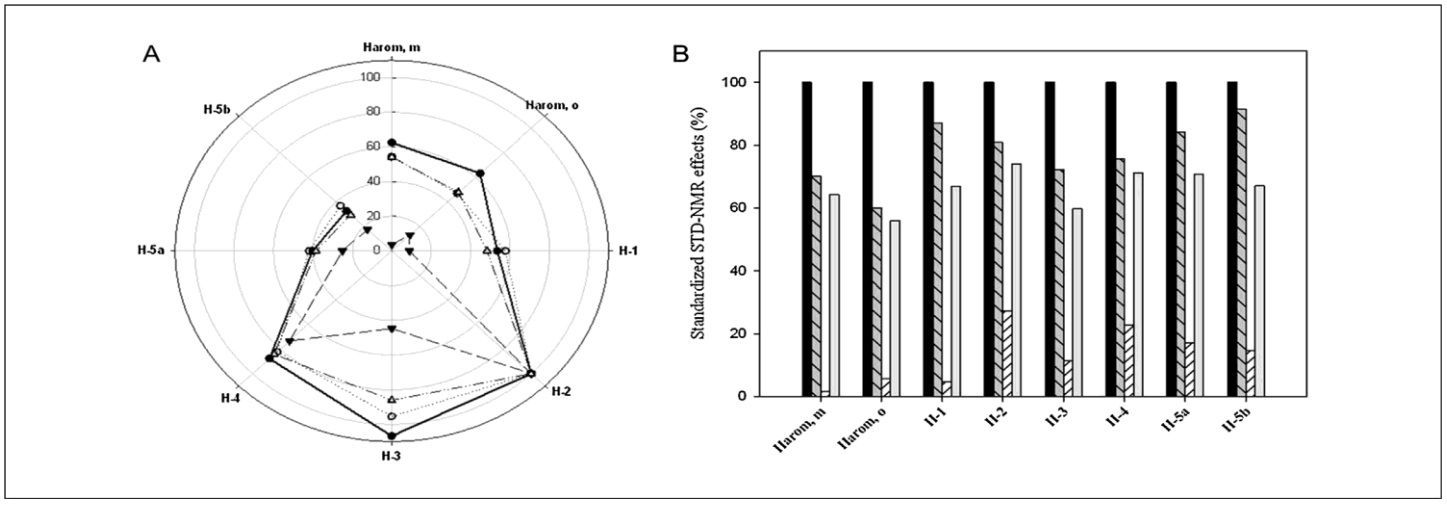


FIGURE 5

STD-NMR analysis of enzyme-ligand interactions. (a) STD-NMR ligand fingerprinting of *p*NP-α-L-Ara*f* bound to *Tx*Abf† (●, solid) and mutants 6† (○, dotted), 1364† (▼, dashed), and 1753† (△, dash-dot). The STD effects, expressed as percentages, are relative to the highest proton signal of the bound substrate for each enzyme, that is, H-2 for all enzymes. (b) Standardized STD-NMR effects expressed for each proton as relative percentages of the effects measured for *Tx*Abf† (black bars) and *p*NP-α-L-Ara*f* with mutants 6† (grey half-hatched bars), 1364† (white half-hatched bars) and 1753† (grey).

TABLE 2. Amino acid (AA) substitutions associated with each of the mutants

Mutant identifier	Total number of AA substitutions	AA substitutions in catalytic domain	AA substitutions in C-terminal domain
6	4	N106S, W178R, T329S	D453N
1364	7	I34V, E89G, Y220C, F257L, K290E, N344S, E359K	None
1753	3	F26L, S319P	D471N

employed to map those substrate protons that are in close contact with the protein [34,41,42]. This revealed that the interactions between the protons of the donor with *Tx*Abf† and mutants 6† and 1753† were globally conserved in comparison to mutant 1364† (Fig. 5a). Additionally, the analysis of the standardized STD inten- sities, relative to *Tx*Abf†, revealed that the values measured for mutants 6^{y} and 1753† were moderately modified when compared to the mutant 1364† (Fig. 5b). Indeed, the latter displayed decreased STD intensities for all of the sugar-associated protons with values varying between 5 and 27% of those obtained with *Tx*Abf†. This effect was particularly strong for H1 (5%), H3 (11%), and aromatic protons (2 and 6%). Contrarily, for mutants 6† and 1753†, *p*NP-*α*-L-Ara*f* protons are globally slightly affected (60–92% and 56–74% of the reference enzyme, respectively), without major differences among sugar protons and aromatics moieties.

Identification and analysis of mutants

DNA sequence analysis of mutants 6, 1364, and 1753 revealed the presence of 4, 7, and 3 amino acid substitutions respectively (Table 2). It is noteworthy that all of the mutations in mutant 1364 were localized within the $(\beta/\alpha)_8$ catalytic domain and only two out of a total of 14 amino acid substitutions mutations, were actually localized in the C-terminal jelly-roll domain. Nevertheless, apart from this observation, no actual structural hotspots could be identified, because 12 amino substitutions were quite widely spread over the catalytic domain, with some being located within approximately 5 Å of C*α* of either the catalytic acid/base, E176, or the nucleophile, E298 (PDB: 2VRK, chain C) [43]. These include N344 (C*α*-C*α* E298, 5.5 Å), F26 (CE2-C*α* E298, 5.3 Å) and W178 (C*α*-C*α* E176, 5.4 Å), with the latter being the only residue that appears to be capable of being in direct contact with a D-Xyl*p* unit bound in the acceptor subsite +1 (Fig. 6). Interestingly, residue N344, which lies close to the catalytic nucleophile, was also altered in mutant 3 (N344K), which was among the eight mutants initially selected by TLC. However, mutant 3 was not retained for further study, because of its low expression. Finally, it is noteworthy that although, in wild-type *Tx*Abf, F26 has not been described as interacting with the L-Ara*f* moiety [43], it does contribute to the topology of the subsite 1 and is close to E28, previously described as a catalytic helper residue [44].

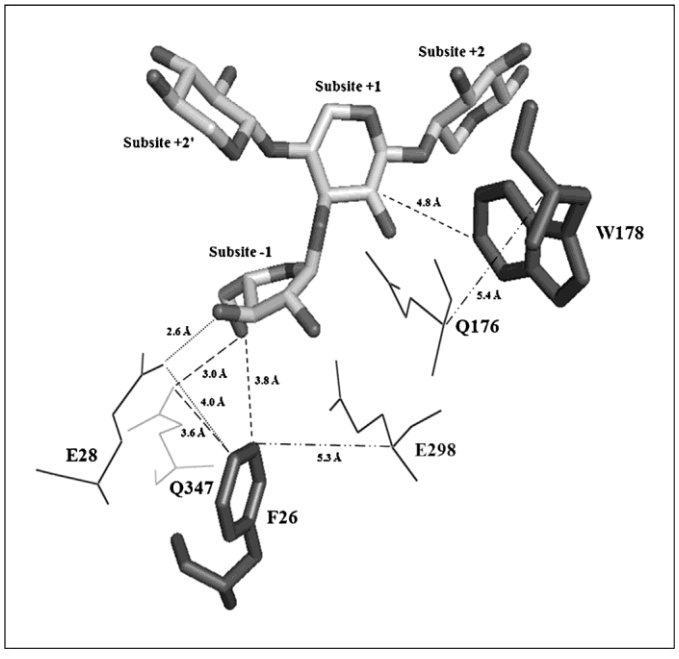


FIGURE 6

View of the *Tx*Abf-E176Q:XA^{3}XX complex active site showing the positioning of F26 and W178 residues, close to the catalytic helper E28 and Q347, and to the D-Xyl*p* moiety in subsite +1, respectively (PDB code 2VRQ, chain A) [43].

Discussion

GHs that perform catalysis using a two-step retaining mechanism form a covalent glycosyl-enzyme intermediate, or equivalent oxazoline, that in certain circumstances can be deglycosylated by nucleophiles other than activated water molecules. In reactions performed using substrates that possess good leaving groups, it is has been demonstrated that the formation of the covalent glycosyl-enzyme intermediate is a fast process and that consequently enzyme deglycosylation is the rate-limiting step in the reaction [45]. In this study, we believe that this assumption can be applied to *Tx*Abf and mutants thereof, because kinetic studies were performed using *p*NP-*α*-L-Ara*f*, which possesses a rather good leaving group (pK_a < 8), with k_{cat} reflecting the efficiency of the deglycosylation step and k_{cat}/K_M the initial enzyme glycosylation step [28,46].

The kinetic analysis of the three mutants selected for further study in this work clearly revealed that the catalytic properties of all of the mutants were severely affected compared to the parental enzyme. Nevertheless, 1753 stands out because of its drastically altered K_M value (7.49 1.14 mM). Taking into account the Arrhenius relationship, one can assume that the theoretical k_{cat} value of 1753, operating at 60 °C, would be less than one order of magnitude lower than that of *Tx*Abf, leading to the tentative conclusion that 1753 deglycosylates the covalent glycosyl-enzyme intermediate almost as efficiently as wild-type *Tx*Abf. In this case, the altered K_M value might be more attributable to changes in substrate binding and/or enzyme glycosylation, although this is not totally reflected in the STD-NMR measurements, which only revealed modest perturbations with regard to the binding of *p*NP-*α*-L-Ara*f*. Nevertheless, STD-NMR effects only reflect the proximity of substrate-associated protons to protein-associated ones, thus subtle changes at this level could still have serious consequences for catalysis. Moreover, the notion that the catalytic properties of 1753 could be linked to altered binding in subsite -1 can be drawn

from a tentative analysis of the mutations that are likely to actually affect catalysis. The mutation F26L stands out as the probable culprit, because long range effects from S319P and D417N appear highly improbable. Significantly F26 is located in subsite 1 and, according to crystallographic data acquired for *Tx*Abf bound to the oligosaccharide XA^{3}XX, lies close (3.8 Å) to the *O*-5 position of the L-Ara*f* moiety, and to the catalytic helper residues E28 and Q347. Applying a similar logic to mutants 6 and 1364 indicates that both glycosylation and deglycosylation steps are severely affected in these mutants, which is coherent with the fact that the K_M values are less affected. Nevertheless, STD-NMR analysis has revealed that the binding of *p*NP-α-L-Ara*f* to mutant 1364 is also quite perturbed, suggesting that an alteration in the ratio k_{-1}/k_1 is extremely likely. In this respect, it is noteworthy that the sugar protons whose STD signals are most affected are H-1 and H-3, these being located on the same face of the substrate and orientated toward the active site of the enzyme. Consequently, this result might be indicative of a slight relocation of the sugar moiety. Regarding mutant 6, it is highly tempting to assume that its properties are determined by the mutation W178R, because W178 probably makes a direct contact with the substrate/product in subsite +1 (Fig. 6). The large hydrogen bonding capacity of arginine could well impede the availability of water required for deglycosylation.

Clearly, improvement of the T/H ratio is a prerequisite for obtaining mutants with good synthetic capability [32], especially if one considers that hydrolysis strongly influences the final product yield, both by competing with sugar acceptors for the deglycosylation of the glycosyl-enzyme and by hydrolysing the transglycosylation product. In this respect, in the initial stage of the reaction mutant 1753 exhibited a modest, but significant increase in transglycosylation and then lowered secondary hydrolysis of the arabinoxyloside product, whereas mutant 6 displayed a markedly greater diminution of secondary hydrolysis. Plotting the transglycosylation yield as a function of donor consumption (Fig. 4) confirms the increased level of transglycosylation for mutants 6 and 1753 and underlines the early phase intervention of this reaction, consistent with the fact that transglycosylation is a kinetically controlled phenomenon [47].

Previous studies that report on increased transglycosylation in glycosidases have supplied a variety of hypotheses, including decreased binding in the donor subsite [48,49], and thus extended lifetime of the glycosyl-enzyme intermediate, increased affinity for acceptor moieties [31,50–57] and modified binding of catalytic water molecules [58–61]. However, it is quite difficult to draw generic conclusions, especially because the efficient glycosynthetic activity displayed by 'true' transglycosylases is likely to be a result of a combination of all of these phenomena [62,63]. Nevertheless, in the case of *Tx*Abf using *p*-nitrophenyl β-D-galactofuranoside (a 5-hydroxymethyl analogue of *p*NP-α-L-Ara*f*), transglycosylation is significantly increased while hydrolysis is characterized by a high K_M value (>50 mM) and a low performance constant (k_{cat}/K_M) [25]. Moreover, recent work using STD-NMR has revealed that no significant STD effects can be measured when *p*-nitrophenyl *b*-D-galactofuranoside is bound to *Tx*Abf [34], suggesting that this donor molecule adopts a quite different position in the subsite 1 compared to *p*NP-α-L-Ara*f* and that the protons of the sugar moiety do not have any vicinal protein counterparts. Therefore, relating these findings to the present data suggest that transglycosylation in *Tx*Abf might be favoured by alterations to the configuration of the glycosyl-enzyme intermediate rather than intrinsic donor binding. Presumably, modifications of this nature would reduce water-mediated deglycosylation and/or favour sugar acceptor-mediated deglycosylation. Moreover, it appears likely that increased acceptor binding would also accentuate the transglycosylation reaction.

Conclusion

In conclusion, using random mutagenesis combined with a simple and efficient screen, we have shown that it is possible to modulate the catalytic properties of *Tx*Abf and thus advance towards the ultimate goal of creating an efficient glycosynthetic tool. This preliminary study has once again confirmed that random mutagenesis is a useful first approach to pinpoint functionally relevant amino acids that would be tedious to identify using rational methods [32,64] and, likewise, opens opportunities for further more targeted structure–function studies that will be vital to clarify the role of the different mutations detected herein.

Acknowledgements

The authors thank the French National Institute for Agricultural Research (INRA) and the French Environment and Energy Management Agency (ADEME) for providing a PhD stipend to F. Arab-Jaziri. B. Bissaro benefits from a PhD stipend from INRA. NMR experiments were performed at the 'Integrated Screening Platform of Toulouse' (PICT, IPBS, CNRS – Université de Toulouse), using equipment funded by GIS IBiSA and the French Research Ministry, CNRS, Université Paul Sabatier, the Région Midi-Pyrénées and European structural funds and at the Toulouse Center for Metabolomics (MetaToul), which is supported by grants from the Région Midi-Pyrénées, the European regional development fund, SICOVAL, CNRS and INRA.

References

[1] Rakić B, Withers SG. Recent developments in glycoside synthesis with glycosynthases and thioglycoligases. Aust J Chem 2009;62(6):510–20.
[2] Kittl R, Withers SG. New approaches to enzymatic glycoside synthesis through directed evolution. Carbohydr Res 2010;345(10):1272–9.
[3] Kadokawa J. Precision polysaccharide synthesis catalyzed by enzymes. Chem Rev 2011;111(7):4308–45.
[4] Schmaltz RM, Hanson SR, Wong C-H. Enzymes in the synthesis of glycoconjugates. Chem Rev 2011;111(7):4259–307.
[5] Fauré R, Saura-Valls M, Brumer III H, Planas A, Cottaz S, Driguez H. Synthesis of a library of xylogluco-oligosaccharides for active-site mapping of xyloglucan *endo*-transglycosylase. Journal of Organic Chemistry 2006;71(14):5151–61.
[6] Teeri TT, Brumer III H, Daniel G, Gatenholm P. Biomimetic engineering of cellulose-based materials. Trends Biotechnol 2007;25(7):299–306.
[7] Lopez M, Fort S, Bizot H, Buléon A, Driguez H. Chemo-enzymatic synthesis of xylogluco-oligosaccharides and their interactions with cellulose. Carbohydrate Polymers 2012;88(1):185–93.
[8] Ebringerova A, Heinze T. Xylan and xylan derivatives – biopolymers with valuable properties. Naturally occurring xylans structures, isolation procedures and properties. Macromolecular Rapid Communications 2000;21(9):542–56.
[9] Köhnke T, Ostlund A, Brelid H. Adsorption of arabinoxylan on cellulosic surfaces: influence of degree of substitution and substitution pattern on adsorption characteristics. Biomacromolecules 2011;12(7):2633–41.
[10] Neyrinck AM, Possemiers S, Druart C, Van de Wiele T, De Backer F, Cani PD, et al. Prebiotic effects of wheat arabinoxylan related to the increase in Bifidobacteria, Roseburia and Bacteroides/Prevotella in diet-induced obese mice. PLoS ONE 2011;6(6):e20944.
[11] Zhang Y, Pitkänen L, Douglade J, Tenkanen M, Remond C, Joly C. Wheat bran arabinoxylans: chemical structure and film properties of three isolated fractions. Carbohydrate Polymers 2011;86(2):852–9.
[12] Phan The D, Debeaufort F, Voilley A, Luu D. Biopolymer interactions affect the functional properties of edible films based on agar, cassava starch and arabi- noxylan

blends. Journal of Food Engineering 2009;90(4):548–58.
[13] Boukari I, Putaux J-L, Cathala B, Barakat A, Saake B, Rémond C, et al. In vitro model assemblies to study the impact of lignin–carbohydrate interactions on the enzymatic conversion of xylan. Biomacromolecules 2009;10(9):2489–98.
[14] Collins HM, Burton RA, Topping DL, Liao M, Bacic A, Fincher GB. Variability in fine structures of noncellulosic cell wall polysaccharides from cereal grains: potential importance in human health and nutrition. Cereal Chemistry 2010;87(4):272–82.
[15] Bouxin F, Marinkovic S, Le Bras J, Estrine B. Direct conversion of xylan into alkyl pentosides. Carbohydr Res 2010;345(17):2469–73.
[16] Broekaert WF, Courtin CM, Verbeke K, Van de Wiele T, Verstraete W, Delcour JA. Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides. Critical Reviews in Food Science and Nutrition 2011;51(2):178–94.
[17] Utille J-P, Jeacomine I. Synthesis of a library of allyl *a*-L-arabinofuranosyl-*a*- or *b*-D-xylopyranosides; route to higher oligomers. Carbohydr Res 2007;342: 2649–56.
[18] Helm RF, Ralphlb J, Anderson L. Regioselective protection strategies for D-xylopyranosides. Journal of Organic Chemistry 1991;56(25):7015–21.
[19] Koto S, Morishima N, Takenaka K, Uchida C, Zen S. Pentoside synthesis by dehydrative glycosylation. Synthesis of *O*-*a*-L-arabinofuranosyl-(1 ! 3)-*O*-*b*-D-xylopyranosyl-(1 ! 4)-D-xylopyranose. Bulletin of the Chemical Society of Japan 1985;58(5):1464–8.
[20] Hirsch J, Petrakova E, Schraml J. Stereoselective synthesis and ^{13}C-N.M.R. spectra of two isomeric methyl *b*-glycosides of trisaccharides related to arabinoxylan. Carbohydr Res 1984;131:219–26.
[21] Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, et al. The path forward for biofuels and biomaterials. Science 2006;311(5760):484–9.
[22] Dumon C, Song L, Bozonnet S, Fauré R, O'Donohue MJ. Progress and future prospects for pentose-specific biocatalysts in biorefining. Process Biochemistry 2012;47(3):346–57.
[23] Rémond C, Ferchichi M, Aubry N, Plantier-Royon R, Portella C, O'Donohue MJ. Enzymatic synthesis of alkyl arabinofuranosides using a thermostable *a*-L-arabinofuranosidase. Tetrahedron Lett 2002;43:9653–5.
[24] Rémond C, Plantier-Royon R, Aubry N, Maes E, Bliard C, O'Donohue MJ. Synthesis of pentose-containing disaccharides using a thermostable *a*-L-arabinofuranosidase. Carbohydr Res 2004;339(11):2019–25.
[25] Rémond C, Plantier-Royon R, Aubry N, O'Donohue MJ. An original chemoenzymatic route for the synthesis of *b*-D-galactofuranosides using an *a*-L-arabinofuranosidase. Carbohydr Res 2005;340(4):637–44.
[26] Chlubnová I, Filipp D, Spiwok V, Dvoráková H, Daniellou R, Nugier-Chauvin C, et al. Enzymatic synthesis of oligo-D-galactofuranosides and L-arabinofurano- sides: from molecular dynamics to immunological assays. Organic & Biomole- cular Chemistry 2010;8(9):2092–102.
[27] Sakamoto T, Fujita T, Kawasaki H. Transglycosylation catalyzed by a *Penicillium chrysogenum exo*-1,5-*a*-L-arabinanase. Biochim Biophys Acta 2004;1674:85–90.
[28] Wan C-F, Chen W-H, Chen C-T, Chang MD-T, Lo L-C, Li Y-K. Mutagenesis and mechanistic study of a glycoside hydrolase family 54 *a*-L-arabinofuranosidase from *Trichoderma koningii*. Biochemical Journal 2007;401(2):551–8.
[29] Marmuse L, Asther M, Fabre E, Navarro D, Lesage-Meessen L, Asther M, et al. New chromogenic substrates for feruloyl esterases. Organic & Biomolecular Chemistry 2008;6(7):1208–14.
[30] Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. Genome Res 1992;2:28–33.
[31] Feng H-Y, Drone J, Hoffmann L, Tran V, Tellier C, Rabiller C, et al. Converting a *b*-glycosidase into a *b*-transglycosidase by directed evolution. Journal of Biolo- gical Chemistry 2005;280(44):37088–97.
[32] Osanjo G, Dion M, Drone J, Solleux C, Tran V, Rabiller C, et al. Directed evolution of the *a*-L-fucosidase from *Thermotoga maritima* into an *a*-L-transfucosidase. Biochemistry 2007;46:1022–33.
[33] Koné FMT, Le Béchec M, Sine J-P, Dion M, Tellier C. Digital screening methodology for the directed evolution of transglycosidases. Protein Engineering Design & Selection 2009;22(1):37–44.
[34] Arab-Jaziri F, Bissaro B, Barbe S, Saurel O, Débat H, Dumon C, et al. Functional roles of H98 and W99 and *β2α2* loop dynamics in the *a*-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. FEBS Journal 2012;279(19):3598–611.
[35] Debeche T, Cummings N, Connerton I, Debeire P, O'Donohue MJ. Genetic and biochemical characterization of a highly thermostable *a*-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. Applied and Environment Microbiology 2000;66(4):1734–6.
[36] Gottlieb HE, Kotlyar V, Nudelman A. NMR chemical shifts of common laboratory solvents as trace impurities. Journal of Organic Chemistry 1997;62(21): 7512–5.
[37] Mayer M, Meyer B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J Am Chem Soc 2001;123(25):6108–17.
[38] Piotto M, Saudek V, Sklenár V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR 1992;2(6):661–5.
[39] Sinnott ML. Catalytic mechanisms of enzymic glycosyl transfer. Chem Rev 1990;90:1171–202.
[40] Rye CS, Withers SG. Glycosidase mechanisms. Curr Opin Chem Biol 2000;4:573–80.
[41] Mayer M, Meyer B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angewandte Chemie International Edition 1999;38(12):1784–8.
[42] Meyer B, Peters T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angewandte Chemie International Edition 2003;42(8):864–90.
[43] Paës G, Skov LK, O'Donohue MJ, Rémond C, Kastrup JS, Gajhede M, et al. The structure of the complex between a branched pentasaccharide and *Thermoba- cillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determi- nants and induced fit. Biochemistry 2008;47:7441–51.
[44] Debeche T, Bliard C, Debeire P, O'Donohue MJ. Probing the catalytically essential residues of the *a*-L-arabinofuranosidase from *Thermobacillus xylanily- ticus*. Protein Eng 2002;15(1):21–8.
[45] Wang Q, Trimbur D, Graham R, Warren RAJ, Withers SG. Identification of the acid/base catalyst in *Agrobacterium faecalis b*-glucosidase by kinetic analysis of mutants. Biochemistry 1995;34(44):14554–62.
[46] Bravman T, Belakhov V, Solomon D, Shoham G, Henrissat B, Baasov T, et al. Identification of the catalytic residues in family 52 glycoside hydrolase, a *b*-xylosidase from *Geobacillus stearothermophilus* T-6. Journal of Biological Chem- istry 2003;278(29):26742–49.
[47] Monsan P, Paul F. Enzymatic synthesis of oligosaccharides. FEMS Microbiology Review 1995;16(2–3):187–92.
[48] Aronson Jr NN, Halloran BA, Alexeyev MF, Zhou XE, Wang Y, Meehan EJ, et al. Mutation of a conserved tryptophan in the chitin-binding cleft of *Serratia marcescens* chitinase A enhances transglycosylation. Bioscience Biotechnology and Biochemistry 2006;70(1):243–51.
[49] Mori H, Bak-Jensen KS, Svensson B. Barley *a*-amylase Met53 situated at the high-affinity subsite 2 belongs to a substrate binding motif in the *b* ! *a* loop 2 of the catalytic $(b/a)_8$-barrel and is critical for activity and substrate specificity. Eur J Biochem 2002;269(22):5377–90.
[50] Johansson P, Brumer III H, Baumann MJ, Kallas AM, Henriksson H, Denman SE, et al. Crystal structures of a poplar xyloglucan endotransglycosylase reveal details of transglycosylation acceptor binding. Plant Cell 2004;16(4):874–86.
[51] Moreau A, Shareck F, Kluepfel D, Morosoli R. Alteration of the cleavage mode and of the transglycosylation reactions of the xylanase A of *Streptomyces lividans* 1326 by site-directed mutagenesis of the Asn173 residue. Eur J Biochem 1994;219(1–2):261–6.
[52] Armand S, Andrews SR, Charnock SJ, Gilbert HJ. Influence of the aglycone region of the substrate binding cleft of *Pseudomonas xylanase* 10A on catalysis. Biochemistry 2001;40(25):7404–9.
[53] Taira T, Fujiwara M, Dennhart N, Hayashi H, Onaga S, Ohnuma T, et al. Transglycosylation reaction catalyzed by a class V chitinase from cycad, *Cycas revoluta*: a study involving site-directed mutagenesis, HPLC, and real-time ESI- MS. Biochim Biophys Acta 2010;1804(4):668–75.
[54] Lü Y, Yang H, Hu H, Wang Y, Rao Z, Jin C. Mutation of Trp137 to glutamate completely removes transglycosyl activity associated with the *Aspergillus fumi- gatus* AfChiB1. Glycoconj J 2009;26(5):525–34.
[55] Umekawa M, Huang W, Li B, Fujita K, Ashida H, Wang L-X, et al. Mutants of *Mucor hiemalis endo-b-N*-acetylglucosaminidase show enhanced transglycosyla- tion and glycosynthase-like activities. Journal of Biological Chemistry 2008;283(8):4469–79.
[56] Fan S-Q, Huang W, Wang L-X. Remarkable transglycosylation activity of glycosynthase mutants of *endo*-D, an *endo-b-N*-acetylglucosaminidase from *Strepto- coccus pneumoniae*. Journal of Biological Chemistry 2012;287(14):11272–81.
[57] Champion E, André I, Moulis C, Boutet J, Descroix K, Morel S, et al. Design of *a*-transglucosidases of controlled specificity for programmed chemoenzymatic synthesis of antigenic oligosaccharides. J Am Chem Soc 2009;131:7379–89.
[58] Kuroki R, Weaver LH, Matthews BW. Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. Proc Natl Acad Sci USA 1999;96(16):8949–54.
[59] Honda Y, Fushinobu S, Hidaka M, Wakagi T, Shoun H, Taniguchi H, et al. Alternative strategy for converting an inverting glycoside hydrolase into a glycosynthase. Glycobiology 2008;18(4):325–30.
[60] Hidaka M, Fushinobu S, Honda Y, Wakagi T, Shoun H, Kitaoka M. Structural explanation for the acquisition of glycosynthase activity. J Biochem (Tokyo) 2010;147(2):237–44.
[61] Kuriki T, Kaneko H, Yanase M, Takata H, Shimada J, Handa S, et al. Controlling substrate preference and transglycosylation activity of neopullulanase by manipulating steric constraint and hydrophobicity in active center. Journal of Biological Chemistry 1996;271(29):17321–29.
[62] Mark P, Baumann MJ, Eklöf JM, Gullfot F, Michel G, Kallas AM, et al. Analysis of nasturtium *Tm*NXG1 complexes by crystallography and molecular dynamics provides detailed insight into substrate recognition by family GH16 xyloglucan *endo*-transglycosylases and *endo*-hydrolases. Proteins Structure Function and Bioinformatics 2009;75(4):820–36.
[63] Baumann MJ, Eklöf JM, Michel G, Kallas AM, Teeri TT, Czjzek M, et al. Structural evidence for the evolution of xyloglucanase activity from xyloglucan *endo*-transglycosylases: biological implications for cell wall metabolism. Plant Cell 2007;19(6):1947–63.
[64] Kelly RM, Leemhuis H, Rozeboom HJ, van Oosterwijk N, Dijkstra BW, Dijkhuizen L. Elimination of competing hydrolysis and coupling side reactions of a cyclodextrin glucanotransferase by directed evolution. Biochemical Journal 2008;413(3):517–25.

Supporting Information A.3: ***Submitted Article***

Enhancing the the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α-L-arabinofuranosidase

Faten Arab-Jaziri,[a,b,c,d] Bastien Bissaro,[a,b,c] Charles Tellier,[e] Michel Dion,[e] Régis Fauré,[a,b,c] and Michael J. O'Donohue[a,b,c,#]

Université de Toulouse; INSA,UPS,INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France[a]; INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France[b]; CNRS, UMR5504, F-31400 Toulouse, France[c]; Agence de l'environnement et de la Maîtrise de l'Energie, 20 Avenue du Grésillé - BP 90406, 49004 Angers Cedex 01, France[d]; and UFIP, UMR CNRS 6286, Faculté des Sciences et Techniques, Université de Nantes, 2 rue de la Houssinière, F-44322 Nantes, France[e]

ABSTRACT

Random mutagenesis was performed on the α-L-arabinofuranosidase of *Thermobacillus xylanilyticus* in order to enhance its ability to perform transarabinofuranosylation using natural xylo-oligosaccharides as acceptors. To achieve this goal, a high-throughput digital imaging screening protocol was used to study a library of 30,000 mutants. Thereby, 199 mutants displaying lowered hydrolytic activity and modified behavior in the presence of a xylo-oligosaccharide mixture were detected. In the presence of these acceptors, most of the 199 (i.e. 70%) enzymes were less inhibited and some (18) mutants displayed an unambiguous alleviation of inhibition (< 25% loss of activity). More precise monitoring of reactions, performed in the presence of xylobiose (acceptor) and catalyzed by the most promising mutants, revealed a significant improvement of the synthesis yields of transglycosylation products (up to 18% compared to 9% for the parental enzyme). Sequence analysis of the improved mutants revealed that many of the amino acid substitutions are putatively responsible for the new phenotype are located in the vicinity of the active site, and precisely in the subsite -1. Consequently, we hypothesize that these mutations modify the active site topology or the molecular interaction network of the L-arabinofuranoside donor substrate, thus crippling the hydrolytic mechanism and concomitantly promoting transglycosylation onto natural acceptors.

Keywords: Glycoside Hydrolase, Transglycosylation, Pentose/Furanose, Xylo-oligosaccharide, Inhibition

Address correspondance to Michael J. O'Donohue, michael.odonohue@insa-toulouse.fr

Running title: Transarabinofuranosylation of short XOS by Abf mutants

Abbreviations: Abf, α-L-arabinofuranosidase; AXOS, arabinoxylo-oligosaccharide; epPCR, error-prone PCR; FH, furanoside hydrolase; GH, glycoside hydrolase; HTS, high-throughput screening; IPTG, isopropyl 1-thio-β-D-galactopyranoside; L-Ara*f*, L-arabinofuranosyl; *p*NP, *para*-nitrophenol; *p*NP-α-L-Ara*f*, *para*-nitrophenyl α-L-arabinofuranoside; *Tx*Abf, α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*; X-α-L-Ara*f*, 5-bromo-4-chloro-3-indolyl α-L-arabinofuranoside; XOS, xylo-oligosaccharide.

1. Introduction

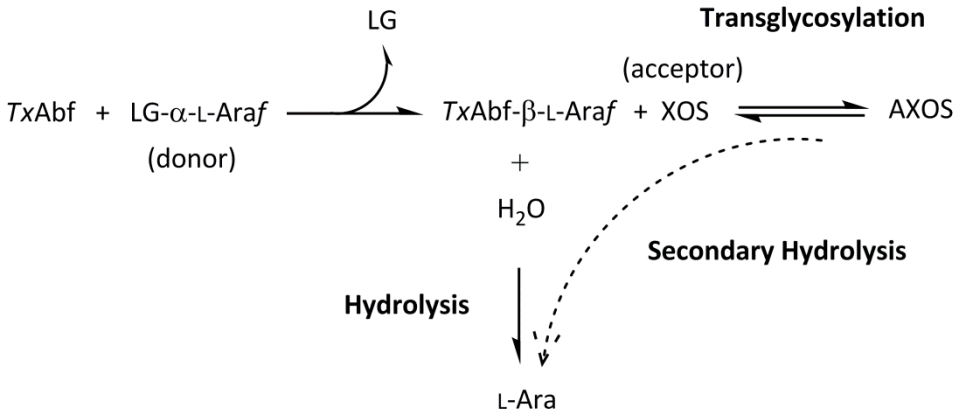
Despite considerable progress in the field of synthetic glycochemistry, the practical and scalable access to complex carbohydrate structures remains a major challenge, involving labour-intensive processes and generally leading to limited overall yields. In recent years, biotechnology has increasingly penetrated organic chemistry, providing complementary or alternative tools, notably for the synthesis of sugar-based compounds (1). This is because enzymes present several advantages when compared to traditional catalysts. Notably, carbohydrate-active enzymes perform stereospecific (i.e. anomeric specificity) and often regioselective reactions, operate in mild conditions and obviate the need for protecting groups, which are required in classical glycochemistry to target the formation of glycosidic bonds. In Nature, glycosyltransferases are the major players in synthesis of carbohydrates and glycomotifs, but for *in vitro* reactions, 'glycan phosphorylases' and glycoside hydrolases (GHs) have also proved to be useful biocatalysts for the synthesis of tailor-made glycosidic assemblies (2, 3).

GHs have been quite extensively used for glycosynthesis, especially because these enzymes are abundant and often readily available in recombinant protein form. Nevertheless, when functioning in transglycosylation mode, the drawback of GHs is that glycosynthesis is in competition with primary and secondary hydrolysis (i.e. hydrolysis of the donor substrate and the transglycosylation product, respectively), which usually remains the dominant activity, thus limiting the overall yield of synthesis products (scheme 1).

The development of GHs as glycosynthetic tools has mainly focused on hexose-acting enzymes, although the use of xylanases for the synthesis of D-xylopyranosyl-containing compounds has also been described by several authors (4–9). However, examples of glycosynthesis catalyzed by furanoside hydrolases (FHs), and in particular pentose-acting FHs, are quite scarce mainly because FHs represent a relatively small subgroup of GHs and furanose chemistry is not a mainstream topic, in part because the modification of these sugars can be quite tricky. Nevertheless, the synthesis of furanose-based compounds is useful for the synthesis of chemotherapeutic agents (10, 11) and for the production of well-defined plant cell wall motifs containing L-arabinofuranosyl (L-Ara*f*) moieties (12–15), and could be valuable in the future to transform L-arabinose, arising from biomass-based biorefineries, into products such as prebiotics and health-related derivatives (16–18), films, biopolymers and biomaterials (19, 20), as well as surfactants.

The first report of a FH-catalyzed synthesis reaction described the use of an α-L-arabinofuranosidase (EC 3.2.1.55) from *Thermobacillus xylanilyticus* (*Tx*Abf) for the synthesis of alkyl arabinofuranosides (21). Since then, the glycosynthetic potential of *Tx*Abf has been further detailed (22–25) and glycosynthetic reactions catalyzed by other L-arabinofuranose-specific enzymes have been reported (26–29). Regarding *Tx*Abf, this thermostable enzyme is one of the well-characterized members of family GH51 of the CAZy classification (30). The hydrolytic activity of *Tx*Abf has been characterized and its 3D structure is known (31). Moreover, in another recent study, variants of *Tx*Abf have been created and screened using a random mutagenesis approach coupled to a high-throughput assay, which has led to the successful identification of mutant *Tx*Abf that display slightly improved glycosynthetic activity in the presence of *para*-nitrophenyl α-L-arabinofuranoside (*p*NP-α-L-Ara*f*) and benzyl α-D-xylopyranoside, which act as donor and acceptor sugars respectively, combined with a reduced secondary hydrolysis of the transglycosylation product (32).

In the present work, we have once again focused on the improvement of the glycosynthetic potential of *Tx*Abf, using a random mutagenesis approach and a high-throughput screening (HTS) strategy that is



Scheme 1. Reactions catalyzed by the retaining GH51 *Tx*Abf (leaving group: LG = *p*NP, X or sugar; xylo-oligosaccharide: XOS; arabinoxylo-oligosaccharide: AXOS)

based on the approach developed by Koné *et al.* (33). The aim of this work was to isolate, using a two-step screening protocol, mutated variants that display an increased ability to synthesize small arabinoxylan-type motifs using *p*NP-α-L-Ara*f* as the donor and xylo-oligosaccharides (XOS) as acceptors.

2. Experimental

2.1. Materials. Unless otherwise stated, routine experimental was performed using chemicals purchased mainly from Sigma-Aldrich (France) and molecular biology reagents purchased from New England Biolabs (USA). The substrates, *p*NP-α-L-Ara*f* and 5-bromo-4-chloro-3-indolyl α-L-arabinofuranoside (X-α-L-Ara*f*) were synthesized in-house according to established or adapted protocols (21, 34). Xylotriose and the XOS were purchased from Wako Chemicals GmbH (Germany). Xylobiose was prepared in-house using XOS as the starting material and sequential per-*O*-acetylation, purification and *O*-deacetylation as previously described (35).

2.2. Analysis of xylo-oligosaccharides. To determine the composition of XOS and assess the purity of xylobiose, high performance anion exchange chromatography, with pulsed amperometric detection (HPAEC-PAD), was performed on XOS solutions (20-25 mg l^{-1} in water) using XOS standards and a Dionex system (Dionex corporation, CA, USA), equipped with a CarboPac PA-100 Guard Column (Dionex, 4 x 50 mm) combined with a CarboPac PA-100 Analytical Column (Dionex, 4 x 250 mm) thermostated at 30°C. Separation was achieved by applying a linear gradient of NaOAc (5 to 45 mM) over 15 min, in a 150 mM NaOH solution, at a flow rate of 1 ml min^{-1}. Analysis of XOS mixture revealed that the major species were xylobiose (26.9%, w/w) and xylotriose (9.6%, w/w), while the remaining material (63.5%, w/w) was a mixture of xylose and other XOS (DP > 3).

2.3. Subcloning, expression and purification of *Tx*Abf. For mutagenesis, the sequence encoding *Tx*Abf (GenBank CAA76421.2) was C-terminally fused to a $(His)_6$ tag in the plasmid pET24. The coding sequence was thus amplified by PCR using the following oligonucleotide primer pair, 5'-GGAATTCCATATGAACGTGGCAAGCC GGGTAGTCG (forward primer containing a *Nde* I site) and 5'-CCCAAGCTTCCCCGCCGTCAGCTCAA GCACC (reverse primer containing a *Hind* III site), and then introduced into pET24, between the *Nde* I and *Hind* III sites and downstream of the T7 promoter. The resulting vector (pTXABF) was used to transform *E. coli* BL21 DE3 star strain (Invitrogen). IPTG-induced expression and subsequent purification, using cobalt-based

metal affinity chromatography, of *Tx*Abf-$(His)_6$ (hereafter referred to as *Tx*Abf) and mutants thereof were performed as previously described (36).

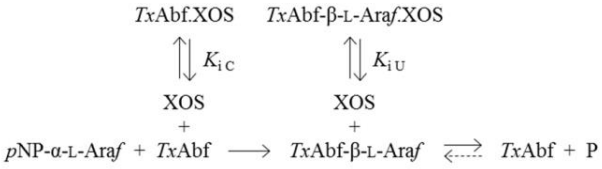
2.4. Creation of randomly mutated DNA libraries. To randomly mutagenize the *Tx*Abf encoding sequence, a previously described epPCR-based protocol (37) was employed, using two primers that flank the target sequence. The forward primer was the same as the one described above, while the reverse primer (5'-CGATGGCCCACTACGTGAACCATC) was specific to the experiment. To achieve random mutagenesis, reaction mixtures (100 µl total volume) containing pTXABF (12 ng), 0.2 mM dATP and dGTP, 1 mM dTTP and dCTP, 7 mM $MgCl_2$, $MnSO_4$ (either 0.05, 0.1 or 0.15 mM) and 5 IU of *Taq* polymerase were prepared in 1x Thermopol buffer and submitted to the following thermocycling protocol: one initial denaturing step at 94°C for 3 min, followed by 30 cycles at 94°C for 30 s, 55°C for 45 s, 72°C for 2 min and a final elongation step at 72°C for 7 min. Afterwards, the epPCR products obtained in the different reactions were ligated to pET24a as before. At the ligation step, 8 to 10 ligation reactions (10 µl each), each containing 400 IU of T4 DNA ligase, per library were prepared and incubated for 1 h at room temperature (25°C). After, a second aliquot (400 IU) of T4 DNA ligase was added to each reaction and incubation was pursued for 1 h, before being used to transform competent *E. coli* Top10 cells. Fourteen transformation reactions per library were performed (1 µl of ligation reaction per transformation, using 50 µl of competent bacteria) and transformants were spread onto Q-trays (22 x 22 cm) containing selective LB agar medium (including kanamycin at 40 µg ml^{-1}). Colonies were allowed to develop for 16 h at 37°C before being pooled and suspended in 50 ml LB liquid medium. Each *Tx*Abf library was prepared in plasmid DNA form using a plasmid purification kit (Maxi kit, Qiagen) and stored at -20°C.

2.5. Implementation of the primary two-step screening protocol. To screen *Tx*Abf libraries, 10 ng of plasmid DNA stock were used to transform *E. coli* BL21 DE3 star cells, which were then grown at 37°C for 12 h on a nitrocellulose membrane (0.45 µm, 20 x 20 cm) placed on selective LB agar medium. Following the development of small colonies (Ø < 0.5 mm), the membrane was transferred onto selective minimal medium (100 mM sodium phosphate buffer pH 7, agar 15 g l^{-1}, kanamycin 40 µg ml^{-1}) containing IPTG (0.4 mM as final concentration) and X-α-L-Ara*f* (0.1 mM in DMSO), and incubated for 1 h at 30°C. After, the membrane was transferred onto fresh minimal medium containing X-α-L-Ara*f* (0.1 mM in DMSO) and XOS (11.3 g l^{-1}) and immediately scanned (Epson 4490, 16 bits grey levels, 600 dpi), generating an image (I_0) that could be analysed using Visilog 6.3 image analysis software (Noesis, France). Afterwards, the membrane was incubated at 37°C for 4 h, before a second image (I_f) was obtained. Images I_0 and I_f were then subjected to processing according to Koné *et al.* (33), which led to colour detection, colony recognition and allowed the exact superposition of the two images. Finally, the ratio $R = (I_f - I_0)/I_0$ was calculated for each colony and classified in a descending order.

2.6. Characterization of the wild-type *Tx*Abf and mutants. For rapid secondary screening of the acceptor impact on the activity of mutants from pre-purified cell extracts (32), following their initial selection on solid medium, a single concentration of XOS (11.3 g l^{-1}) was used to acquire a measurement of residual activity in presence of *p*NP-α-L-Ara*f* (5 mM) at 40°C. To further evaluate acceptors impact on the activity of the wild-type *Tx*Abf and mutants with improved residual activity, the hydrolysis of *p*NP-α-L-Ara*f* (from 0.1 to 7 mM) at 40°C in the presence of different concentrations of XOS (from 0 to 36.7 g l^{-1}), xylobiose (from 0 to 80 mM) and xylotriose (at 40 mM) was quantified. Routinely, the hydrolytic activity

of wild-type *Tx*Abf and mutants was spectrophotometrically assayed as previously described (38). To model the inhibitory properties of the different xylo-oligosaccharides, classical equations for uncompetitive and non-competitive inhibition were employed, making the assumption that because transglycosylation is a minor reaction (< 10% for the wild-type) it is unnecessary to include it in scheme 2. Obviously, this implies that values for K_{iC} and K_{iU} are only approximations.

2.7. Monitoring transglycosylation. To investigate the ability of the different enzymes to perform transglycosylation, two techniques were employed. First, analytical TLC (as described above) was used to monitor reactions performed at 40°C, using 0.11 IU of enzyme, *p*NP-α-L-Ara*f* (5 mM) and xylobiose (10 mM) (donor/acceptor ratio of 1:2). Reactions were stopped when the donor was fully consumed. For finer analysis, reactions were monitored by ^{1}H NMR, collecting spectra using a Bruker Avance 500 spectrometer (36). Transglycosylation reactions (600 µl) were performed at 40°C in D_2O (99.90%) using *p*NP-α-L-Ara*f* (5 mM), xylobiose (10 mM) and enzyme (0.05 IU - corresponding to enzyme concentration of 8, 524, 89, 436 nM and 3.06 and 1.62 µM for the *Tx*Abf and mutants 1, 8, 169, 2032 and 1068, respectively). For second shell mono-mutants, reactions were carried out at 45°C in deuterated sodium acetate buffer (pD 5.8, 25 mM) at 45°C with *p*NP-α-L-Ara*f* (5 mM), xylotriose (10 mM) and enzyme (0.05 IU - corresponding to concentrations of 8, 13, 27, 141, 11 and 50 nM for the *Tx*Abf and mutants V173D, Q316L, N344I, I345L and L352M). All reagents, including enzymes, were prepared in D_2O or deuterated buffer. For enzymes, enzyme solutions were diluted 10-fold in D_2O and then concentrated (10-fold) using an Amicon® Ultra 10 kDa device (Millipore). This procedure was performed twice. ^{1}H NMR spectra were accumulated continuously over 5.5 min periods (32 scans) during 3 to 10 hours, depending on the activity of the mutants. Chemical shifts (δ) were expressed in parts per millions (ppm) and the residual water peak was used as the internal reference (39). High-resolution mass spectra (HRMS) were acquired at the CRMPO (Rennes University, France) using a Waters Q-Tof 2 spectrophotometer, equipped with an electrospray ionization source, operating in positive ion detection mode. HRMS (ESI) for the transarabinofuranosylation onto xylobiose: *m/z* : calcd for $C_{15}H_{26}O_{13}Na$: $[M+Na]^+$ 437.12711; found: 437.1278-437.1283 (2-3 ppm).



$$v = \frac{V_{max}.[S]}{K_M.(1 + [I]/K_{i\,C}) + [S].(1 + [I]/K_{i\,U})} \quad \textbf{(Eq. 1)}$$

Scheme 2. Kinetic mechanism for non-competitive inhibition (i.e. a special case of mixed inhibition, with $K_{i\,C} = K_{i\,U}$) of *Tx*Abf by XOS. The fact that XOS can act as acceptors for transglycosylation was not specifically considered (because this is a minor occurrence) and thus product (P) is a global term that includes the outcomes of hydrolysis and transglycosylation.

3. Results

3.1. Library characteristics and primary screening.

Three randomly mutagenized libraries were created. A preliminary assessment of each library revealed that global mutation rates increased with Mn^{2+} concentration (up to 7.0 point mutations per kb). Concomitantly, the number of white colonies (i.e. inactive mutants) increased from 47 to 69% of the population and proportion of dark blue colonies (i.e. fully active mutants and residual wild-type background) decreased from 38 to 25%. Concerning light blue colonies, which are of particular interest in

the context of this study (29), these represented 15% of the population in a library prepared using a Mn^{2+} concentration of 0.05 mM. This library, which was retained for further study and screening, also displayed less than 50% of inactive mutants and an overall mutation rate of 3.8/kb (i.e. mean average of 4.0 amino acid substitutions per *Tx*Abf-encoding sequence). Generally, mutation rates correlated well with the observed phenotypes. Accordingly, the sequencing of plasmid DNA from dark, light blue and white colonies revealed average mutation rates of 1.6, 4.2 and 5.8 amino acid substitutions per *Tx*Abf-encoding sequence. Moreover, when only considering mutations in the $(\beta/\alpha)_8$ catalytic domain, mutants displaying the light blue phenotype were characterized by an average mutation of 3.1 amino acid substitutions per catalytic domain.

A two-step procedure developed by Koné *et al.* (33) was employed to screen 30,000 library clones (i.e. from the $[Mn^{2+}]$ = 0.05 mM library), indirectly monitoring the release of the chromogenic indolyl aglycon from X-α-L-Ara*f* (scheme 1), first in the absence and then in the presence of XOS (acceptor sugars). The logic of the screen was first to detect hydrolytically-impaired mutants, which were further differentiated according to their behavior in the presence of sugar acceptors. Colonies whose coloration increased in the second step were deemed to bear potentially interesting mutants, since previous work has shown that this is a quite good way to detect mutants that display a greater ability to catalyze transglycosylation (32, 40, 41). Accordingly, after classification of the different colonies according to the measured R values (Fig. S1 and Table 1), 199 colonies were selected for further analysis.

3.2. Inhibition of *Tx*Abf by xylo-oligosaccharides. When using the release of the chromogenic leaving group *p*NP from *p*NP-α-L-Ara*f* as an indicator of *Tx*Abf activity, one is actually exclusively monitoring the progress of the first step (glycosylation) of the double-displacement mechanism (Scheme 1), which is common to hydrolysis and transglycosylation. Therefore, to extend the informational content of the available data and reveal preliminary details about enzyme-acceptor molecule interactions (i.e. involved in the deglycosylation step), *p*NP release was also monitored in the presence of different concentrations of XOS (i.e. putative acceptors). This revealed that *Tx*Abf was strongly inhibited by XOS (Fig. 1), similar to *Tt*βgly that was shown to be inhibited by maltose and cellobiose (33, 40). Moreover, no detectable hydrolysis of the XOS by *Tx*Abf was observed, indicating that XOS do not occupy subsite -1 in the active site.

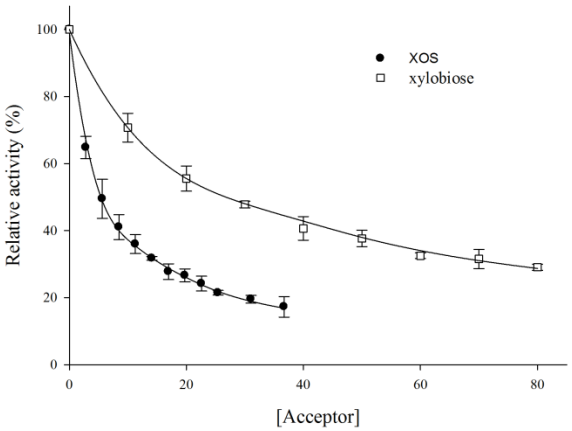


Figure 1. Wild-type *Tx*Abf inhibition profile with XOS mixture and xylobiose. Overall enzyme activity was determined by monitoring the release of *p*NP during *Tx*Abf-catalyzed reactions containing *p*NP-α-L-Ara*f* (5 mM) and different concentrations of XOS (from 0 to 36.7 g/L) or xylobiose (from 0 to 80 mM). Relative activities were derived by comparing enzyme activity in the presence and absence (i.e. 152 IU/mg, defined as 100% activity) of acceptors. Measurements were performed at least in duplicate, at 40 °C in 50 mM sodium acetate buffer (pH 5.8).

The effect of increasing XOS concentrations on the *Tx*Abf-catalyzed hydrolysis of *p*NP-α-L-Ara*f* caused a decrease in apparent V_{max} values, while apparent K_M values remained relatively unchanged, an observation that is

consistent with a non-competitive inhibition model (Scheme 2, Eq. 1 and Fig. S2) in which XOS can bind to both the free enzyme and the glycosyl-enzyme intermediate (*Tx*Abf-β-L-Ara*f*), thus forming unproductive and/or highly rate-limiting complexes. This is further supported by the fact that calculation of the inhibition constants $K_{i\ C}$ and $K_{i\ U}$ (representing competitive and uncompetitive modes respectively) revealed very similar values (10.9 ± 3.1 and 13.0 ± 4.0 g/L, respectively). Finally, it is noteworthy that the results of preliminary experiments investigating inhibition by pure xylobiose or xylotriose suggested that inhibition becomes uncompetitive, suggesting that a low productivity complex of the type *Tx*Abf-β-L-Ara*f*.xylobiose (or xylotriose) was being formed (Figs. S3 and S4). However, these observations could not be fully confirmed due to the requirement for very high concentrations of pure xylo-oligosaccharides.

3.3. Secondary screening of *Tx*Abf mutants. Of the 199 mutants selected for secondary screening, three failed to grow. Therefore only 196 were retained for enzyme expression and the preparation of cell lysates, which were used to perform reactions with *p*NP-α-L-Ara*f* in the presence or absence of XOS. This screening phase revealed that all mutants exhibited significantly diminished hydrolytic activity and that approximately 70% displayed improved relative activity (*p*NP release; data not shown) in the presence of XOS at 11.3 g l^{-1} (i.e. 10.8 mM of xylobiose), compared to *Tx*Abf, which was strongly inhibited (38% of residual activity; Fig 1). Among the cell lysates that were characterized by higher activity in the presence of XOS, 38 also displayed quite weak (i.e. < 30% reduction in activity) or no XOS-mediated inhibition. Within the latter group, 18 mutants showed a residual activity > 75% (referred as 1, 1d, 8, 169, 195, 215, 353, 449, 522, 555, 575, 639, 668, 1013, 1352, 1421, 2032 and 2068) and some even revealed no signs of XOS-mediated inhibition (i.e. activity > 100%). All of these were expressed, purified and further investigated, measuring reaction kinetics in the presence of XOS, xylobiose and xylotriose. While *Tx*Abf was strongly inhibited by both XOS and xylobiose (Fig. 1), the 18 selected mutants exhibited an alleviation of acceptor-mediated inhibition (Table 1) and mutants 1, 169, 449, 639, 1013, 2032 and 2068 even appeared to be slightly activated (> 100% residual activity) by xylobiose at 40 mM, compared to the wild-type (59% loss of activity). Additionally, mutants 1 and 639 exhibited no apparent inhibition in the presence of xylotriose at 40 mM, whereas parental *Tx*Abf only displays 17% residual activity in identical conditions.

3.4. Further characterization of selected mutants. The optimal temperatures for activity of the 18 mutants were determined and their corresponding specific activities were measured (Table 1). Interestingly, compared to *Tx*Abf, the optimal temperature was lowered for all mutants and hydrolytic activity was generally low and sometimes barely detectable. In contrast, three mutants (1d, 522 and 555) displayed higher residual hydrolytic activity. Regarding the potency of transglycosylation, qualitative TLC analysis of products formed in reactions using *p*NP-α-L-Ara*f* (5 mM) as donor and xylobiose (10 mM) as acceptor revealed that half of the 18 mutants (1, 1d, 8, 169, 639, 1013, 1421, 2032 and 2068) produced products that were only poorly synthesized by *Tx*Abf itself, whose activity was mainly directed towards the hydrolysis of *p*NP-α-L-Ara*f* and the very weak formation of a self-condensation product. ^{1}H NMR monitoring of the reactions catalyzed by the aforementioned nine mutants in the presence of *p*NP-α-L-Ara*f* and xylobiose as the donor and acceptor respectively revealed that no free xylose was detected, but in some cases xylobiose was consumed and

Table 1 Characterization of mutants. Residual hydrolytic activities of mutants, expressed as a percentage of wild-type *Tx*Abf activity, was measured in the presence of *p*NP-α-L-Ara*f* and various acceptors.[a]

Enzyme	XOS	Xylobiose	Xylobiose	Xylotriose	R value[c]	O.T.[d]	S.A.[d]
wt	38	71	41	17		70	587
1	91	100	106	102	μ + 3σ	50	6
1d	76	87	96	66	μ + 4σ	60	155
8	70	90	94	71	μ + 3σ	60	26
169	90	95	101	47	μ + 3σ	30	22
195	63	94	93	33	μ + 5σ	50	84
215	75	95	91	69	μ + 4σ	40	9
353	89	93	99	85	μ + 6σ	40	59
449	89	88	115	92	μ + 4σ	40	8
522/555[e]	82	90	84	78	μ + 6σ/ μ + 5σ	60	262
575	76	88	83	56	μ + 8σ	50	38
639	89	88	103	102	μ + 1σ	60	24
668	63	89	69	32	μ + 6σ	30	19
1013	95	101	107	88	μ + 2σ	40	14
1352	73	94	77	46	μ + 3σ	40	26
1421	78	97	100	99	μ + 3σ	50	21
2032	88	89	102	78	μ + 4σ	50	6
2068	81	94	108	n.d.[f]	μ + 5σ	40	4

[a] Characterization was performed on 18 purified enzymes. The *p*NP-α-L-Ara*f* concentration was 5 mM and reactions were performed at 40°C in sodium acetate buffer (50 mM, pH 5.8). Relative errors were <12%, though assays using xylotriose were only performed once.

[b] At 11.3 g l^{-1}, the concentrations of xylobiose and xylotriose were 10.8 and 2.6 mM, respectively.

[c] Ratio mean value of the Q-tray (μ) and standard deviation (σ) corresponding to the primary two-steps screening.

[d] Optimal temperatures for activity (O.T.) and specific activities (S.A.) were determined using *p*NP-α-L-Ara*f* (5 mM). S.A. were determined in triplicate at the O.T. of each mutant. Relative errors were < 8% (except for mutant 2068, 16%). Wild-type activity at 40°C was 152 IU.mg^{-1}.

[e] The genetic analysis of mutants 522 and 555 revealed that these are identical.

[f] n.d., not determined.

transglycosylation products, distinct from the self-condensation product, were formed. These transglycosylation products were identical to those produced by *Tx*Abf in the same conditions, with the wild type enzyme producing XA^3, A^3X and A^2X (describing an α-(1,3)-linked L-Ara*f* unit of the reducing D-Xyl*p* unit of xylobiose, an α-(1,3) and an α-(1,2)-linked L-Ara*f* substituents of the non-reducing D-Xyl*p* unit of xylobiose, respectively, according to nomenclature based on Fauré et al. (42)) at 1, 6 and 2% yields respectively. Mutants 1d and 639 did not display glycosynthetic activity, while mutants 1013 and 1421 actually produced slightly less transglycosylation products (8 and < 3% maximum overall yield, respectively) than *Tx*Abf. However, in the case of mutants 8, 169 and 2032, the yield of transglycosylation products was increased (11 to 13% overall yields) and actually attained 18 and 17% in reactions catalyzed by 1 and 2068 respectively (Table 2).

Table 2 Transglycosylation product yields calculated from time-course NMR kinetics of wild-type *Tx*Abf and mutants thereof.[a]

		Transglycosylation product Yield (%)			Maximum overall yield[b]
		XA^3	A^3X	A^2X	
	δ (ppm)	5.40	5.32	5.28	
Enzyme	wt	1	6	2	9
	1	3	10	5	18
	8	2	7	3	12
	169	1	6	4	11
	2032	2	8	3	13
	2068	3	10	4	17

[a] Transglycosylation reactions were performed in D_2O at 40°C with *p*NP-α-L-Ara*f* (5 mM) as donor, xylobiose (10 mM) as acceptor and 0.05 IU of enzyme.

[b] Maximum transglycosylation product yield is the sum of individual product yields, expressed as a percentage of the limiting donor concentration, which can be reached at different times depending on the regioisomer. The relative mean error was 1.5%.

Unsurprisingly, in most cases once formed the transglycosylation products were subject to secondary hydrolysis, although this was not the case for mutant 169, whose maximum yield remained stable even after prolonged incubation (Fig. 2). The arabinoxylo-oligosaccharides (AXOS) formed were characterized by mass spectrometry, revealing a mass of 414 g mol^{-1} that corresponds to that of an arabinoxylobioside. Furthermore, the structures of the transglycosylation products were consistent with those previously reported (43–46). Indeed, the signals of anomeric protons at 5.39, 5.33, and 5.28 ppm are fully consistent with XA^3, A^3X and A^2X respectively.

3.5. Structure-function analysis. DNA sequence analysis performed on the 38 mutants that displayed significantly reduced XOS-mediated inhibition revealed a total of 112 different mutations located within the $(\beta/\alpha)_8$ catalytic domain, with 33 of these being localized at only 11 amino acid positions (i.e. 11 amino positions were targets for multiple mutations). Significantly, when considering the subgroup of 18 mutants that was submitted to more detailed characterization, some of its members were found to be mutated at these hotspots. In particular mutations at F113, K251, V264, M272, Q316 and N344 occur three to five times among the 18 mutants. Moreover, it is noteworthy that among the 68 mutated amino acid positions that characterize the 18 mutant-subgroup, two thirds of the mutations are located within the catalytic $(\beta/\alpha)_8$ barrel and 30% within 15 Å of the Cα of the catalytic nucleophile, E298. In particular, three mutations are located close to W248, and six are in the vicinity of W302, two important residues that interact through van der Waals forces with the acceptor moiety in subsite +1 (Fig. 3) (31). Furthermore, four mutations were found to be close to E28 and three close to Q347, two amino acids that hydrogen bond with the L-Ara*f* moiety in subsite -1. Finally, six mutations are localized behind the catalytic nucleophile.

Regarding the 18 mutant-subgroup, mutants 1, 8, 169, 2032 and 2068 appear to be particularly interesting from the standpoint of the point mutations that they contain (Table 3). In mutant 1, the amino acid substitution L352M is positioned close to the primary hydroxyl group of the L-Ara*f* moiety, the distance of the Cδ (L352) from the OH-5 group being 3.5 Å, and near to W302, W248 and Q347 (distances between the different side chains are about 3.5, 4.8 and 5.0 Å). Regarding mutant 8, residue M272 is positioned behind the catalytic nucleophile E298, with the distance Cε (M272)-Cα (E298) being 7.5 Å. Mutant 169

contains two so-called second shell mutations, Q316L and I345L that are located near to W302 (4.3 Å between Cα (Q316) and Cα (W302)) and Q347 (Cα (I345)-Cα (Q347) being 6.1 Å) respectively, the latter lying just behind the catalytic nucleophile. Among the point mutations that characterize mutant 2032, V173D is in the vicinity of the catalytic dyad, the distances Cα (V173)-Cα (E176) and Cα (V173)-Cα (E298) being 10.1 and 14.9 Å respectively. Finally, mutations at N344 characterize the mutants 2032 and 2068. The presence of N344 mutations is significant, because this residue has already been pinpointed in previous mutagenesis work (29) and has been shown to be a pH modulator and a key residue interacting with E298 (36), the Cα (N344)-Cα (E298) distance being 5.5 Å. Taken together, these findings reveal that mutations that can potentially confer new transglycosylation properties to *Tx*Abf occupy quite different locations within the *Tx*Abf structure, but are nevertheless positioned in and around the subsites that constitute the active site (i.e. either in the first or second shells).

3.6. Analysis of site-specific mutations. The analysis of the conservation of amino acids V173, M272, Q316, N344, I345 and L352 in a GH51 family subset composed of 488 sequences of bacterial origin (Fig. S5) reveals that none of these residues are totally conserved (36, 28, 51, 55, 50 and 52% identity respectively). However, generally (> 86%) these amino acids are substituted in other GH51 members by amino acids with related physicochemical properties. Therefore, in these cases, we mainly assumed that the observed mutations at these positions would not influence the target phenotype. However, the substitutions V173D, Q316L, N344I or Y and L352M do not form part of the natural sequence diversity observed in GH51 and were thus further analyzed. Additionally, although the substitution I345L occurs in 33% of the sequences, it was targeted for analysis in order to better dissociate the potential effect of this mutation from that of Q316L, both of these substitutions being present in mutant 169 (Table 3). On the other hand, although mutation M272L (a substitution that occurs in 11% of GH51 sequences; Fig. S5) is present in mutant 8, its impact on the overall properties was not studied.

Investigation of the catalytic properties of the point mutations V173D, Q316L, N344I/Y, L352M and I345L revealed that to some extent all of the mutants display improved transglycosylation (Table 4), with L352M standing out as a particularly interesting candidate. This latter synthesized the same product regioisomers as wild type *Tx*Abf, but procured higher (up to 3-fold) yields.

4. Discussion

The glycosynthetic activity of *Tx*Abf has been extensively characterized, especially using synthetic glycoside acceptors such as benzyl α-D-xyloside (22, 23, 32). However, although these studies have revealed the multiple facets of this activity and shown how *Tx*Abf can be used to prepare a variety of compounds, sometimes in high yields, they have also shown that *Tx*Abf only displays low synthetic activity in the presence of natural XOS acceptors.

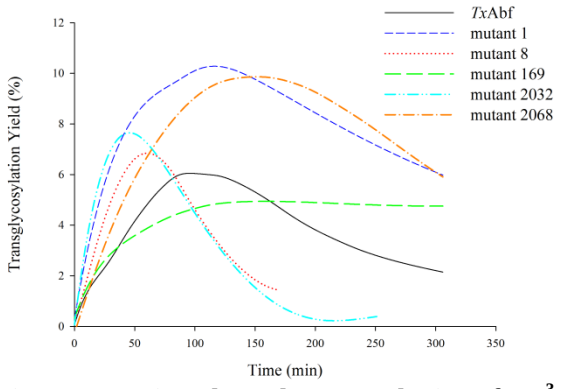


Figure 2 Time-dependant synthesis of A^3X catalyzed by wild-type *Tx*Abf and mutants thereof. Assays were conducted by ^1H NMR in D_2O at 40°C with *p*NP-α-L-Araf (5 mM) as donor, xylobiose (10 mM) as acceptor and enzyme (0.05 IU).

Table 3 Amino acid (AA) substitutions associated with each of the mutants.

Mutant identifier	Total number of AA substitutions	AA substitutions in catalytic domain[a]	AA substitutions in C-terminal domain
1	3	I103N, R158W, L352M	None
8	2	Y125H, M272L	None
169	5	V264D, Q316L, I345L	T17K, L399P
2032	5	I145T, V173D, N344I, T358S	I452S
2068	5	Q136R, N344Y, Q379L	D430V, L474H

[a] Underlined amino acid substitutions represent mutations located in the first and second shell of the active site (almost exclusively related to subsite -1).

This is an important observation, because the overarching aim of our work is to develop chemoenzymatic strategies for the preparation of AXOS, which could have a variety of applications. Nevertheless, the fact that *Tx*Abf actually possesses a highly under-developed, but measurable catalytic potential for the synthesis of AXOS provides a foothold for activity improvement using protein engineering, notably molecular evolution.

In this work, using random mutagenesis coupled to an appropriate HTS, we set out to identify original mutants that display enhanced ability to synthesize AXOS and thereby contribute to our fundamental of the molecular determinants of transglycosylation in *Tx*Abf. Underlying to this overarching goal are two different but related challenges, the first being the modulation of the transglycosylation/hydrolysis partition toward the glycosynthetic activity, while the second is the improvement of transglycosylation using natural XOS as acceptors. To simultaneously address these two challenges, we adopted a previously described screening strategy (33), which in a first step selects for hydrolytically-impaired variants and in the second step for variants whose ability to release indolyl moieties from X-α-L-Ara*f* is improved in the presence of XOS. However, because the screening strategy is based on the increase of catalytic turnover in the presence of acceptors, it also can select mutants that display reduced sensitivity to acceptor inhibition.

Gratifyingly the mutagenesis and screening protocol implemented in this study provided a large collection of potentially interesting mutants that display high R values relative to the mean R value and the standard deviation. Even more significantly, a majority of these were hydrolytically-impaired and appeared to be rescued by XOS, suggesting that XOS-mediated inhibition was reduced and/or that transglycosylation onto XOS acceptors was increased. Overall, these data validate the screening methodology, even if transglycosylation events are not directly monitored. The ^{1}H NMR-monitored transglycosylation reactions, confirmed by mass spectrometry analyses, indicate that several of the mutants displayed improved synthesis yields, although this contrasts with previous results obtained for *Tt*βgly mutants (33, 40), which revealed much higher acceptor-mediated activation. However, our results also indicate that acceptor-mediated activation cannot be exclusively correlated with improved glycosynthetic activity. Indeed, although the release of *p*NP from *p*NP-α-L-Ara*f* by the mutants 449, 639 and 1013 was slightly activated in the presence of xylobiose , these mutants displayed neither better transglycosylation activity, nor significantly increased synthesis yields, suggesting that only inhibition was alleviated. Most importantly, this study

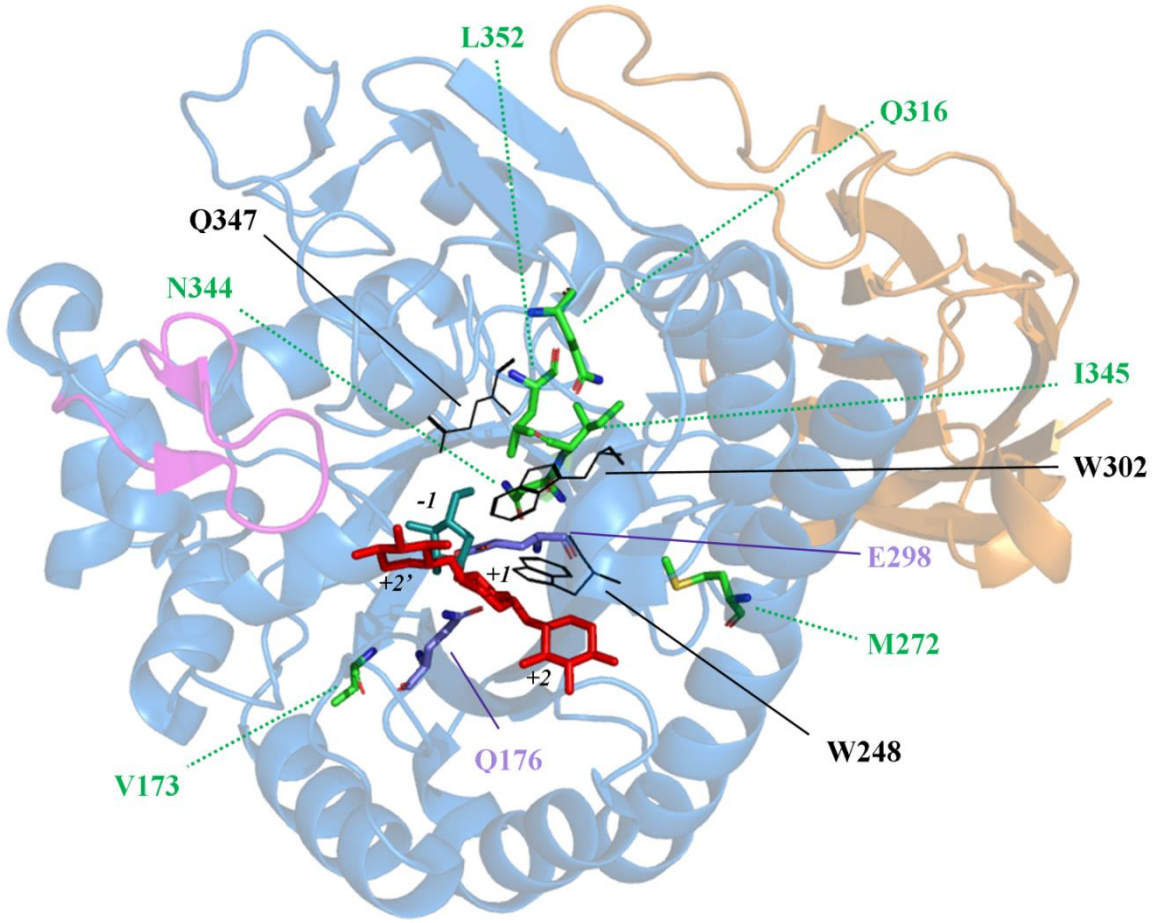


Figure 3 Cartoon representation of *Tx*Abf structure (PDB ID: 2VRQ) (31) showing the mobile β2α2 loop (pink), the $(\beta/\alpha)_8$ barrel (blue) and the C-terminal domain (orange). Mutations located in the first and second shells present in selected mutants are represented as sticks (green). Active site residues that are potentially affected by mutations are shown as lines (black), whereas the acid/base E176 (mutated into Q) and nucleophile E298 catalytic residues are shown in stick format (purple).

achieved its principal aim, which was the identification of several mutations that procured improved abilities (up to 2-fold higher than the wild-type enzyme) to synthesize AXOS. In the case of mutants 169 and 1013 this was achieved to some extent by a severe reduction in secondary hydrolysis, which preserves the transglycosylation products by reaching a stable plateau-phase for the endpoint product yields. A similar effect has already been attributed to point mutations at N344, a residue that modulates the nucleophilicity of E298 and thus affects the transglycosylation/hydrolysis partition (36). Accounting for the highly similar properties of mutant 169 and the single mutant Q316L, it appears reasonable to suggest that this latter substitution is solely responsible for the catalytic behavior that is ascribed to mutant 169. In this respect, it is interesting to note that although Q316 is located between subsites -1 and +1, this effect might be related to long-range interactions (the Cε (Q316)-Cα (E298) distance being 13.1 Å). Regarding mutant 1013, lack of hard evidence prevents us from proposing such an affirmative attribution of roles. However, the drastic spatial changes that are likely to be caused by the substitution T243P, which lies near to the key residue Y242 (conserved

Table 4 Transglycosylation yields calculated from time-course NMR kinetics of second shell mono-mutants (0.05 IU, each) in deuterated sodium acetate buffer (pD 5.8, 25 mM) at 45°C with *p*NP-α-L-Ara*f* (5 mM) as donor, xylotriose (10 mM) as acceptor.[a]

		Transglycosylation product Yield (%)			Maximum
		XA^3X	A^3XX	A^2XX	overall yield
	δ (ppm)	5.40	5.32	5.28	
Enzyme	wt	3	4	6	13
	V173D	4	6	6	16
	Q316L	6	8	9	23
	N344I	6	8	8	22
	I345L	4	6	6	16
	L352M	9	9	11	29

[a] The maximum overall yield represents the sum of each product maximum. Notably, maximum yield is not reached exactly at the same time for each mutant.

throughout GH-A) (47), designates T243P as a possible determinant of the properties of mutant 1013.

In the case of mutants (1, 8, 169, 2032 and 2068) that actually increased maximum instantaneous transglycosylation yields, albeit moderately, it seems likely that their properties are determined by mutations that mainly alter interactions between the enzyme and the donor substrate through a significant decrease of hydrolysis. In this respect it is significant that many of the point mutations that characterize these mutants are located in the first and second shell of subsite -1. Other studies that have focused on the improvement of transglycosylation in glycosidases have revealed that alterations in the donor subsite -1 might favor this property (40, 48–50).

In summary, the data presented here suggest that transglycosylation in *Tx*Abf, which is an *exo*-type GH that is characterized by an extensive hydrogen bonding network in the subsite -1 (31), might be favored by the alteration of these interactions. Presumably, modifications of this nature reduce water-mediated deglycosylation and/or slightly favor sugar acceptor-mediated deglycosylation. However, it is also clear crippling hydrolysis alone is insufficient to create strong transglycosylation activity, because it is also vital to increase acceptor binding. In this respect, the mutations pinpointed in this study are rather interesting starting points for directed evolution. This work is now underway to create a *bona fide* transarabinofuranosylases.

Acknowledgements

F. Arab-Jaziri and B. Bissaro received PhD fellowships from the French National Institute for Agricultural Research (INRA), though FAJ stipend was co-funded by the French Environment and Energy Management Agency (ADEME). The authors express gratitude to the staff members of MetaToul (Metabolomics & Fluxomics Facitilies, Toulouse, France, www.metatoul.fr) for technical support and access to NMR facility. MetaToul is part of the national infrastructure MetaboHUB (The French National infrastructure for metabolomics and fluxomics, www.metabohub.fr). MetaToul is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, SICOVAL, IBiSa-France, CNRS and INRA. Some NMR experiments were also performed at the PICT facility (IPBS, CNRS - Université de Toulouse) using equipment funded by the French Research Ministry, CNRS, Université Paul Sabatier, the Région Midi-Pyrénées and European structural funds. We thank Nic Lindley (LISBP) for

critical comments during the preparation of this manuscript.

REFERENCES

1. **Muthana S, Cao H, Chen X**. 2009. Recent progress in chemical and chemoenzymatic synthesis of carbohydrates. Curr. Opin. Chem. Biol. **13**:573–581.
2. **Kittl R, Withers SG**. 2010. New approaches to enzymatic glycoside synthesis through directed evolution. Carbohydr. Res. **345**:1272–1279.
3. **Schmaltz RM, Hanson SR, Wong C-H**. 2011. Enzymes in the synthesis of glycoconjugates. Chem. Rev. **111**:4259–4307.
4. **Jiang Z, Zhu Y, Li L, Yu X, Kusakabe I, Kitaoka M, Hayashi K**. 2004. Transglycosylation reaction of xylanase B from the hyperthermophilic *Thermotoga maritima* with the ability of synthesis of tertiary alkyl β-D-xylobiosides and xylosides. J. Biotech. **114**:125–134.
5. **Oshima H, Kimura I, Izumori K**. 2006. Synthesis and structure analysis of novel disaccharides containing D-psicose produced by endo-1,4-β-D-xylanase from *Aspergillus sojae*. J. Biosci. Bioeng. **101**:280–283.
6. **Kim Y-W, Fox DT, Hekmat O, Kantner T, McIntosh LP, Warren RAJ, Withers SG**. 2006. Glycosynthase-based synthesis of xylo-oligosaccharides using an engineered retaining xylanase from *Cellulomonas fimi*. Org. Biomol. Chem. **4**:2025–2032.
7. **Kadi N, Crouzet J**. 2008. Transglycosylation reaction of endoxylanase from *Trichoderma longibrachiatum*. Food Chem. **106**:466–474.
8. **Watanabe S, Viet DN, Kaneko J, Kamio Y, Yoshida S**. 2008. Cloning, expression, and transglycosylation reaction of *Paenibacillus* sp. strain W-61 xylanase 1. Biosci. Biotechnol. Biochem. **72**:951–958.
9. **Kim DY, Ham S-J, Kim HJ, Kim J, Lee M-H, Cho H-Y, Shin D-H, Rhee YH, Son K-H, Park H-Y**. 2012. Novel modular endo-*β*-1,4-xylanase with transglycosylation activity from *Cellulosimicrobium* sp. strain HY-13 that is homologous to inverting GH family 6 enzymes. Bioresour. Technol. **107**:25–32.
10. **Houseknecht JB, Lowary TL**. 2001. Chemistry and biology of arabinofuranosyl- and galactofuranosyl-containing polysaccharides. Curr. Opin. Chem. Biol. **5**:677–682.
11. **Peltier P, Euzen R, Daniellou R, Nugier-chauvin C, Ferrières V**. 2008. Recent knowledge and innovations related to hexofuranosides: structure, synthesis and applications. Carbohydr. Res. **343**:1897–1923.
12. **Hirsch J, Petrakova E, Schraml J**. 1984. Stereoselective synthesis and ^{13}C-N.M.R. spectra of two isomeric methyl *β*-glycosides of trisaccharides related to arabinoxylan. Carbohydr. Res. **131**:219–226.
13. **Koto S, Morishima N, Takenaka K, Uchida C, Zen S**. 1985. Pentoside synthesis by dehydrative glycosylation. Synthesis of *O*-*α*-L-arabinofuranosyl-(1→3)-*O*-*β*-D-xylopyranosyl-(1→4)-D-xylopyranose. Bull. Chem. Soc. Jpn. **58**:1464–1468.
14. **Helm RF, Ralphlb J, Anderson L**. 1991. Regioselective protection strategies for D-xylopyranosides. J. Org. Chem. **56**:7015–7021.
15. **Utille J-P, Jeacomine I**. 2007. Synthesis of a library of allyl α-L-arabinofuranosyl-α- or β-D-xylopyranosides; route to higher oligomers. Carbohydr. Res. **342**:2649–2656.
16. **Bang MH, Riep T Van, Thinh NT, Song LH, Dung TT, Truong L Van, Don L VAn, Ky TD, Pan D, Shaheen M, Ghoneum M**. 2010. Arabinoxylan rice bran (MGN-3) enhances the effects of interventional therapies for the treatment of hepatocellular carcinoma: a three-year randomized clinical trial. Anticancer Res. **30**:5145–5152.
17. **Grootaert C, Van den Abbeele P, Marzorati M, Broekaert WF, Courtin CM, Delcour JA, Verstraete W, Van de Wiele T**. 2009. Comparison of prebiotic effects of arabinoxylan oligosaccharides and inulin in a simulator of the human intestinal microbial ecosystem. FEMS Microbiol. Ecol. **69**:231–242.
18. **Broekaert WF, Courtin CM, Verbeke K, Van de Wiele T, Verstraete W, Delcour JA**. 2011. Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides. Crit. Rev. Food Sci. Nutr. **51**:178–194.
19. **Phan The D, Debeaufort F, Voilley A, Luu D**. 2009. Biopolymer interactions affect the functional properties of edible films based on agar, cassava starch and arabinoxylan blends. J. Food Eng. **90**:548–558.
20. **Boukari I, Putaux J-L, Cathala B, Barakat A, Saake B, Rémond C, O'Donohue M, Chabbert B**. 2009. In vitro model assemblies to study the impact of lignin-carbohydrate interactions on the enzymatic conversion of xylan. Biomacromolecules **10**:2489–2498.
21. **Rémond C, Ferchichi M, Aubry N, Plantier-Royon R, Portella C, O'Donohue MJ**. 2002. Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α-L-

arabinofuranosidase. Tetrahedron Lett. **43**:9653–9655.

22. **Rémond C, Plantier-Royon R, Aubry N, Maes E, Bliard C, O'Donohue MJ**. 2004. Synthesis of pentose-containing disaccharides using a thermostable α-L-arabinofuranosidase. Carbohydr. Res. **339**:2019–25.
23. **Rémond C, Plantier-Royon R, Aubry N, O'Donohue MJ**. 2005. An original chemoenzymatic route for the synthesis of β-D-galactofuranosides using an α-L-arabinofuranosidase. Carbohydr. Res. **340**:637–644.
24. **Euzen R, Lopez G, Nugier-Chauvin C, Ferrières V, Plusquellec D, Rémond C, O'Donohue M**. 2005. A chemoenzymatic approach for the synthesis of unnatural disaccharides containing D-galacto- or D-fucofuranosides. Eur. J. Org. Chem. 4860–4869.
25. **Lopez G, Nugier-Chauvin C, Rémond C, O'Donohue M**. 2007. Investigation of the specificity of an α-L-arabinofuranosidase using C-2 and C-5 modified α-L-arabinofuranosides. Carbohydr. Res. **342**:2202–2211.
26. **Sakamoto T, Fujita T, Kawasaki H**. 2004. Transglycosylation catalyzed by a *Penicillium chrysogenum* exo-1,5-α-L-arabinanase. Biochim. Biophys. Acta **1674**:85–90.
27. **Wan C-F, Chen W-H, Chen C-T, Chang MD-T, Lo L-C, Li Y-K**. 2007. Mutagenesis and mechanistic study of a glycoside hydrolase family 54 α-L-arabinofuranosidase from *Trichoderma koningii*. Biochem. J. **401**:551–558.
28. **Chlubnová I, Filipp D, Spiwok V, Dvoráková H, Daniellou R, Nugier-Chauvin C, Králová B, Ferrières V**. 2010. Enzymatic synthesis of oligo-D-galactofuranosides and L-arabinofuranosides: from molecular dynamics to immunological assays. Org. Biomol. Chem. **8**:2092–2102.
29. **Chlubnová I, Králová B, Dvořáková H, Hošek P, Spiwok V, Filipp D, Nugier-Chauvin C, Daniellou R, Ferrières V**. 2014. The versatile enzyme Ara*f*51 allowed efficient synthesis of rare pathogen-related β-D-galactofuranosyl-pyranoside disaccharides. Org. Biomol. Chem. **12**:3080–3089.
30. **Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B**. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. **42**:D490–D495.
31. **Paës G, Skov LK, O'Donohue MJ, Rémond C, Kastrup JS, Gajhede M, Mirza O**. 2008. The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. Biochemistry **47**:7441–7451.
32. **Arab-Jaziri F, Bissaro B, Dion M, Saurel O, Harrison D, Ferreira F, Milon A, Tellier C, Fauré R, O'Donohue MJ**. 2013. Engineering transglycosidase activity into a GH51 α-L-arabinofuranosidase. New Biotechnol. **30**:536–544.
33. **Koné FMT, Le Béchec M, Sine J-P, Dion M, Tellier C**. 2009. Digital screening methodology for the directed evolution of transglycosidases. Protein Eng., Des. Sel. **22**:37–44.
34. **Marmuse L, Asther M, Fabre E, Navarro D, Lesage-Meessen L, Asther M, O'Donohue M, Fort S, Driguez H**. 2008. New chromogenic substrates for feruloyl esterases. Org. Biomol. Chem. **6**:1208–1214.
35. **Vrsanska M, Nerinckx W, Claeyssens M, Biely P**. 2008. An alternative approach for the synthesis of fluorogenic substrates of *endo*-β-(1→4)-xylanases and some applications. Carbohydr. Res. **343**:541–548.
36. **Bissaro B, Saurel O, Arab-Jaziri F, Saulnier L, Milon A, Tenkanen M, Monsan P, O'Donohue MJ, Fauré R**. 2014. Mutation of a pH-modulating residue in a GH51 α-L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products. BBA - Gen. Subjects. **1840**:626–636.
37. **Cadwell RC, Joyce GF**. 1992. Randomization of genes by PCR mutagenesis. Genome Res. **2**:28–33.
38. **Arab-Jaziri F, Bissaro B, Barbe S, Saurel O, Débat H, Dumon C, Gervais V, Milon A, André I, Fauré R, O'Donohue MJ**. 2012. Functional roles of H98 and W99 and β2α2 loop dynamics in the α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. FEBS J. **279**:3598–3611.
39. **Gottlieb HE, Kotlyar V, Nudelman A**. 1997. NMR chemical shifts of common laboratory solvents as trace impurities. J. Org. Chem. **62**:7512–7515.
40. **Feng H-Y, Drone J, Hoffmann L, Tran V, Tellier C, Rabiller C, Dion M**. 2005. Converting a β-glycosidase into a β-transglycosidase by directed evolution. J. Biol. Chem. **280**:37088–37097.
41. **Osanjo G, Dion M, Drone J, Solleux C, Tran V, Rabiller C, Tellier C**. 2007. Directed evolution of the α-L-fucosidase from *Thermotoga maritima* into an α-L-transfucosidase. Biochemistry **46**:1022–1033.
42. **Fauré R, Courtin CM, Delcour JA, Dumon C, Faulds CB, Fincher GB, Fort S, Fry SC,**

Halila S, Kabel MA, Pouvreau L, Quemener B, Rivet A, Saulnier L, Schols HA, Driguez H, O'Donohue MJ. 2009. A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls. Aust. J. Chem. **62**:533–537.

43. **Pastell H, Tuomainen P, Virkki L, Tenkanen M**. 2008. Step-wise enzymatic preparation and structural characterization of singly and doubly substituted arabinoxylo-oligosaccharides with non-reducing end terminal branches. Carbohydr. Res. **343**:3049–3057.
44. **Correia MAS, Mazumder K, Brás JLA, Firbank SJ, Zhu Y, Lewis RJ, York WS, Fontes CMGA, Gilbert HJ**. 2011. Structure and function of an arabinoxylan-specific xylanase. J. Biol. Chem. **286**:22510–22520.
45. **Anders N, Wilkinson MD, Lovegrove A, Freeman J, Tryfona T, Pellny TK, Weimar T, Mortimer JC, Stott K, Baker JM, Defoin-Platel M, Shewry PR, Dupree P, Mitchell RAC**. 2012. Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. Proc. Natl. Acad. Sci. U.S.A. **109**:989–993.
46. **McKee LS, Peña MJ, Rogowski A, Jackson A, Lewis RJ, York WS, Krogh KBRM, Viksø-Nielsen A, Skjøt M, Gilbert HJ, Marles-Wright J**. 2012. Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. Proc. Natl. Acad. Sci. U.S.A. **109**:6537–6542.
47. **Zhang Y, Ju J, Peng H, Gao F, Zhou C, Zeng Y, Xue Y, Li Y, Henrissat B, Gao GF, Ma Y**. 2008. Biochemical and structural characterization of the intracellular mannanase AaManA of *Alicyclobacillus acidocaldarius* reveals a novel glycoside hydrolase family belonging to clan GH-A. J. Biol. Chem. **283**:31551–31558.
48. **Aronson Jr NN, Halloran BA, Alexeyev MF, Zhou XE, Wang Y, Meehan EJ, Chen L**. 2006. Mutation of a conserved tryptophan in the chitin-binding cleft of *Serratia marcescens* chitinase A enhances transglycosylation. Biosci., Biotechnol., Biochem. **70**:243–251.
49. **Placier G, Watzlawick H, Rabiller C, Mattes R**. 2009. Evolved β-galactosidases from *Geobacillus stearothermophilus* with improved transgalactosylation yield for galacto-oligosaccharide production. Appl. Environ. Microbiol. **75**:6312–6321.
50. **Teze D, Hendrickx J, Czjzek M, Ropartz D, Sanejouand Y-H, Tran V, Tellier C, Dion M**. 2014. Semi-rational approach for converting a GH1 β-glycosidase into a β-transglycosidase. Protein Eng., Des. Sel. **27**:13–19.

SUPPORTING INFORMATION of the Article: "Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α-L-arabinofuranosidase"

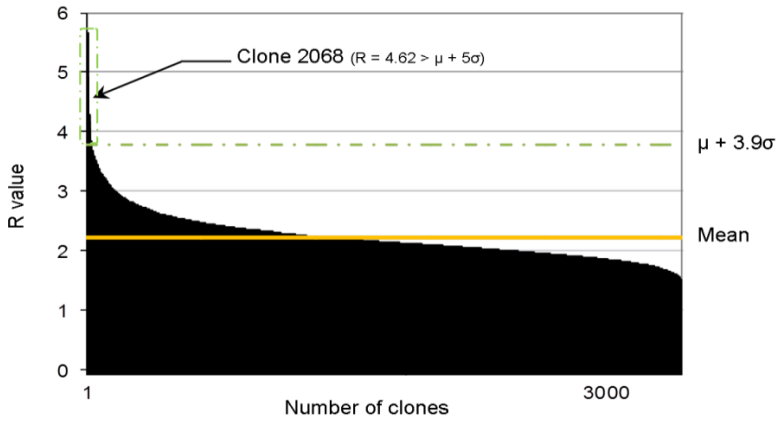


Figure S1. Digital image screening of a library of mutant *Tx*Abf in the presence of XOS mixture as acceptors. Activities are expressed as R= $(I_f - I_0)/I_0$, where I_0 and I_f are the colony intensities registered after the first (i.e. incubation on X-α-L-Ara*f*, 1 h at 30°C) and second steps (incubation on X-α-L-Ara*f* in the presence of XOS, 4 h at 37°C) in the procedure respectively. Ratio mean value of the Q-tray (μ) = 2.22 (yellow line) and standard deviation (σ) = 0.40; fifteen clones with a R value ≥ 3.78 (μ + 3.9σ) were selected (green dash dot line), including the one expressing mutant 2068 (4.62).

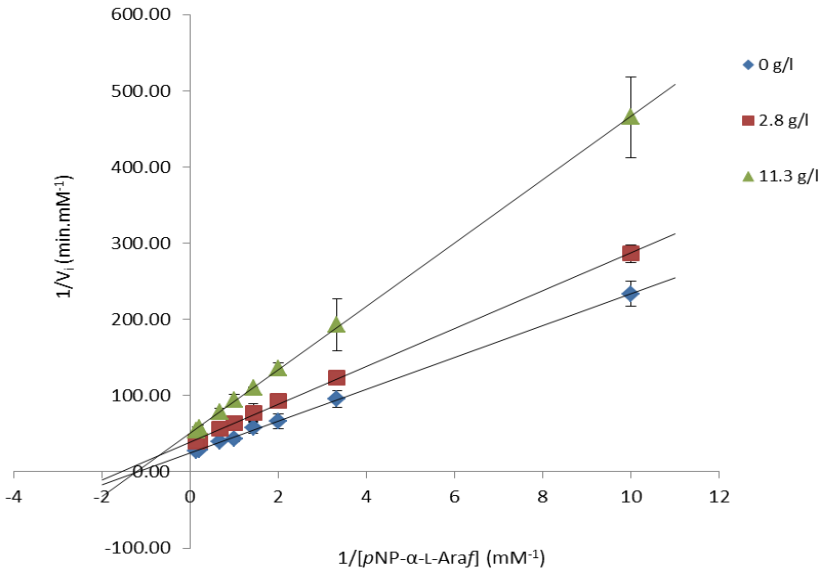


Figure S2. Kinetics of wild-type *Tx*Abf-catalyzed consumption of various amounts of *p*NP-α-L-Ara*f* (0.1 to 7 mM) in the presence of different concentrations of XOS mixture. Results were analysed using a double reciprocal plot. Reactions were performed in 50 mM sodium acetate buffer pH 5.8 at 60 °C.

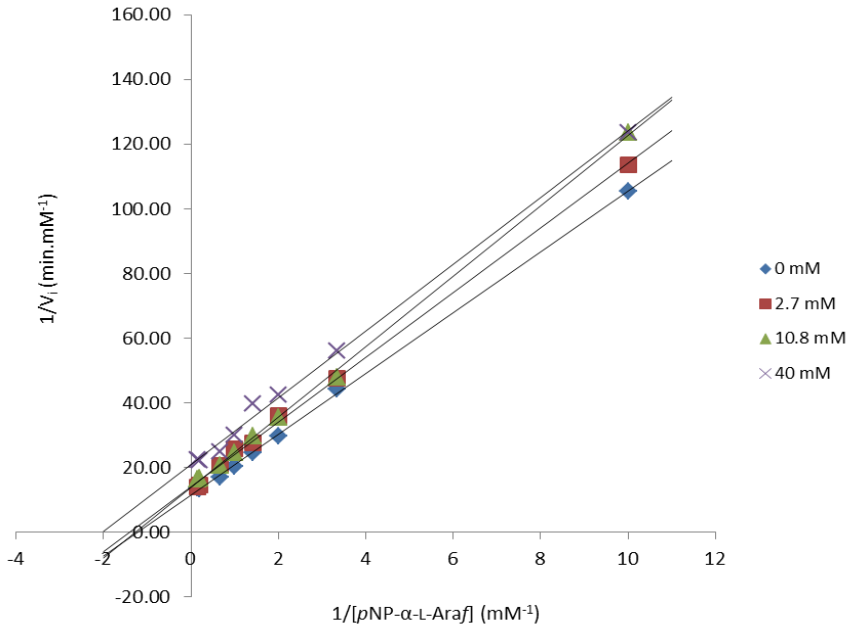


Figure S3. Kinetics of wild-type *Tx*Abf-catalyzed consumption of various amounts of *p*NP-α-L-Ara*f* (0.1 to 7 mM) in the presence of different concentrations of xylobiose. Results were analysed using a double reciprocal plot. The data suggest uncompetitive inhibition, but accurate determination of the $K_{i\,U}$ value (approximately 56 mM) was rendered difficult taking into account the need to use very high xylobiose concentrations. Reactions were performed in 50 mM sodium acetate buffer pH 5.8 at 60 °C.

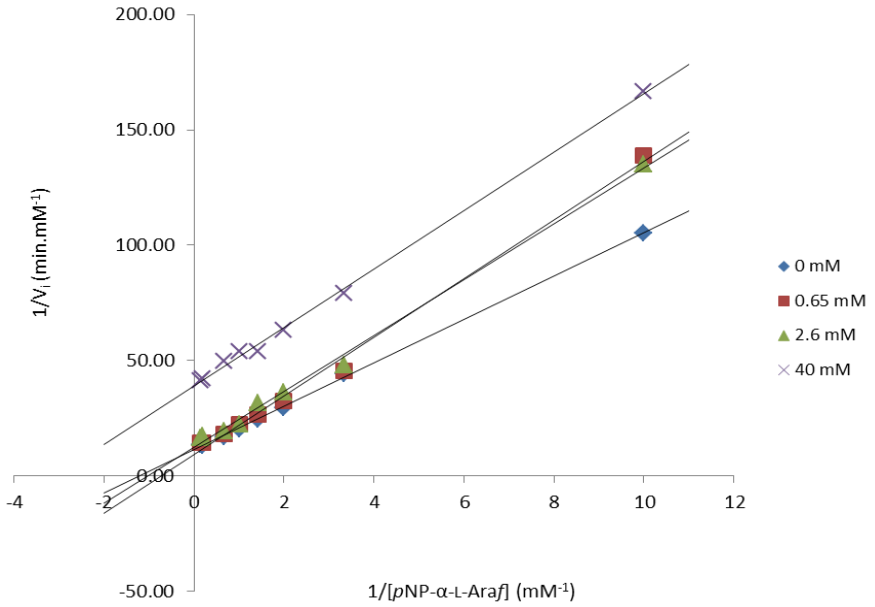
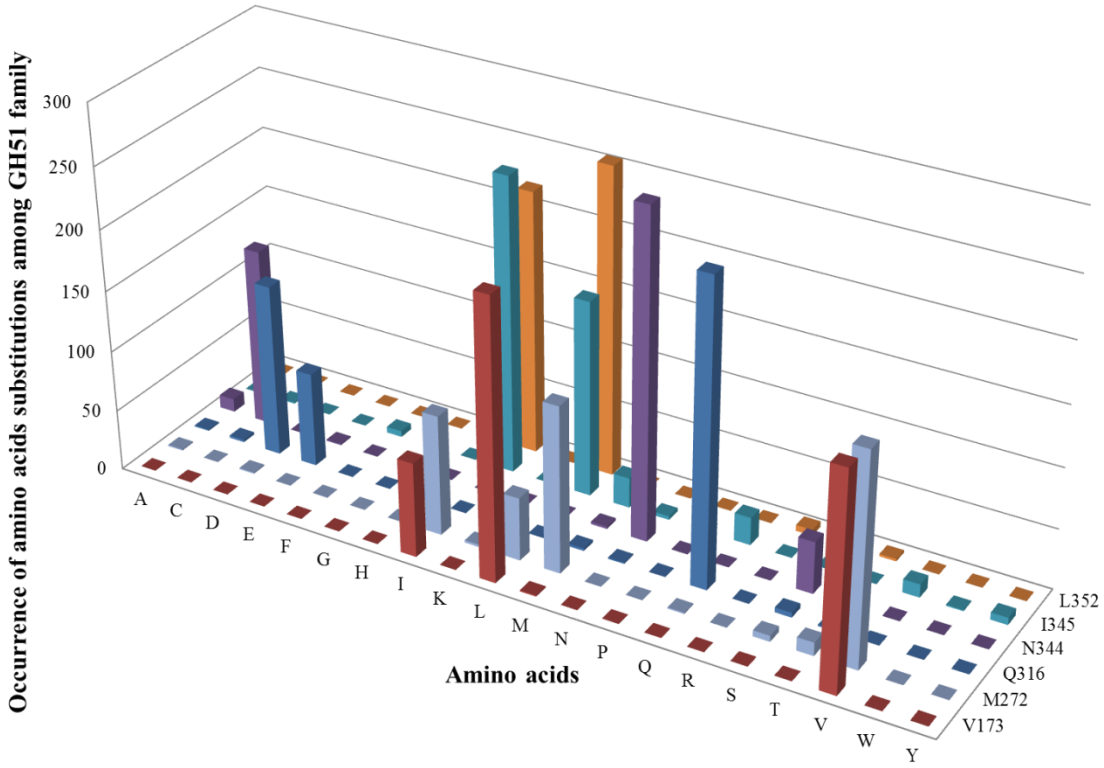


Figure S4. Kinetics of wild-type *Tx*Abf-catalyzed consumption of various amounts of *p*NP-α-L-Ara*f* (0.1 to 7 mM) in the presence of different concentrations of xylotriose. Results were analysed using a double reciprocal plot. The data suggest uncompetitive inhibition, but accurate determination of the $K_{i\,U}$ value (approximately 14 mM) was rendered difficult taking into account the need to use high xylotriose concentrations. Reactions were performed in 50 mM sodium acetate buffer pH 5.8 at 60 °C.



	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
V173	0	0	0	0	0	0	0	78	0	230	1	0	0	0	0	0	0	178	0	0
M272	1	0	1	0	0	0	0	99	2	52	137	0	0	1	0	5	11	174	0	0
Q316	1	2	142	78	0	0	1	0	0	0	2	1	0	250	0	4	1	0	0	0
N344	11	147	0	1	0	0	0	0	0	0	2	270	0	0	0	43	8	0	0	0
I345	0	1	0	0	5	0	0	245	0	162	25	3	0	23	0	0	0	11	1	6
L352	0	0	0	0	0	0	1	218	0	255	0	0	0	0	5	1	2	0	0	0

Figure S5. Amino acid substitutions at positions identified through random mutagenesis. For each position identified through random mutagenesis, the occurrence of alternative amino acid was determined by examining 488 GH51 sequences of bacterial origin.