



HAL
open science

Conception “ à façon ” d’enzymes pour la synthèse chimio-enzymatique programmée d’oligosaccharides antigéniques

Mounir Benkoulouche

► **To cite this version:**

Mounir Benkoulouche. Conception “ à façon ” d’enzymes pour la synthèse chimio-enzymatique programmée d’oligosaccharides antigéniques. Microbiologie et Parasitologie. INSA de Toulouse, 2019. Français. NNT : 2019ISAT0050 . tel-03653082

HAL Id: tel-03653082

<https://theses.hal.science/tel-03653082>

Submitted on 27 Apr 2022

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

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



THÈSE

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de
Toulouse

Présentée et soutenue par
Mounir BENKOULOUCHE

Le 27 septembre 2019

Conception « à façon » d'enzymes pour la synthèse chimio-enzymatique programmée d'oligosaccharides antigéniques

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries**

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :

TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering

Thèse dirigée par
Isabelle ANDRE

Jury

M. Sébastien FORT, Rapporteur

M. Antoni PLANAS, Rapporteur

M. Uwe BORNSCHEUER, Examineur

Mme Magali REMAUD-SIMÉON, Examinatrice

Mme Claire MOULIS, Examinatrice

Mme Laurence MULARD, Examinatrice

Mme Isabelle ANDRE, Directrice de thèse

*Je dédie cette thèse à mes parents,
qui ont nourri ma curiosité et m'ont amené sur les chemins de la Science.*

REMERCIEMENTS

Je tiens tout d'abord à remercier ma directrice de thèse Isabelle André, et la leader du pôle CIMES Magali Remaud-Siméon qui m'ont donné l'opportunité de réaliser cette thèse au sein du TBI/LISBP en m'accueillant dans les sous-groupes MMD et Enzynov. A l'interface entre ces deux groupes, j'ai réussi à apprendre plus que je ne l'aurais jamais imaginé et à m'épanouir dans un environnement propice à l'innovation, à la prise de risques, et à la liberté d'expérimenter ses idées et c'est exactement ce dont j'avais besoin pour développer mon goût pour la recherche. Merci à toi Isabelle en particulier, tu m'as rendu plus fort et plus résistant, et grâce à ton mentorat, je serai prêt à affronter l'adversité et la frustration présents dans tout projet de recherche. Et merci à toi Magali pour toutes ces discussions scientifiques passionnantes. Bien sûr je tiens également à remercier ma co-encadrante Claire Moulis. Merci pour ton support indéfectible au quotidien, ta gentillesse et ta compétence qui ont aussi fait de cette thèse ce qu'elle est aujourd'hui.

Je tiens ensuite à remercier Akli Ben Imeddourene, mon très cher binôme qui a évidemment sa place en haut de l'affiche tant ces travaux sont indissociables de ton dévouement et de ton acharnement à produire des données de qualité. Tu n'as pas été simplement un collaborateur mais également un professeur pour moi, grâce à ta patience pour m'expliquer tous les concepts de modélisation et de RMN qui m'étaient si peu familiers, et ta disponibilité systématique pour automatiser mes analyses. Merci également à Dorian, Valentin et Elodie, qui ont participé directement à ce travail au travers de leur stage. C'était un plaisir de vous encadrer.

Merci ensuite à Yannick, car tu étais présent lors de mon entretien et c'est avec toi que j'allais travailler au début. Merci d'avoir cru en moi, c'est aussi grâce à toi que je suis là. Et je suis ravi d'avoir pu initier le développement de cette nouvelle souche avec toi. Merci pour toutes nos longues discussions sur la bioéconomie et nos partages d'idées.

Merci, ensuite, à tous nos collaborateurs qui ont permis à cette thèse de voir le jour. A l'Institut Pasteur, Laurence Mulard pour avoir mené si loin cette collaboration et pour ta volonté de fer à éradiquer le fléau *Shigella*, à l'origine de tellement de projets intéressants. Si la synergie entre approches chimiques et enzymatiques n'est plus à prouver, il en va de même de celle entre Pasteur et TBI. Merci à Louis-Antoine (et avant toi, à Guillaume et Zhaoyu) pour votre ténacité à synthétiser cette molécule dont j'aurais apprécié chaque milligramme. Merci ensuite à Gianluca pour nos (trop rares) moments à travailler ensemble, j'espère que tu pourras porter encore plus loin l'histoire de BRS-B ! Merci également à nos collaborateurs de toutes les plateformes, qui ont été indispensables pour les nombreuses méthodes analytiques développées durant ces travaux mais aussi pour le support technique. De la plateforme MetaToul, Floriant et Hanna, de la plateforme ICEO, Nelly, Sandra et Sophie ; et puis bien sûr Etienne, toujours présent pour aider et proposer une solution à mes (nombreux) problèmes. Merci enfin à David G et David R pour cette fructueuse collaboration avec la plateforme BIA-

BIBS qui j'en suis persuadé pourra apporter des analyses d'une grande qualité sur tous les projets 'sucre' de l'équipe.

Merci à Gaby de m'avoir donné une opportunité pour la suite de la thèse, j'ai hâte de poursuivre le travail sur les enzymes à sucre, et bien sûr merci à mon futur complice en crime, Julien. J'ai déjà eu les prémisses lors de nos galères à purifier les sucres, et je sais que ça va être juste génial de travailler avec toi ! Tu es un sacré modèle de persévérance et je t'admire pour ton parcours,

Merci ensuite bien sûr à tous mes amis de Toulouse, pour ces soirées, ces rires, ces échanges. Si la thèse m'a fait grandir scientifiquement, vous m'avez fait grandir humainement.

Merci donc à « la Ligue », à toi Maher pour ton soutien, tu m'as pris sous ton aile et tu m'as donné goût au sport et au dépassement, merci pour tous tes conseils avisés de « grand frère ». Merci à Yassim, surtout garde toujours ton optimisme et ta bonne humeur qui m'ont tellement de fois redonné le sourire. Au bordelais frère d'armes, Arthur, avec ton cœur sur la main. A Moli, mon modèle de thésard à l'époque et maintenant de chercheur, pour ton soutien et les fous-rires de « Blonde time », vivement les prochains jeux de société ! Merci également à toi Cathy, c'est toujours un plaisir de partager nos galères et d'en rire.

Merci ensuite au « Gang ». Carmo, on est arrivé au labo en même temps et c'est grâce à toi que j'ai rencontré mes premiers amis à Toulouse, aujourd'hui si proches. N'arrête jamais les carmonites, car ta bonne humeur est contagieuse et fait oublier tous les soucis ! Merci pour tout, depuis 4 ans, et j'espère pour encore longtemps. Merci donc aussi à Damien, Julien, Paul pour nos sessions JdR endiablés, et à Marina en addition lors des soirées pubs/burger ou Baraka. Merci enfin à toi Laure, également toujours d'un grand soutien dans les moments difficiles, je suis heureux que tu sois de retour à Toulouse après avoir baroudé loin de la civilisation.

Merci à tous les collègues et amis du laboratoire, passés et présents, qui m'auront permis de vivre cette aventure avec un vrai plaisir d'aller travailler chaque jour. Merci donc à Alex G, Alex T, Alvaro, Amandine, Angeline, Audrey, Awilda, Barbara, Benoît, Betty, Camille, Clément F, Clément V, Coraline, David S, Eli, Emma, Emna, Florent, Franck, Gilles, Jelena, Jérémy, Jiao, Julian, Laurence T, Louise, Maéva G, Marie, Marion, Marlène, Maxant, Neil, Pablo, Pauline, Sabine, Sayani, Tarun, Thomas, Victor, Vincent, Vinciane, Zhi, Zhongpeng, et à toute l'équipe CIMEs ! Enfin un merci tout particulier à toi, Laurence F. Tu es arrivée alors que j'étais en pleine rédaction, et ton soutien depuis lors compte beaucoup pour moi, tout comme tu pourras compter sur le mien pendant ta (future) thèse. Merci pour tout.

Merci aux copains de l'école, Lucie, Antoine, Jojo, Maeva, Sacha et Marine, pour nos soirées raclettes, nos sessions escalades ou canoë, ou encore ces vacances tous ensemble à Milan ! Vous avez été si importants à mon équilibre !

Aux amis loin, mais finalement pas tant que ça : Thomas, pour ton accueil toujours chaleureux à Bruxelles, ma thèse finit, la tienne commence, et tu es clairement fait pour ça, alors je te souhaite simplement bon courage. Merci aussi à Co, Titi, Pâté, Bassem... Merci pour ces bons moments, ce super TomorrowLand et ce voyage au Japon génial ! Merci ensuite évidemment aux « bros », Alex, Florent, Romain, Floriane, Diane, Marion, Fanny, chacune de nos réunions est juste un retour en arrière, comme si on ne s'était jamais quitté. Et si la vie avance et chacun se construit, je suis heureux de vous compter parmi mes amis, encore pour très longtemps. Merci à mes expats préférés, Flo et Délia, nos sessions Skype sont un grand bol d'air frais à chaque fois.

A tous ceux que j'ai côtoyés pendant ma thèse et qui sont tous une pierre construisant cette aventure pendant 4 ans et donc moi-même par la même occasion : les amis des doctoriales (Eva, Nico, Cyril, Julie, Paul) mais aussi du théâtre, du crossfit...

Et enfin, merci à ma famille. Papa, Maman, merci pour votre soutien permanent, pour votre accueil à la maison à chaque fois, pour m'avoir donné un refuge et un lieu de coupure dans les moments difficiles, et pour avoir pris sur vous pour me faciliter la vie au maximum dans ces moments-là. Merci à toi Nassim, mon cher frère, pour nos longues conversations au téléphone et d'être toujours présent, en toutes circonstances, et à toi Pauline car tu t'es toujours tenue au courant de mes avancées.

SUMMARY

Tailored enzymes for programmed chemo-enzymatic synthesis of antigenic oligosaccharides

Combined with chemical synthesis, the use of **biocatalysts** holds great potential to open the way to molecular diversity. Nonetheless, the lack of appropriate enzymatic tools with requisite properties has hampered extensive exploration of **chemo-enzymatic** routes to complex carbohydrates. The research work presented in this thesis aimed at exploring **multidisciplinary** and **synergistic** strategies to develop “programmed” chemo-enzymatic pathways that take advantage of innate substrate promiscuity and computer-aided engineering of α -transglucosylases to produce glyco-bricks, easy-to-assemble into biologically active molecules mimicking the antigenic polysaccharide moiety – **O-antigen (O-Ag)** – of the *Shigella flexneri* **lipopolysaccharides**. These latter have been identified as **major targets** of protection against reinfection and could be used in the composition of vaccines to prevent shigellosis, an endemic form of **bacillary dysentery**. Many *S. flexneri* serotypes presented a tremendous O-Ag structural diversity, varying mainly in the regio-selectivity of the 1,2-*cis* α -D-glucosylation. We focused herein on the common structure of the ABCD tetrasaccharide defining the backbone repeating unit of most *S. flexneri* O-Ags. A lightly protected ABC'D' was first chemically synthesized by our collaborators from the Pasteur Institute. Next, we evaluated the potential of **branching sucrases** (BRS) from the Glycoside Hydrolase family 70 to glucosylate ABC'D'. They present the advantage of displaying a large **promiscuity** toward acceptor substrates.

A first achievement was reached using 6 native BRS enzymes whose ability to recognize and glucosylate ABC'D' was shown, leading to the production and the structural characterization of a **first glyco-brick**, ABC'[E(1→6)]D', representative of *S. flexneri* **serotypes 4a/4b**.

Then a collection of 22 pre-existing mutants of ΔN_{123} -GBD-CD2 was screened. Depending on the amino acid substitutions introduced in the active site, the glucosylation regio-selectivity was redirected compared to parental enzyme. Three products were identified, of which a **second glyco-brick**, [E(1→3)]ABC'D', is representative of prevalent *S. flexneri* **serotype 3a**.

To further enhance the accessible diversity, we then decided to purposely redesign the active site of another branching sucrose, BRS-B $\Delta 2$, using **computer-aided engineering methods**. We followed a computational framework combining molecular modelling and automated computational protein design techniques. A set of 49 sequences containing up to 15 mutations was then proposed and the 49 mutants were recombinantly produced in *E. coli*, then assayed for glucosylation of ABC'D'. Products were characterized using sensitive MS/MS techniques. Impressively these mutants produced as much as six different mono-glucosylated products out of the eight possible ones, of which one glyco-brick, A[E(1→3)]BC'D', is representative of *S. flexneri* **serotype 5a**.

Overall, our results demonstrate the versatility of branching sucrases and their mutants to access a broad range of **glucosylation patterns from a common tetrasaccharide acceptor** molecule, ABC'D'. It also opens new paths for improving catalytic performances of these mutants and better understanding molecular and dynamical determinants that could play a role in acceptor recognition and its site-selective glucosylation that will hopefully offer new opportunities for the synthesis of *S. flexneri* haptens.

Keywords: Transglycosylase, Glucosylation, Enzyme design and engineering, Chemo-enzymatic synthesis, Antigenic oligosaccharides, *Shigella flexneri*

Doctoral school: SEVAB (Ecological, Veterinary, Agronomic Sciences and Bioengineering).

Field: Enzymatic and Microbial Engineering

Laboratory: Toulouse Biotechnology Institute (UMR CNRS 5504, UMR INRA 792), INSA Toulouse

Conception à façon d'enzymes pour la synthèse chimio-enzymatique programmée d'oligosaccharides antigéniques

Combinée à la synthèse chimique, l'utilisation de **biocatalyseurs** offre l'opportunité d'accéder à une nouvelle diversité moléculaire. Néanmoins, le manque d'outils enzymatiques possédant les propriétés requises complique l'exploration de voies **chimio-enzymatiques** pour la synthèse de sucres complexes. Les travaux de recherche présentés dans cette thèse visent à explorer des stratégies **multidisciplinaires** et **synergiques** afin de développer des voies chimio-enzymatiques «programmées» qui tirent parti de la promiscuité de substrat innée et de l'ingénierie assistée par ordinateur des α -transglucosylases. Le but est ainsi de produire des glycobriques, facilement assemblables en molécules biologiquement actives et mimant le fragment polysaccharidique antigénique - **O-antigène (O-Ag)** - des **lipopolysaccharides** de *Shigella flexneri*. Ces derniers ont été identifiés comme étant des **cibles majeures** de protection contre la réinfection et pourraient être utilisés dans la composition de vaccins de prévention contre la shigellose, une forme endémique de **dysenterie bacillaire**. De nombreux sérotypes de *S. flexneri* ont révélé une grande diversité structurelle de leurs O-Ags, qui varient principalement de par la régio-sélectivité de la 1,2-*cis* α -D-glucosylation. Nous nous sommes concentrés ici sur la structure commune du tétrasaccharide ABCD, définissant l'unité de répétition du squelette de la plupart des O-Ags de *S. flexneri*. Une version protégée (ABC'D') a tout d'abord été synthétisée par nos collaborateurs de l'Institut Pasteur. Nous avons ensuite évalué le potentiel des sucrases de branchement (BRS) de la famille 70 des Glycoside Hydrolases pour la glucosylation de ABC'D'. Ces enzymes présentent une grande **promiscuité** vis-à-vis des substrats accepteurs. Un premier résultat majeur a été obtenu grâce à l'utilisation de 6 enzymes natives dont la capacité à reconnaître et à glucosyler ABC'D' a été démontrée, conduisant à la production et à la caractérisation structurale d'une **première glycobrique**, ABC'[E(1→6)]D', caractéristique des **sérotypes 4a / 4b** de *S. flexneri*.

Par la suite une collection de 22 mutants préexistants de ΔN_{123} -GBD-CD2 a été criblée. En fonction des substitutions d'acides aminés introduites dans le site actif, nous avons observé la redirection de la régio-sélectivité de la glucosylation par rapport à l'enzyme parentale. Trois produits ont été identifiés, dont une **seconde glycobrique**, [E(1→3)]ABC'D', représentative du **sérotipe prévalent 3a** de *S. flexneri*.

Pour étendre davantage la diversité accessible, nous avons ensuite entrepris de remodeler le site actif d'une autre BRS à l'aide de méthodes **de design d'enzymes assistée par ordinateur**. Nous avons mis en place une stratégie associant la modélisation moléculaire et des techniques automatisées de conception de protéines. Un ensemble de 49 séquences contenant jusqu'à 15 mutations a ensuite été proposé et les 49 mutants ont été produits par voie recombinante dans *E. coli*, puis testés pour la glucosylation de ABC'D'. Les produits ont été caractérisés à l'aide de techniques sensibles telle que la MS/MS. De manière impressionnante, ces mutants produisent jusqu'à six produits mono-glucosylés différents sur les huit possibles, dont une glycobrique, A[E(1→3)]BC'D', représentative du **sérotipe 5a** de *S. flexneri*. Au final, nos résultats démontrent la polyvalence des BRS et de leurs mutants pour accéder à une large gamme de **profils de glucosylation à partir d'une molécule accepteur commune**, le tétrasaccharide ABC'D'. Ces travaux ouvrent également de nouvelles voies pour améliorer les performances catalytiques de ces mutants et pour mieux comprendre les déterminants moléculaires et dynamiques qui pourraient jouer un rôle dans la reconnaissance des accepteurs et leur glucosylation sélective, ce qui pourra offrir de nouvelles possibilités pour la synthèse des haptènes de *S. flexneri*.

Mots-clés: Transglycosylase, Glucosylation, Design et ingénierie d'enzymes, Synthèse chimio-enzymatique, Oligosaccharides antigéniques, *Shigella flexneri*

Ecole doctorale: SEVAB (Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries)

Axe : Ingénieries microbiennes et enzymatiques

Laboratoire: Toulouse Biotechnology Institute (UMR CNRS 5504, UMR INRA 792), INSA de Toulouse

SCIENTIFIC PRODUCTION

Publications

Dudnik, A. Almeida F., Andrade R., Avila B., Bañados P., Bassard J.-E., **Benkoulouche M.**, ... Förster J. BacHBerry: BACterial Hosts for production of Bioactive phenolics from bERRY fruits. *Phytochemistry Reviews* (2017), pp.1-36. <http://doi.org/10.1007/s11101-017-9532-2>

Benkoulouche M., Fauré R., Remaud-Siméon M., Moulis C. and André I. Harnessing glycoenzyme engineering for synthesis of bioactive oligosaccharides *Interface Focus* (2019) 9: 20180069. <http://doi.org/10.1098/rsfs.2018.0069>

Hu# Z., Benkoulouche M.#, Barel L-A.#, et al., Submitted

A convergent chemoenzymatic strategy to deliver a diversity of *Shigella flexneri* serotype-specific O-antigen segments from a unique lightly protected tetrasaccharide core.

Benkoulouche M.#, Ben Imeddourene A.# et al., Draft is publication-ready

Redirecting substrate regioselectivity using engineered ΔN_{123} -GBD-CD2 branching sucrose for the production of *S. flexneri* 3a, 4a and 4b specific pentasaccharide repeating units

Benkoulouche M.#, Ben Imeddourene A.# et al., Draft is publication-ready

Computer-aided redesign of a branching sucrose active site for glycosylation diversification of a tetrasaccharide backbone for access to *S. flexneri* antigenic oligosaccharides

Benkoulouche M. et al., In preparation

Structural characterization of a GH70 branching sucrose with $\alpha 1,3$ linkage specificity

#Co-first authors

Patent

L. Mulard, G. Le Heiget, Z. Hu, L.A. Barel, I. André, C. Moulis, M. Remaud-Siméon, S. Barbe, **M. Benkoulouche**, A. Ben Imeddourene
WO/2019007999: Protected tetrasaccharides, their process of preparation and their use in the synthesis of bacterial oligosaccharides

Oral communications

Benkoulouche M., Ben Imeddourene A., Barel L.A., Hu Z., Le Heiget G., Barbe S., Mulard L., Remaud-Siméon M., Moulis C., André I. Enzymatic glycosylation of a chemically-protected tetrasaccharide to access antigenic oligosaccharides.

Carbohydrate Bioengineering Meeting 13th edition (CBM13). Toulouse, France. May 19-22, 2019.

Benkoulouche M.

“Les Doctoriales”, 20th edition. Albi, France. June 18-23, 2017.

Nominee for best oral communication.

Moulis C., Claverie M., Molina M, Grimaud F., **Benkoulouche M.**, Severac E., Cioci G., Remaud-Simeon M. Sucrose bioconversion in linear and branched α -glucans by discovery and engineering of bacterial α -transglucosylases. International Symposium in Green Chemistry. La Rochelle, France. May 13-17, 2019.

Ben Imeddourene A., **Benkoulouche M.**, Esque J., Barbe S., Mulard L.A, Remaud-Siméon M., Moulis C. and André I. Computer-aided engineering carbohydrate-active enzymes for the synthesis of antigenic carbohydrates. IR-RMN meeting. Orléans, France. Oct 09,. 2018

Barel L.-A., Hu Z., Le Heiget G., **Benkoulouche M.**, Imeddourene A., Guerreiro C., Moulis C., Remaud-Siméon M., André I., Mulard L.A. An innovative chemoenzymatic synthesis of type-specific pentasaccharides toward *Shigella flexneri* vaccine development. Institut Pasteur - Structural Biology and Chemistry Departmental Days, Pornichet, France. October 15-17, 2018.

Barel L.-A., Hu Z., Le Heiget G., **Benkoulouche M.**, Imeddourene A., Guerreiro C., Moulis C., Remaud-Siméon M., André I., Mulard L.A. Chemical synthesis and enzymatic glucosylation of lightly protected tetrasaccharides toward a *Shigella* glycovaccine. 27^{èmes} Journées du Groupe Français des Glycosciences, Nouan-Le-Fuzelier, France. May 21-25, 2018.

Barel L.-A., Hu Z., Le Heiget G., **Benkoulouche M.**, Imeddourene A., Guerreiro C., Moulis C., Remaud-Siméon M., André I., Mulard L. A. Divergent chemo-enzymatic synthesis of *Shigella flexneri* pentasaccharides for vaccine development. 19th International Carbohydrate Symposium, Lisbon, Portugal. July 15-19, 2018.

Posters

Benkoulouche M., Ben Imeddourene A., Hu Z., Le Heiget G., Barbe S., Mulard L.A., Remaud-Siméon M., Moulis C., André I. Enzymatic glucosylation of a chemically-protected tetrasaccharide scaffold to access *Shigella flexneri* antigenic oligosaccharides. 19th European Carbohydrate Symposium (EuroCarb). Barcelone, Espagne. July 2-6, 2017.

Ben Imeddourene A., Barbe S., Esque J., **Benkoulouche M.**, Moulis C., Mulard L.A., Remaud-Siméon M., André I. Computational study and prediction of enzyme-carbohydrate interactions to assist the development of novel chemo-enzymatic routes towards O-antigenic oligosaccharides. TouCAM : Toulousaines du Calcul Atomique et Moléculaire, Toulouse, France. Nov. 9-10, 2017.

Ben Imeddourene A., Barbe S., Esque J., **Benkoulouche M.**, Moulis C., Mulard L.A., Remaud-Siméon M. and André I. Computational study and prediction of enzyme-carbohydrate interactions to assist the development of novel chemo-enzymatic routes towards O-antigenic oligosaccharides. Groupe de Graphisme et de Modélisation Moléculaire, GGMM 2017. Reims, France. May 9-11, 2017.

Ben Imeddourene A., **Benkoulouche M.**, Barbe S., Moulis C., Mulard L.A., Remaud-Siméon M., André I. Computer-aided Design and Chemo-enzymatic Synthesis Working in Synergy to Develop a Carbohydrate-based Vaccines. Workshop Enzyme Engineering: Bright Strategies from Theory and Experiments, CECAM-HQ-EPFL, Lausanne, Switzerland. June 2016.

TABLE OF CONTENTS

SUMMARY.....	1
SCIENTIFIC PRODUCTION.....	7
TABLE OF CONTENTS.....	11
LIST OF ABBREVIATIONS.....	15
INTRODUCTION	19
General introduction	21
Objectives of the thesis.....	22
BIBLIOGRAPHY.....	25
CHAPTER I.....	27
PART I: Harnessing glycoenzyme engineering for synthesis of bioactive oligosaccharides	29
Introduction.....	29
Utilization of engineered Leloir-type glycosyltransferases	33
Engineered Glycoside-Hydrolases as synthetic tools	39
Exploitation of engineered Glycoside-Phosphorylases in synthesis.....	47
Outlook and future directions	49
PART II: An overview on shigellosis-causing bacteria, vaccine strategies and previous projects .	51
Generalities and epidemiological distribution of <i>Shigella</i> bacteria.....	51
<i>Shigella flexneri</i> polysaccharide antigens.....	52
Vaccine strategies to fight shigellosis.....	56
Highlights on the prior work done within the frame of TBI-Pasteur Institute collaboration	58
PART III: The α-transglucosylases from GH70 family	65
Classification of sucrose-active α -transglucosylases	65
Reactions catalyzed by GH70 enzymes	65
Mechanistic insights of the transglucosylation reaction	69
General description of sequence, structure and dynamics features	71
Latest evolutionary evidences on GH70 enzymes.....	78
A quick look at structure-function relationships of the branching sucrase GBD-CD2.....	79
Examples of engineered GH70 enzymes for glucodiversification	81
RESULTS	85
CHAPTER II.....	87
A convergent chemoenzymatic strategy to deliver a diversity of <i>Shigella flexneri</i> serotype-specific O-antigen segments from a unique lightly protected tetrasaccharide core	89
Introduction.....	89
Design and chemical synthesis of tetrasaccharide 12 (ABC'D')	91
Enzymatic glucosylation of tetrasaccharide 12 (ABC'D')	95
Conclusion	99
Acknowledgements	99
Supporting Materials & Methods	100
Chemical synthesis	100
Methods for enzymology	116
Supporting Schemes and Figures	119
Supporting Tables.....	122
MS data.....	123

Complementary Results	130
CHAPTER III.....	135
Redirecting substrate regioselectivity using engineered ΔN_{123}-GBD-CD2 branching sucrases for the production of <i>S. flexneri</i> 3a, 4a and 4b pentasaccharide repeating units.....	137
Introduction.....	137
Results & Discussion.....	138
Conclusion	150
Material & Methods.....	151
Supplementary Information.....	156
Supporting Figures	156
Supporting Tables.....	161
Complementary Results	162
CHAPTER IV	167
Computer-aided engineering of a branching sucrose for the glucodiversification of a tetrasaccharide precursor of <i>S. flexneri</i> antigenic oligosaccharides.....	169
Introduction.....	169
Results & Discussion.....	171
Conclusion	181
Material & Methods.....	182
Supplementary Information.....	186
Supporting Methods.....	186
Supporting Figures	188
Supporting Tables.....	194
CONCLUSIONS AND PERSPECTIVES	197
General conclusion.....	199
Perspectives	205
REFERENCES	211

LIST OF ABBREVIATIONS

A

ABC'D': allyl α -L-rhamnopyranosyl-(1 \rightarrow 2)- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-*O*-chloroacetyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-deoxy-2-trichloroacetamido- β -D-glucopyranoside
ACN/MeCN: acetonitrile
Asp: aspartate/aspartic acid
ASR: alternansucrase
AU: absorbance unit

B

BRS: branching sucrase

C

CA/CIAC: chloroacetyl
CAD: Charged Aerosol Detector
CASTing: Combinatorial Active-Site Saturation Testing
CAZy: Carbohydrate-active enzyme
CD2: Catalytic Domain 2

D

DNS: acid 3,5-dinitrosalicylic
DP: degree of polymerization
DSF: Differential Scanning Fluorimetry
DSR: dextransucrase

E

EC: Enzyme Commission

F

FACS/FADS: fluorescence activated cell/droplet sorting
FEL: free energy landscape
Fru: fructose

G

GAG: glycosaminoglycan
GBD: Glucan Binding Domain
GC: Gas Chromatography
GH: glycoside hydrolase
Glc: glucose
GlcNAc: *N*-acetyl- glucosamine
Glu: glutamate/glutamic acid
GP: glycoside phosphorylase
GS: glucansucrase (or glycosynthase when specified)
GT: glycosyltransferase
GTF: glucanotransferase

H

HMBC: Heteronuclear Multiple-Bond Correlation
HMO: Human Milk Oligosaccharide
HPAEC-PAD: High Pressure Anion Exchange Chromatography - Pulsed Amperometric Detection
HPLC: High Performance Liquid Chromatography
HRMS: High Resolution Mass Spectrometry

HSQC: Heteronuclear Single Quantum Coherence

I

IM: isomaltose
ISM: iterative saturation mutagenesis

L

LPS: lipopolysaccharide

M

MD: Molecular Dynamics
MM/GBSA: Molecular Mechanics/Generalized Born Surface Area
MS/MS: Tandem Mass Spectrometry

N

NMR: Nuclear Magnetic Resonance

O

O-Ag: O-Antigen

P

PCA: Principal Component Analysis
PCR: Polymerase Chain Reaction
PDB: Protein Data Bank

Q

QDF COSY: Double Quantum Filtered COrrrelation SpectroscopY

R

REACH: Registration, Evaluation and Authorization of Chemicals
Rha: rhamnose
RMSD: Root Mean Square Deviation
RU: Repeating Unit

S

SDS-PAGE: Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
Sf: *Shigella flexneri*
Suc: sucrose

T

TCA/Cl₃Ac: trichloroacetamide
TFA: trifluoroacetic acid
TG: transglycosylase
TLC: Thin Layer Chromatography
T_m: Melting Temperature
TOF: Time-Of-Flight

U

U: enzymatic unit
UDP: uridine diphosphate
UV: ultraviolet

W

WHO: World Health Organization

INTRODUCTION

General introduction

Since the very first chemical syntheses at industrial scale, synthetic routes have been widely explored to access molecules of defined structures and able to fit the needs of our growing society. They can find applications in all fields that support our modern way of living, whether it is technology, healthcare, food and feed, textile or materials. These molecules can be quite simple and constitute versatile building blocks or precursors of value-added molecules such as propylene (C_3H_6) for polypropylene (PP) (Ipatieff and Pines 1936), or ethylene dichloride ($C_2H_2Cl_2$) for polyvinyl chloride (PVC) (Wickson and Grossman 2008). But they can also be more complex, such as carbohydrates that are of great use in many fields thanks to their unique properties: pharmaceuticals (e.g. glycans used as probe for diagnosis (Hu et al. 2019) or entering the composition of vaccines (Zimmermann and Lepenies 2015); dextran as blood plasma substitute (Grönwall and Ingelman 1945)), agroindustry (oligosaccharides acting as prebiotics (Gibson et al. 2017)), cosmetics (sulfated polysaccharides with antioxidant activity (Jiao et al. 2011)), materials (cellulose derivative (Klemm et al. 2011)), heavy industries (biofuels (Kawai and Murata 2016)) etc.

Synthetic chemistry of carbohydrates has benefited from tremendous improvements over the last decades, especially related with the combined use of Liquid and Gas Chromatography techniques with Nuclear Magnetic Resonance spectroscopy and Mass Spectrometry that have speeded up the structural characterization and the isolation of the molecules in pure forms (Linhardt and Bazin 2001). Usually extracted from natural sources (e.g. cellulose for papermaking), carbohydrates can also be produced through synthetic technologies, which remain however challenging. Indeed, carbohydrate assembly requires stereo- and regio-selective control of donor and acceptor reactivity to achieve specific synthesis, often needing to use and manipulate protecting groups to mask the reactivity of the numerous hydroxyl functions. Despite these impediments, carbohydrate synthesis has been used at the industrial scale, from simple forms to more complex specialty carbohydrates (e.g. kilogram-scale production of synthetic heparin used as anticoagulant and obtained in 30+ steps (Trouilleux et al. 2019)).

The 21st century has seen a drift towards a bio-based society. Since 2007, the REACH (Registration, Evaluation and Authorization of Chemicals) regulation aims at reducing risks and hazards posed by chemicals on human health and the environment by limiting the use of toxic solvents and chemicals, hazardous wastes, and reaction conditions (Čihák 2009). It also promotes alternative methods that may lay in the use of either “greener” chemistry, or enzyme-based biocatalysis -that can also work in combination with organic synthesis toward more environmentally friendly processes.

Not only enzymes have the ability to work in aqueous environment and in mild temperature and pH conditions, some of them also display a high flexibility toward the substrates they can

utilize (so-called substrate promiscuity) making them versatile tools able to act on a wide range of relevant precursors, while keeping high regio and stereo-specificities. Thus, the last decades have witnessed the emergence of enzymatic and chemo-enzymatic pathways, leading to the synthesis of carbohydrate-based molecules. The association of robust chemical synthesis with versatile enzyme catalysts, has been increasingly used to tackle challenging carbohydrate syntheses.

This is of particular interest as many carbohydrates have sought-after bioactivities, such as prebiotic activity or antigenicity, which can be exploited for the development of relevant industrial products (respectively food additive and glycovaccines etc.). However, the enzymes do not always have the requisite properties especially when used with exogenous donor or acceptor substrates. In that context, directed evolution and molecular engineering techniques have emerged as a promising methodology to improve biocatalysts in terms of catalytic activity and stability in given reaction conditions (temperature, solvent, pH ...) but also to further extend the repertoire of modifications applicable to various molecules, including carbohydrates. Associated with computer-aided enzyme design that can guide the re-shaping of the enzyme active site toward novel substrates, engineering of enzymes able to assemble saccharidic blocks via regio and stereo-specific osidic linkages is constantly increasing. Thanks to a better comprehension of the sequence-structure-dynamics-function relationships, and to a more accurate modelling of the catalytic binding mode of a substrate in the enzyme active site, it is now possible to focus mutagenesis to key amino acid residue regions, and thus limit drastically the size of mutant libraries to be constructed and screened, avoiding then unnecessary efforts to explore very large libraries of millions or even billions of variants. Outstandingly, *de novo* computational enzyme design strategies have led in recent years to the impressive creation of completely new-to-Nature enzymes with unprecedented catalytic properties, showing the tremendous potential of these approaches to access greater molecular diversity (Jiang et al. 2008).

Objectives of the thesis

This thesis work is part of the project CarbUniVax funded by the French National Agency (ANR, 2016-2019) that associates the Catalysis and Enzyme Molecular Engineering team of the Toulouse Biotechnology Institute (TBI, formerly LISBP) at INSA Toulouse and the Glycochemistry Unit led by Laurence Mulard at the Pasteur Institute, Paris. In the context of this thesis, we have more particularly explored the α -transglucosylases from the glycoside hydrolase family 70 (GH70), which are excellent transferases that use very cheap and abundant sucrose as donor substrate, to develop programmed chemo-enzymatic pathways to access antigenic oligosaccharides representative of *Shigella flexneri* serotypes. These bacteria are responsible for endemic shigellosis in developing countries and kill thousands of people each year, in particular young children. Previous collaborative projects between TBI and

the Pasteur Institute, respectively named OptiGluc (2005-2008) and GlucoDesign (2009-2012), both funded by the ANR, revealed the potential of using sucrose-active α -transglucosylases to regio-selectively glucosylate exogenous lightly chemically protected mono- and disaccharides (Champion et al. 2009a; Salamone et al. 2015a; Vergès et al. 2015), which were subsequently converted by chemical means into precursors of haptens specific of prevalent *S. flexneri* 3a, 1b and 2a serotypes (Champion et al. 2009a; Salamone et al. 2015b).

Taking on a larger challenge, our goal in the current project was to identify enzymes able to regio-selectively glucosylate a chemically modified tetrasaccharide precursor at relevant positions, characteristic of glucosylation patterns encountered in the various *Shigella flexneri* serotypes. Whenever native GH70 sucrose-active transglucosylases did not present the needed specificity, a computer-aided design strategy was followed in order to further diversify and extend the range of oligosaccharides that could be accessed from one single enzyme template using one single tetrasaccharide acceptor precursor. Ultimately, the goal of the CarbUniVax project is to use the oligosaccharides produced either via chemical or chemo-enzymatic processes, to develop a synthetic multi-serotype glycoconjugate vaccine candidate containing molecules representative of serotypes *Sf2a*, *Sf3a/3b* and *Sf1a/1b*, which are prevalent and represent almost 50% of all *Shigella* infection in sub-Saharan Africa and South Asia, and potentially including less common serotypes 4a/4b and 5a.

In this context, the main objectives of my PhD project were thus to screen α -transglucosylases from the GH70 family, and more particularly so-called branching sucrases, either native or engineered, for their ability to produce glucosylated tetrasaccharides representative of the serotype repeating unit mentioned above, and at more fundamental level to unravel the determinants involved in recognition and glucosylation of the tetrasaccharide precursor at specific positions.

Chapter I is the introduction section of this thesis and comprises first a published review of the recent literature on the use of engineered enzymes for the synthesis of bioactive oligosaccharides. Then, further information is given on the context of the project, *Shigella* infection and vaccine strategies, with emphasis on the previous collaborative projects between TBI and the Pasteur Institute. The last part of the introduction provides an overview on the enzymes of the GH70 family which were used and engineered in the frame of the project.

The results obtained during this PhD project are reported in chapters II-IV that also constitute three research articles that should be submitted shortly. The first article is ready to be submitted for publication and the two last ones are in preparation.

Chapter II outlines the general concept of the programmed chemo-enzymatic pathway as well as the design and synthesis of the chemically modified tetrasaccharide acceptor that will be targeted within the project. More particularly, it describes the screening of 6 native GH70 branching sucrases which were envisioned for the glucosylation of the target tetrasaccharide.

The main glucosylation product was synthesized, purified and characterized by NMR spectroscopy as being characteristic of *S. flexneri* serotype 4a/4b.

Chapter III describes the screening of a pre-existing collection of 22 mutants of one particular branching sucrose, ΔN_{123} -GBD-CD2. Mutants revealed distinct product profiles, forming majorly one out of two mono-glucosylated tetrasaccharides, depending on the mutations introduced. Two of these mutants were selected for in-depth biochemical characterization. This revealed that depending on the mutant, glucosylation occurs preferably at either one or the other extremity of the tetrasaccharide chain, producing pentasaccharides specific of *S. flexneri* serotypes 4a/4b and prevalent 3a. These mutants were further investigated by molecular modelling in order to decipher the molecular determinants that were responsible for glucosylation at one or the other extremity of the tetrasaccharide.

Chapter IV reports the computer-aided re-design of a branching sucrose named BRS-B $\Delta 2$. Overall, up to 15 mutations were introduced in the active site of the enzyme, and a set of 49 mutants was proposed for experimental construction. Upon optimization of the recombinant production of the mutants, they were screened and their product profiles were analyzed in details. Altogether, 6 mono-glucosylated products out of the 8 possible ones, -revealing thus a tremendous glucodiversification-, were identified by a combination of NMR spectroscopy and MS/MS techniques. Out of the 6 mono-glucosylated products, a novel molecule representative of *S. flexneri* 5a serotype was identified. Molecular determinants and role of mutations on the regio-selective glucosylation of the tetrasaccharide were further investigated by molecular modelling.

Finally, the last chapter draws the main conclusions of this PhD thesis and future prospects. Overall, we managed to constitute an enzymatic toolbox able to glucosylate the tetrasaccharide backbone at 6 out of the 8 possible hydroxyl groups, with 3 products representative of *S. flexneri* serotypes 4a/4b, 3a and 5a. The best enzymes for glucosylation at each position were either native or engineered enzymes. In particular specific re-design of BRS-B $\Delta 2$ led to the largest glucodiversification of the tetrasaccharide. Strong of these results, we then propose novel perspectives for enzyme-based synthetic processes able to produce non-natural carbohydrates that could have a vaccine interest. The enzymes developed in this study constitute good starting points for further optimization, which would greatly benefit from obtaining crystallographic structures and a larger-scale screening method. To that extent, we started a work to attempt crystallization of the enzyme BRS-B $\Delta 2$, and the development of an engineered strain for high-throughput selection of sucrose-active enzymes.

BIBLIOGRAPHY

CHAPTER I

In this chapter, we first surveyed the utilization of engineered glycoenzymes, sometimes in combination with chemical steps, for the synthesis of natural bioactive oligosaccharides or their precursors. This led to the publication of a review article entitled “Harnessing glycoenzyme engineering for synthesis of bioactive oligosaccharides” in the journal Interface Focus in February 2019, that constitutes Part I of this bibliography section.

The two subsequent parts of this section present respectively the context of the project with an emphasis on Shigella and vaccine strategies and the family of enzymes used in this study.

PART I: Harnessing glycoenzyme engineering for synthesis of bioactive oligosaccharides

This section was published as a review article in *Interface Focus* (2019) 9: 20180069. <http://doi.org/10.1098/rsfs.2018.0069>

Authored by **Benkoulouche M.**, Fauré R., Remaud-Siméon M., Moulis C. and André I.

Abstract

Combined with chemical synthesis, the use of glycoenzyme biocatalysts has shown great synthetic potential over the last decades thanks to their remarkable versatility in terms of substrates and regio- and stereoselectivity that allows structurally-controlled synthesis of carbohydrates and glycoconjugates. Nonetheless, the lack of appropriate enzymatic tools with requisite properties in the natural diversity has hampered extensive exploration of enzyme-based synthetic routes to access relevant bioactive oligosaccharides, such as cell-surface glycans or prebiotics. With the remarkable progress in enzyme engineering, it has become possible to improve catalytic efficiency and physico-chemical properties of enzymes but also considerably extend the repertoire of accessible catalytic reactions and tailor novel substrate specificities. In this review, we intend to give a brief overview of the advantageous use of engineered glycoenzymes, sometimes in combination with chemical steps, for the synthesis of natural bioactive oligosaccharides or their precursors. The focus will be set on examples resulting from the three main classes of glycoenzymes specialized in carbohydrate synthesis: glycosyltransferases, glycoside-hydrolases and glycoside-phosphorylases.

Introduction

Carbohydrate chemical synthesis: limits & challenges

Carbohydrates are known to play major and complex roles in many biological systems and recognition processes (Varki 2017). As such, they have emerged as invaluable tools to probe carbohydrate-protein recognition and uncover novel therapeutic targets. Carbohydrates also find more and more applications in the food or feed industries as prebiotic compounds able to stimulate the growth of beneficial bacteria of human and animal microflora (Monsan and Paul 1995; Zhu et al. 2016; Schmid 2018). Development of synthetic routes to access these carbohydrate structures are thus highly demanded. However, in spite of many great advances in the field, conventional organic chemical synthesis of carbohydrates, especially oligosaccharides, remains highly challenging due to their tremendous structural diversity and complexity in terms of constituting monosaccharides with numerous reactive hydroxyl groups and glycosidic linkages. Despite the development of novel methods and tremendous improvements in selectivities and yields (thoroughly reviewed in (Nielsen and Pedersen 2018; Hsu et al. 2011)), chemical methods still suffer from the challenging protecting group

manipulations of the precursors necessary to obtain a regioselective glycosylation, the harsh methods, the use of heavy metal catalysts, and often the requirement of intermediary purification steps to separate stereoisomers. To circumvent some of these limitations, chemical synthesis is being more and more frequently combined with (or sometimes replaced by) the use of highly selective enzymatic catalysts when advantageous for the wanted synthetic transformation. Both chemical and enzymatic approaches used in combination should enable to overcome current synthetic bottlenecks and provide convergent chemo-enzymatic synthetic routes to access more easily carbohydrates. In many cases though, chemical synthesis remains invaluable for either the synthesis of building blocks that are further used for enzymatic glycosylation or for the incorporation of rare monosaccharides for which no enzymatic transfer has been described.

Natural glycoenzyme catalysts for carbohydrate synthesis

Biosynthesis of carbohydrates is usually ensured by a combination of enzymes, mostly glycosyltransferases (GTs), glycoside-hydrolases (GHs) or transglycosylases (TGs), and more rarely glycoside-phosphorylases (GPs). These enzymes have been evolved by Nature to endow them with a wide array of substrate specificities and in some cases, with improved stability in harsh environments (Brown and Babbitt 2014; Antranikian and Egorova 2003). Utilization of these enzyme catalysts has provided over the last decades an alternative to chemical synthesis in order to overcome some difficulties encountered in glycochemistry by taking advantage of their high stereo- and regioselectivities for non-protected substrates, and action in mild, aqueous conditions. However, utilization of natural enzymes is often hampered by a difficult recombinant production and the availability of substrates (in particular nucleotide-activated donors). To avoid these problems, *in vivo* production of carbohydrates, including the use of metabolically engineered whole cells, has been considered but has remained limited so far to the production of few biologically relevant carbohydrate structures (Chen 2018; Geremia and Samain 2000; Priem et al. 2002).

Contribution of enzyme engineering to the development of novel synthetic tools

In spite of the increasing number of glycoenzyme sequences identified from -omics technologies and the progress of automated annotation (Uchiyama and Miyazaki 2009; Cantarel et al. 2009; Yin et al. 2012; Park et al. 2010), the availability of enzymes having the requisite substrate specificity remains critical to perform highly specific glycosylations. With the remarkable progress of *in vitro* enzyme evolution, especially supported by *in silico* methods including computer-aided design (André et al. 2014; Khan et al. 2016), it has been possible to extend the repertoire of accessible reactions and improve catalytic efficiency and physicochemical properties of enzymes.

For a long time, engineering strategies to tune up enzyme properties have relied on site-directed mutagenesis (Hsieh and Vaisvila 2013) guided by various knowledge-based analyses derived from available sequences or 3D-structures to improve mostly catalytic activity or to alter substrate specificity. Directed evolution technologies (reviewed in (Packer and Liu 2015)), mimicking in accelerated the natural evolution of enzymes by combining cycles of mutagenesis using for example error-prone PCR (epPCR) or gene shuffling with subsequent high-throughput screening (reviewed in (André et al. 2014)) have also been increasingly used over the years to enhance enzyme catalytic performances, solubility and stability to temperature or solvents. Controlled randomization at specific gene locations can also be achieved using degenerate oligonucleotides (Chusacultanchai and Yuthavong 2004). However, these methods (summarized in Figure 1) require considerable screening efforts to sort out the requisite enzyme from very large mutant libraries (several thousand up to millions of mutants). With the development of more reliable predictive tools, rational and semi-rational strategies which combine several approaches have been privileged. These methods allow the construction of a limited number of mutants or small-size libraries of controlled diversity focused on key regions for the desired property (Jimenez-Rosales and Flores-Merino 2018) (Figure 1). Successful examples of enzyme engineering are abundant in the literature going up to the remarkable *de novo* computational design of biocatalysts able to catalyze new-to-Nature reactions (Huang et al. 2016). So far, such computational strategies have been poorly explored with carbohydrate-active enzymes although they offer good new perspectives in chemo-enzymatic carbohydrate pathways, notably by tailoring enzymes to chemically modified substrates and allowing programming of most convenient chemical and enzymatic steps (Champion et al. 2009a; Salamone et al. 2015a). Furthermore, these enzymes acting on non-natural substrates could open the way to non-natural glycan structures of potential interest in different fields.

In this review, we aim to provide a concise overview on the advantageous use of engineered glycoenzymes, sometimes in combination with chemical steps, for the synthesis of natural bioactive oligosaccharides or corresponding precursors, leaving aside glycoconjugates, glycopeptides and glycoproteins that have been the subject of many recent studies using native and engineered glycosyltransferases or oligosaccharyltransferases (OSTs) (Oldrini et al. 2018; Ihssen et al. 2015; Ollis et al. 2014; Hunter et al. 2018). Here, the focus will be set on examples resulting from the three main families (GTs, GHs and GPs) of carbohydrate-active enzymes specialized in carbohydrate synthesis (Lombard et al. 2014).

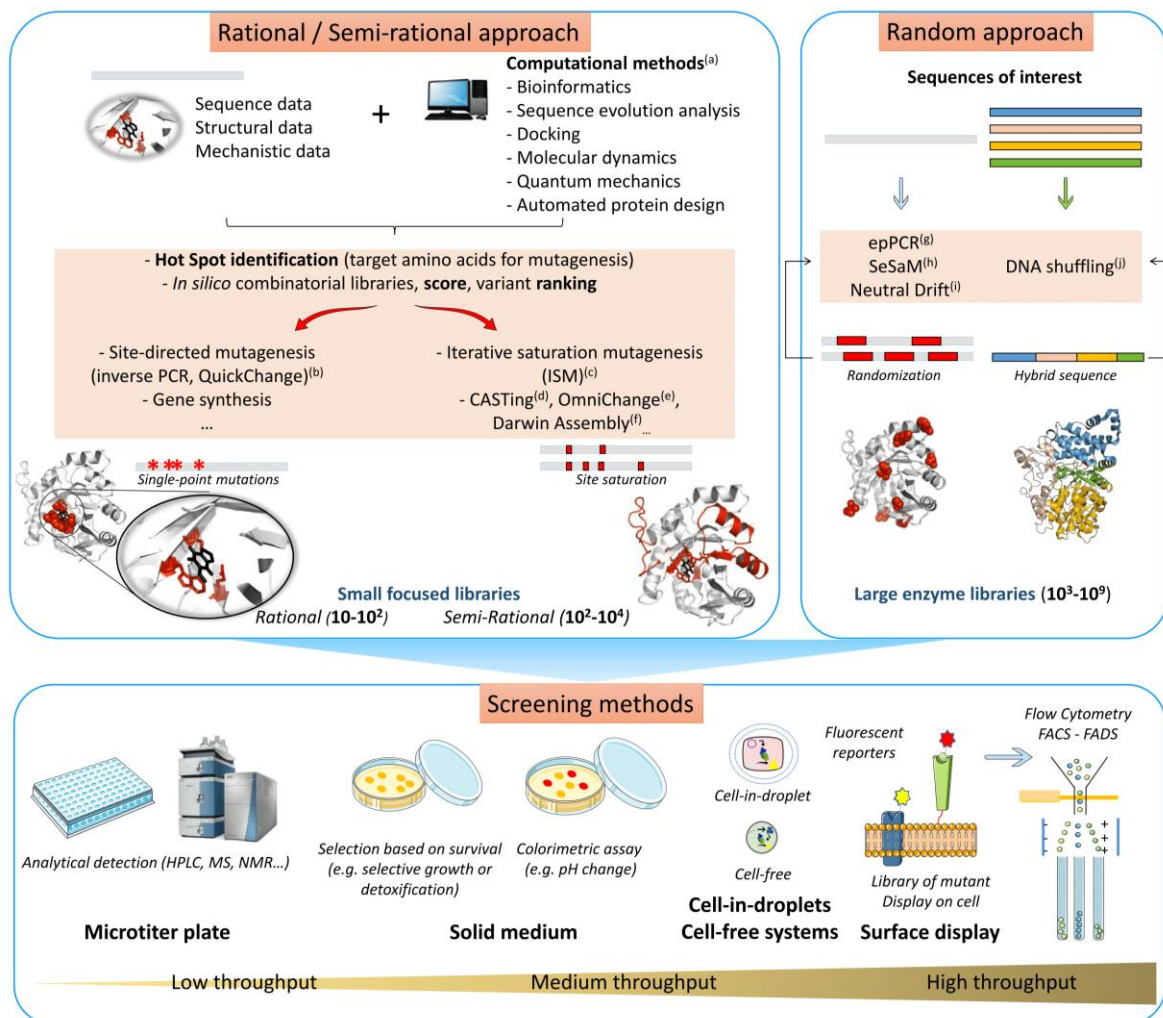


Figure 1: Random and rational/semi-rational engineering approaches to generate genetic diversity combined with screening methods enabling access to enzyme biocatalysts with desired and/or improved properties

Semi-rational and rational approaches make use of sequence, structure, and mechanism knowledge, often combined with (a) computational methods (Khan et al. 2016) in order to target regions or positions where to insert mutations into the DNA template. Rational libraries (~10¹-10² variants) can be constructed using simple techniques such as (b) inverse PCR or QuickChange (Hogrefe et al. 2002) or directly by gene synthesis. Semi-rational libraries (~10²-10⁴ variants) can be constructed using for example NNK or NDT degenerate codons using subsequently different approaches such as (c) Iterative Saturation Mutagenesis (ISM) (Reetz and Carballeira 2007), (d) Combinatorial Active-Site Saturation Testing (CASTing) (Reetz et al. 2005), and using molecular biology methods such as: (e) OmniChange (Dennig et al. 2011), or (f) the recently described Darwin Assembly method (Cozens and Pinheiro 2018). The resulting smaller libraries can then be screened using low or medium throughput screening assays such as liquid assays in microtiter plates coupled with analytical detection or screening on solid medium for the isolation of interesting mutants.

Random approaches use either (g),(h),(i) random mutagenesis techniques that can insert errors randomly into a specific DNA template (Packer and Liu 2015; Mundhada et al. 2011; Aharoni et al. 2005) or (j) shuffling in order to create hybrid sequences from different DNA templates (Stemmer 1994). The generated enzyme libraries (~10³-10⁹ variants) require then a suitable high-throughput screening method such as cell surface display or cell-free systems (Becker et al. 2004; Smith et al. 2015) coupled with Flow Cytometry - Fluorescence Activated Cell/Droplet Sorting (FACS/FADS) to sort out interesting variants based on specific fluorescent reporters.

Utilization of engineered Leloir-type glycosyltransferases

Leloir-type glycosyltransferases are the most common enzymes involved in biosynthesis of carbohydrate synthesis. GTs are responsible for the synthesis of most cell-surface glycoconjugates in mammalian systems and cell-wall carbohydrates in plants, fungi and bacteria (Lairson et al. 2008). These enzymes, which are mostly membrane-associated, act on nucleotide-activated sugars as donor substrates to catalyze glycosylation reactions, contrarily to non-Leloir GTs that use non-nucleotide sugars donors, such as lipid-sugars. Although the *in vitro* utilization of these enzymes for synthetic purpose remains difficult due to a poor expression and the cost of sugar-nucleotides, they remain of utmost interest thanks to the range of utilized glycosyl donors and recognized acceptors. The nine most common glycosyl moieties transferred by Leloir-type GTs are D-galactose (D-Gal), *N*-acetyl-D-galactosamine (D-GalNAc), D-glucose (D-Glc), *N*-acetyl-D-glucosamine (D-GlcNAc), D-glucuronic acid (D-GlcA), L-fucose (L-Fuc), D-mannose (D-Man), D-xylose (D-Xyl) and *N*-acetylneuraminic acid (Neu5Ac, the predominant form of sialic acid) (Lairson et al. 2008). The tremendous diversity of recognized acceptors (carbohydrates, lipids, polypeptides, DNA, antibiotics, ...) is tightly related with the sequence variability and the structural topology of the acceptor binding site, which is itself strongly dependent on the overall fold of the enzyme (Chang et al. 2011). Of great interest for synthetic purpose, GTs catalyze glycosyl transfer reactions with remarkable efficiency, regio-, stereo-, and chemospecificities enabling the formation of requisite glycosidic bonds. The Leloir-type GTs can proceed through either retention (double-displacement S_N1 - or dissociative S_{Ni} -like mechanisms) or inversion (S_N2 mechanism) of the anomeric carbon configuration (Figure 2) (Lairson et al. 2008). GTs isolated from natural organisms have been extensively explored in the synthesis of carbohydrates (Schmaltz et al. 2011; Wen et al. 2018). In addition, many instances of engineering of Leloir-type GTs, resulting either from rational mutagenesis or directed evolution strategies, have been reported in recent years to synthesize efficiently various complex carbohydrates, glycoconjugates, glycolipids, glycoproteins and antibiotics, and further extend the repertoire of accessible reactions (Nidetzky et al. 2018; Gantt et al. 2011)

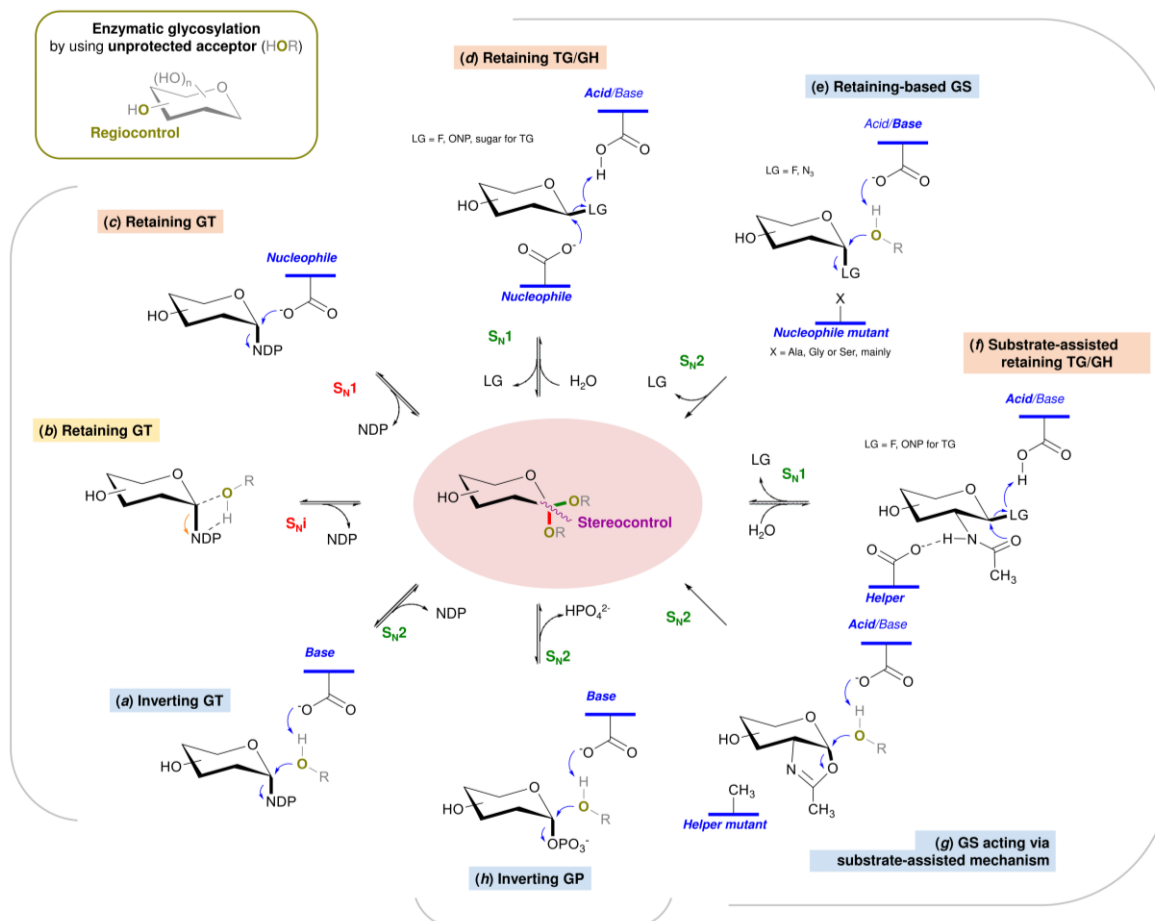


Figure 2: Main enzymatic pathways for glycosidic bond formation by wild-type and engineered glycoenzymes mentioned in the manuscript (GT, TG/GH and GS, and GP)

HOR, unprotected acceptor; LG, leaving group; GH, glycoside-hydrolase; GP, glycoside-phosphorylase; GS, glycosynthase; GT, glycosyltransferase; TG, transglycosylase; NDP, nucleotide diphosphate; ONP = O-nitrophenyl derivative. Only the initial stage of the glycoenzyme-catalyzed reactions are summarized. Therefore, the involvement of the acceptor substrates (ROH) that occurs in the second step of the mechanism for retaining enzymes (c, d and f) is not depicted.

(a) Inverting GTs use a single-displacement S_N2 mechanism through a single oxocarbenium ion-like transition state formed under catalytic assistance from a general base. Retaining GTs follow either (b) a S_N1 -like mechanism, with the leaving phosphate group functioning as the catalytic base to deprotonate the ROH acceptor, or (c) a double-displacement mechanism involving the formation of a covalent glycosyl-enzyme intermediate (Nidetzky et al. 2018; Lairson et al. 2008). (d) Classical Koshland two-step displacement mechanism of retaining TGs with enzyme nucleophile and acid/base residue proceeds via the formation of a covalent glycosyl-enzyme intermediate that opens access to transglycosylation even for innate hydrolytic GHs (Koshland 1953; Bissaro et al. 2015; Li and Wang 2018). α -Retaining mechanism differs only from the depicted β -retaining one by the anomeric configurations of the α -substrate and α -product as well as the more reactive β -covalent glycosyl-enzyme intermediate (André et al. 2010). (e) Single-displacement mechanism of inverting GS combines the use of both an activated mimic of the covalent glycosyl-enzyme intermediate and a catalytic nucleophile mutant of retaining GH (Cobucci-Ponzano et al. 2011a; Li and Wang 2018). (f) The two-step displacement mechanism with neighboring group participation of retaining N-acetylhexosaminidases, also termed substrate-assisted catalysis, proceeds via an oxazoline or an oxazolinium ion intermediate. The 2-acetamido moiety of the donor substrate plays the role of intramolecular nucleophile with the assistance of a helper residue to facilitate the formation of the intermediate (Li and Wang 2018). (g) The mutation of the helper, which promotes the intermediate formation during the catalysis of retaining N-acetylhexosaminidases, coupled with the use of activated oxazoline derivative leads to glycosynthase-like mutants (Umekawa et al. 2008; Li and Wang 2018). (h) Inverting GPs promote synthesis via either axial-to-equatorial or equatorial-to-axial substitution on the anomeric carbon of the glycosyl-1-phosphate donor substrates through the base catalytic residue. Retaining GPs (not depicted here) are classified within both GH and GT families and, therefore, share common S_N1 (c and d) or S_N1 -like (b) mechanisms (Luley-Goedl and Nidetzky 2010; Nakai et al. 2013; O'Neill and Field 2015; Puchart 2015).

In the following, we will exemplify the use of engineered Leloir-type GTs as synthetic tools, sometimes in combination with chemical synthesis, to access bioactive oligosaccharides, leaving aside reactions involving non-carbohydrate aglycones, which are reviewed elsewhere (Nidetzky et al. 2018; Desmet et al. 2012).

Synthesis of cell surface glycans

Involved in a variety of biological functions such as structural and modulatory roles, complex carbohydrates generally called glycans are found in bacterial cell walls or attached to the lipids of the outer membrane of Gram negative bacteria cells. These molecules are often targeted to treat inflammation, or to prevent bacterial infection. Glycans are also found at the surface of mammalian cells and constitute potential targets for the development of therapeutics against cancer (Varki 2017). The structure of these glycans is generally complex, involving a wide variety of glycosidic linkages, sugar types and arrangements of the glycosyl units. *In vivo*, these molecules are mostly assembled by the action of GTs.

As D-galactosyl moiety is often found in glycan structures, linked through α - or β - linkages, some of the most explored GTs in oligosaccharide synthesis are galactosyltransferases (GalTs) that use uridine diphosphate-galactose (UDP-Gal) as glycosyl donors to galactosylate a range of acceptor molecules linked via either α -(1,3), α -(1,4), α -(1,6), β -(1,3) or β -(1,4) bonds. Most of these enzymes have been identified from natural biosynthetic pathways of O-antigens in bacteria (*E. coli*, *Helicobacter pylori*, *Neisseria meningitidis*, *Neisseria gonorrhoeae*, ...) (Weijers et al. 2008; Chen 2018) and they have been advantageously utilized, often in combination with a chemically prepared oligosaccharide precursor, for the synthesis of glycans (Koeller and Wong 2000; Palcic 1999; Niggemann et al. 1998; Fang et al. 1998).

Among them are the antigens of ABO blood-group systems that have shown remarkable differences in antigenicity related with changes in their structure (Stanley and D. Cummings 2017). In particular, oligosaccharide epitopes of antigens A and B composed of a disaccharide containing L-Fuc and D-Gal residues, are differentiated by the presence of an additional D-GalNAc or a D-Gal residue, respectively, leading to α -D-GalNAc-(1,3)-[α -(L-Fuc-(1,2))- β -D-GalOR and α -D-Gal-(1,3)-[α -(L-Fuc-(1,2))- β -D-GalOR antigens. GTs involved in A/B blood-group differentiation have been well-studied, and were shown to be responsible for formation of α -D-GalNAc-(1,3)-D-Gal and α -D-Gal-(1,3)-D-Gal linkages respectively, on the α -L-Fuc-(1,2)- β -D-GalOR moieties of blood group O (H) antigens. The donor substrate specificity (UDP-Gal vs UDP-GalNAc) of human blood group A and B GTs was found to rely on only four amino acid differences located in the binding pocket of the donor substrate (R176→G, G235→S, L266→M and G268→A). Interchanges of these amino acids have led to remarkable changes in donor substrate recognition and catalytic efficiency of the GTs (Seto et al. 1997; Marcus et al. 2003). In particular, mutants, such as R176G/G235S/L266M, were found with dual specificity toward

UDP-Gal and UDP-GalNAc donors. Others, like G235S/L266M, displayed broader promiscuity toward donor substrate, and were found to accept UDP-Glc and UDP-GlcNAc as donor. These enzymes were used for the chemo-enzymatic synthesis of blood group A and B oligosaccharides and their corresponding analogues (Seto et al. 2000).

The α -(2,6)-sialosides found at the terminal or internal D-galactosyl unit of numerous glycan sequences are generally introduced by two β -galactoside α -(2,6)-sialyltransferases (SiaTs) in humans that have a narrow substrate specificity. Conversely, bacterial α -(2,6)-SiaTs such as β -galactoside α -(2,6)-SiaT from *Photobacterium damsela* (*Pd2,6ST*) have been widely used to produce sialosides as they have a more relaxed substrate specificity that could be further extended by enzyme engineering (Xu et al. 2018). Combining a structure and sequence analysis, two non-conserved amino acid residues (A200 and S232) found at the bottom of the active site of *Pd2,6ST*, and potentially differentiating D-Gal/D-GalNAc moieties from the glycan acceptor, were targeted by site-directed mutagenesis to reshape the binding pocket. Out of 15 single and double mutants, a double mutant A200Y/S232Y was found capable of regioselectively sialylating both the D-Gal and D-GalNAc at the non-reducing end of various glycans with higher efficiency than parental wild-type *Pd2,6ST*.

The active-site of a GT80 α -(2,3)-SiaT, from *Pasteurella dagmatis*, was redesigned in order to switch its regioselectivity from (2,3) to (2,6). Sequence-based comparison with α -(2,6)-SiaT guided the identification of putative residues responsible for the (2,3)/(2,6) regioselectivity. The corresponding P7H/M117A mutant was constructed and used for the synthesis of sialyllactose (SL) from CMP-Neu5Ac donor and lactose acceptor in a similar yield (72%) to that of the parental enzyme (75%). SL proved to be mainly 6'-sialyllactose and only trace amount of 3'-sialyllactose was produced compared to wild-type enzyme (Schmolzer et al. 2015).

A random library of $>10^6$ variants of GT42 SiaT from *Campylobacter jejuni*, CStill, was generated by epPCR and expressed in modified *E. coli* strains which were further screened using a fluorescence-based high-throughput methodology (Aharoni et al. 2006). Sialylated products were produced *in vivo* in the engineered cells which were able to transport both donor (CMP-Neu5Ac) and fluorescent galactose-containing acceptor (e.g. bodipy-lactose). Cells accumulating the most transfer products were isolated by FACS, leading to the isolation of a single mutant F91Y with improved transfer activity towards the labeled substrate (Aharoni et al. 2006).

An elegant strategy combining neutral genetic drift and site-directed mutagenesis was applied to engineer a polysialyltransferase (PST) from *Neisseria meningitidis* serogroup B and control polymer products (Keys et al. 2014). This enzyme is responsible for transferring successively α -(2,8)-linked sialic acids onto growing end of a polysaccharide chain, polySia. A single-residue (K69) was found to control the size distribution of the polySia. In particular, mutant K69Q showed a distributive mechanism of chain elongation with a narrow-size product profile.

Several other examples of SiaT engineering have been reported for the production of sialylosides (Ding et al. 2015; Choi et al. 2014; Guo et al. 2015).

Production of Human milk oligosaccharides

Human milk is composed of lipids, proteins, lactose, and more complex oligosaccharides called HMOs, known to have a prebiotic effect, especially by providing an advantage for the growth of beneficial *Bifidobacterium* species (Zivkovic et al. 2011), and also by having various immunomodulating effects (Newburg et al. 2005). The core structures of HMOs always display a lactose (Lac) moiety at the reducing end. This lactose can be branched with sialic acid (α -(2,3) or α -(2,6) linkages) or L-Fuc (α -(1,2) or α -(1,3) linkages) which corresponds to short-chain HMOs (trisaccharides). More complex HMOs are further extended in a linear or branched pattern with repetitions of lacto-*N*-biose (LNB, β -D-Gal-(1,3)-D-GlcNAc) or *N*-acetyl-lactosamine (LacNAc, β -D-Gal- β -(1,4)-D-GlcNAc) and can be decorated with sialic acid and/or L-Fuc residues (Kobata 2010; Sprenger et al. 2017; Han et al. 2012; Petschacher and Nidetzky 2016; Urashima et al. 2018).

In vivo, the pathway for the HMOs biosynthesis involves several GTs (Bode 2012). Bacterial GTs, in native or engineered form, have been privileged to mammalian ones for the *in vitro* synthesis of HMOs (reviewed in (Sprenger et al. 2017)). Synthesis of HMOs has benefited from chemo-enzymatic strategies. Xiao and colleagues produced a library of 31 simple and diverse HMOs from three chemically synthesized simple building blocks, comprising the core structure of more complex HMOs (lactose, D-GlcNAc extension, and the presence or not of an additional D-Galp at the non-reducing end) (Xiao et al. 2016). Four native GTs were used sequentially or in one-pot multi-enzymatic synthesis in the presence of the chemical building blocks: a β -(1,4)-galactosyltransferase, an α -(2,6)-sialyltransferase, an α -(1,3)-fucosyltransferase and an α -(1,2)-fucosyltransferase. This work relied on the beneficial use of substrate promiscuity of these enzymes, but undoubtedly, enzyme engineering technologies could provide access to more extended HMO structural diversity, improve enzyme activity, but also tackle recombinant expression challenges.

For instance, the α -(1,3)-fucosyltransferase from *Helicobacter pylori* that transfers the L-Fuc moiety from GDP-Fuc to LacNAc is of great interest to produce various antigen-like glycans (e.g. Lewis x, Le^x, and Lewis y, Le^y, epitopes) and HMO components, e.g. the trisaccharide block 3-fucosyllactose (3-FL). To improve soluble expression of this enzyme, Lin and colleagues truncated the C-terminal region, which is constituted by 2-10 repeats (depending on the strain) of 7 amino acids named heptad repeats, required for enzyme dimerization (Lin et al. 2006). By removing 5 out of the 10 heptad repeats together with 80 additional residues, the soluble expression of the enzyme was improved without any alteration of the structure or enzyme activity. A similar strategy was followed by Choi and colleagues to improve expression

of an α -(1,3)-fucosyltransferase from another *H. pylori* strain (Choi et al. 2016). The C-terminal 52 amino acids truncated enzyme (Δ 52 FutA) was left with only one heptad. The truncated enzyme gene was then synthesized after codon optimization for expression in *E. coli*. The solubility of the enzyme was found greatly improved, leading to better yields for the production of 3-FL (from 2% to 45%). This template was then used for further enzyme engineering using a two-step strategy of computer-aided analysis and iterative saturation mutagenesis (ISM) (Reetz and Carballreira 2007). Among the several mutants tested, a quadruple mutant A128N/H129E/Y132I/S46F revealed a 15.5-fold increase in specific activity toward 3-FL with a 96% yield achieved (corresponding to a 40-fold improvement) in only 1 h of reaction.

Synthesis of GAG oligosaccharides

Glycosaminoglycans (GAGs, i.e., heparin, heparan sulfate, chondroitin, sulfate, dermatan sulfate, hyaluronans) are composed of repeating disaccharide building blocks formed by an hexosamine (D-GlcNAc or D-GalNAc) and an uronic acid residue (either D-glucuronic acid and/or L-iduronic acid) or D-Gal. Due to the presence in their structure of negatively charged carboxylic groups and often sulfates, these polysaccharides exhibit a hydrophilic and anionic physicochemical nature that confer them key roles in biological interactions (cell adhesion, differentiation, signaling, ...) and allow them to maintain structural integrity of tissues (Zhang et al. 2010). Given the difficulty to produce large and complex GAG oligosaccharides by chemical synthesis, many chemo-enzymatic routes have been proposed using notably bifunctional GTs involved in GAG biosynthesis and sulfotransferases (O- and N-STases) that use high-cost 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as donor as well as epimerases converting D-glucuronic acid (D-GlcA) into L-iduronic acid (L-IdoA) (Lu et al. 2018; Li et al. 2017; Zhang et al. 2017; Suflita et al. 2015; Deangelis et al. 2013). To diversify and better control the accessible structures of GAGs, limited studies yet report the use of engineered enzymes. While investigating the mechanism of the bifunctional chondroitin synthase K4CP that catalyzes β -(1,3)-D-glucuronyl and β -(1,4)-N-acetyl-D-galactosaminyl transfer reactions to polymerize D-GlcA and D-GalNAc into a chondroitin chain, β -D-GlcA-(1,3)- β -D-GalNAc-(1,4), Sobhany and colleagues identified amino acid residues in the N-terminal region that affect polymerase activity while retaining transfer activities (Sobhany et al. 2012). Targeting this region by site-directed mutagenesis, mutants were obtained that enable transfer of both D-GlcA and D-GalNAc moieties without any polymerization, or transfer only D-GalNAc with no D-GlcA transfer or polymerization, allowing a fine tuning of the reaction. Site-directed mutagenesis performed on the DXD amino acid motifs of processive bifunctional GTs involved in heparan sulfate (HS) synthesis resulted in mutants associated with a single action, processing one type of glycan unit only that could be used to better control the oligosaccharide synthesis (Chavaroche et al. 2011).

Engineered Glycoside-Hydrolases as synthetic tools

In Nature, Glycoside-Hydrolases (GHs) are responsible for the cleavage of glycosidic linkages between carbohydrate moieties or between a carbohydrate and a non-carbohydrate moiety. These retaining or inverting enzymes usually proceed to the hydrolysis along two most common types of mechanisms but with several additional variations, that lead to two different stereochemical outcomes, either the retention or the inversion of the anomeric configuration (Figure 2) (Koshland 1953). The catalytic reaction includes two key amino acid residues: a general acid/base (proton donor) and a catalytic nucleophile or a general acid and a general base. Both retaining and inverting glycosidases are also classified as either *endo*- or *exo*-glycosidases, depending on whether they cleave within the middle or at the end of an oligosaccharide chain, respectively. Most glycosidases used for synthetic purposes have an *exo*-mechanism, catalyzing glycosyl transfer to the non-reducing terminal monosaccharide unit of acceptor substrates (Kobata 2013). The use of *endo*-glycosidases is scarcer. These enzymes could be appealing for synthetic purpose because they are quite robust and tolerant to solvents (Li and Wang 2016; Kobata 2013). However, their synthetic efficiency remains limited due to the high activity toward water.

Despite a hydrolytic activity, some retaining-GHs also display naturally efficient transferase activities and have thus been denominated 'transglycosylases'. Out of the 153 GH families currently defined in the CAZy database (Lombard et al. 2014), TGs can be mainly found in families GH2, 13, 16, 31, 70, 77, 23, 102, 103 and 104 (Williams 2015). They are known to function with diverse non-nucleotide sugar donors and preferentially transfer the glycosyl moiety in a regio- and stereoselective way onto a broad variety of acceptors distinct from water to synthesize a wide range of glycosides.

Both *endo*- and *exo*-GHs have also been engineered in order to limit the hydrolytic activity, meaning their ability to transfer the carbohydrate moiety onto water, and favour transglycosylation reactions (Bissaro et al. 2015). Among them are found the so-called 'glycosynthases' (GSs) that have been mutated on the catalytic nucleophile residue and are therefore devoid of their innate hydrolytic activity. The resulting enzymes are still able to catalyze the glycosyl transfer using a fluoride- or azide-activated glycosyl moiety, which mimics the covalent glycosyl-enzyme intermediate of the natural reaction, to perform the formation of glycosidic linkages (Figure 2). GSs can be effective biocatalytic tools for carbohydrate synthesis, achieving high yields in spite of a limited donor substrate selectivity and the relatively poor stability and availability of glycosyl fluorides (Mackenzie et al. 1998; Malet and Planas 1998; Moracci et al. 1998). The glycosynthase approach has been implemented in many chemo-enzymatic pathways for the synthesis of oligosaccharides (Cobucci-Ponzano et al. 2011a).

Overall, engineering of GHs has enabled not only to improve catalytic activity, but also to access diversified oligosaccharides, often new-to-Nature, with altered chain length, polydispersity and glycosidic linkage specificities. It has also helped to remove unwanted side reactions, for example primary or secondary hydrolytic activities, or to improve enzyme expression, stability and solubility. This section will be focused on the use of engineered GHs, mostly transglycosylases and glycosynthases, for the synthesis of a variety of relevant oligosaccharides.

Synthesis of cell surface glycans

Endo-β-N-acetylglucosaminidase (Endo-M) is a retaining GH85 endo-glycosidase that cleaves the β-(1,4)-glycosidic linkage in the *N,N'*-diacetylchitobiose core of N-glycans. The enzyme can also transfer whole sugar chains, onto a D-GlcNAc residue from a peptide or a protein. However, the secondary hydrolysis of the transglycosylation product obtained using endo-M is a key limitation of the approach. To avoid hydrolysis of the product (secondary hydrolysis), the active site of the enzyme was engineered to yield GS-like enzymes. The putative helper residue, which acts to facilitate the formation of the oxazoline or oxazolinium ion intermediate, was substituted by Gln or His amino acids. The corresponding mutants N175H, N175Q and its evolved N175Q/W251N version, showed both diminished primary and secondary hydrolysis activity as well high transglycosylation activity to form the *N,N'*-diacetylchitobiose core in the presence of activated oligosaccharide oxazoline that mimics the intermediate formed during the reaction (Umekawa et al. 2010; Sakaguchi et al. 2016). Therefore, Endo-M and its related enzymes embody powerful tools for the glycoengineering and glycodiversification of glycoproteins (Li and Wang 2018).

Transglycosylating activity of GHs has also been engineered to produce bioactive oligosaccharides (such as HMOs, oligosaccharide-antigens displayed on proteins or lipids, and tumor-associated antigens) that often contain the core galacto-*N*-biose (GNB, β-D-Gal-(1,3)-D-GalNAc) and LNB moieties. As an example, the retaining GH35 β-galactosidase from *Bacillus circulans* (BgaC) suffering from low yields due to re-hydrolysis of the transfer products (Warmerdam et al. 2013) was further engineered by Henze and colleagues into a galactosynthase mutant by substituting the catalytic nucleophile glutamate into a glycine (E233G mutant) enabling transfer of galactosyl residue from α-D-Gal fluoride to different β-linked *N*-acetyl-D-glucosamine acceptor substrates. The resulting β-(1,3)-linked D-galactosides, LNB conjugates, were stereo- and regioselectively synthesized in good yield (40–90%) without -as expected - any trace of secondary hydrolysis (Henze et al. 2014).

In order to recover its innate (1,3) regioselectivity during the β-D-Gal transfer for the synthesis of LNB, the substrate specificity of the *Tb*Gly E338G glycosynthase derived from a GH1 retaining β-glycosidase from *Thermus thermophilus* was expanded. The use of the 2-amino-

2-deoxy- β -D-glucoside instead of its bulky 2-acetamido derivative reoriented the regioselectivity of this glycosynthase-catalyzed reaction toward the formation of (1,3) instead of (1,4) linkages; the use of the thiophenyl group at the anomeric position of the acceptor being already a prerequisite to ensure a correct positioning of the acceptor in the active site and allowing the catalysis to proceed. Following the enzymatic coupling performed in 88% yield, additional sequential selective acylation of nitrogen followed by anomeric deprotection led to the production of pre-activated or free LNB disaccharides in 94 and then 97% yields (D'Almeida et al. 2009).

Fucosynthases were obtained from the retaining GH29 α -(1,3/4)-fucosidase BbAfcB from *Bifidobacterium bifidum*. For instance, the D703S α -L-fucosynthase mutant regioselectively catalyzed the synthesis of valuable biomimetic tri- and pentasaccharides, including the HMOs 3-FL and lacto-*N*-fucopentaose II (LNFP-II, β -D-Gal-(1,3)-[α -L-Fuc-(1,4)]- β -D-GlcNAc-(1,3)- β -D-Gal-(1,4)-D-Glc) in 13 and 41% yield, respectively) (Sakurama et al. 2012). The D242S α -L-fucosynthase based on the GH29 retaining α -L-fucosidase from *Sulfolobus solfataricus* catalyzed fucosylation of *p*NP- β -D-Gal to generate various regioisomers in an overall yield of 26% using the stable β -fucosyl azide as donor. Furthermore, *p*NP- β -D-GlcNAc was regioselectively fucosylated on its O-3 in 86% yield (Cobucci-Ponzano et al. 2009). Likewise, β -D-Gal azide was converted with high regioselectivity to α -Gal-disaccharides in 33-51% yields, depending on the acceptor, using the D327G galactosynthase mutant based on the GH36 α -galactosidase from *Thermotoga maritima* (Cobucci-Ponzano et al. 2011b).

A GH98 enzyme from *Streptococcus pneumoniae* SP3-BS71 was used to cleave respectively the A- and B-trisaccharides from type 2 core chains (LacNAc), responsible for determination of blood groups, with the objective of producing universal blood (Kwan et al. 2015). However, some linkages (e.g. LNB from type 1 core chains) were resistant to the activity of the wild-type enzyme. Following a structure-guided directed evolution strategy, involving several iterations of site-selected randomized mutagenesis combined with high-throughput screening, Kwan and colleagues isolated a mutant displaying 170-fold higher activity towards the cleavage of the LNB linkage compared to parental wild-type enzyme, while preserving a high-activity towards type 2 core-chains (LacNAc). This exemplifies the potential of building specific glycans or glycan building blocks with *exo*-enzymes by de-constructing more complex glycan structures.

Targeting bacterial polysaccharides

Pathogenic bacteria surface carbohydrates, either capsular polysaccharides (CPS) when bacteria produce a capsule, cell wall-associated glycans in Gram positive bacteria or lipopolysaccharides (LPS) in Gram negative bacteria are targets of choice for the development of glycoconjugate vaccines (Micoli et al. 2018b; Phalipon et al. 2009). Various strategies have been proposed to access the relevant immunogenic oligosaccharides that could elicit an

immune response and therefore be used in the composition of glycovaccines: extraction from natural sources, chemical synthesis (which remains the most used approach for the development of glycoconjugate vaccines in general), chemo-enzymatic synthesis and *in vivo* production by whole-cell biocatalysts.

Recently, programmed *in vitro* chemo-enzymatic routes have been reported by our group that take advantage of transglycosylase engineering to produce complex microbial cell-surface oligosaccharides and circumvent synthetic boundaries of glycochemistry. The different studies targeted more specifically the synthesis of oligosaccharides mimicking the O-antigen of several *Shigella flexneri* serotypes. Most known *S. flexneri* O-antigen repeats share a linear tri-L-rhamnosyl-*N*-acetyl-D-glucosamine tetrasaccharide backbone and the serotype specificity is partly defined by the α -D-glucosyl branched to the backbone. In the synthetic pathway foreseen for the production of the O-antigen repeats, regioselective glucosylation was considered using α -transglucosylases in order to overcome the poor α/β stereoselectivity of the chemical glucosylation process. Given that no native enzyme has been reported for this reaction, computer-aided engineering approaches were followed to redesign their active site in order to render them able to act on non-natural substrates, carrying protecting groups compatible with a subsequent chemical elongation. These chemically modified substrates were designed in order to integrate the enzymatic step at different stages of the synthesis, while taking into account the inherent constraints of chemical reactivity and enzyme-based catalysis. Different lightly chemically protected substrates derived from O-antigen repeats, and well suited for further chemical elongation leading to the *S. flexneri* haptens, were thus targeted to set-up the programmed chemo-enzymatic routes of monosaccharide analogs such as methyl α -L-rhamnoside and allyl 2-acetamido-2-deoxy- α -D-glucoside (Champion et al. 2009a, 2012), or disaccharide derivatives such as allyl 2-deoxy-2-trichloroacetamido- β -D-glucosyl-(1,2)- α -L-rhamnoside or allyl α -L-rhamnosyl-(1,2)-2-deoxy-2-trichloroacetamido- β -D-glucoside (Salamone et al. 2015a; Vergès et al. 2015).

In this context, sucrose-utilizing transglucosylases from GH13 and GH70 families, well known for their remarkable versatility regarding the acceptor substrate (Daudé et al. 2013a, 2014; Moulis et al. 2016; André et al. 2010), were selected as starting scaffolds to adapt the active site to the target molecules. Recombinant amylosucrase from *Neisseria polysaccharea* (*NpAS*) was used as a scaffold to tailor enzymes for the glucosylation of the monosaccharides. By randomizing mutations at 7 positions identified by molecular modelling in the acceptor binding site, a library of 133 single mutants followed by pairwise recombination of the mutations led to a library of ~20,000 mutants which was screened. Amylosucrase variants with either completely new specificity toward methyl α -L-rhamnoside or significantly enhanced (by up to a 400 fold compared to *NpAS*) toward allyl 2-*N*-acetyl-2-deoxy- α -D-glucoside were isolated. Yet limited, the catalytic efficiency of the engineered enzymes toward the target acceptors was found considerably improved compared to the parental wild-type enzyme. The best variants

were then used to synthesize glucosylated building blocks that were converted into acceptors and donors compatible with chemical elongation toward oligosaccharide fragments of the O-antigens of the targeted serotypes 1b and 3a (Champion et al. 2012).

A more ambitious design was then undertaken to re-engineer several binding subsites of *NpAS* and render it able to glucosylate a partially protected disaccharide acceptor (allyl (2-deoxy-2-trichloroacetamido- β -D-glucosyl)-(1,2)- α -L-rhamnoside) to generate a precursor of *S. flexneri*. An approach combining molecular docking, computational protein design, and sequence co-evolution derived information was developed to design libraries containing a limited number of authorized mutations at 23 selected positions of the active site. From a designed library of 63,000 clones selected from a set of 20^{23} theoretical combinations, one *NpAS* mutant displaying 7 mutations in the active site of the enzyme (R226L/I228V/F290Y/E300V/V331T/G396S/T398V) was able to glucosylate the disaccharide of interest using sucrose as a readily available donor substrate. A reaction for which there is no equivalent yet reported in the literature was made possible, generating the *S. flexneri* type 1b glucosylation pattern (Vergès et al. 2015).

Targeting another lightly protected disaccharide acceptor: allyl α -D-glucosyl-(1,4)- α -L-rhamnosyl-(1,3)-2-deoxy-2-trichloroacetamido- β -D-glucoside, which is well suited for further chemical elongation leading to the *S. flexneri* serotype 2a, an engineered dextranucrase (GBD-CD2 branching sucrose) was able to achieve a remarkable high yield of site-selective α -D-glucosylation (94%) of the targeted disaccharide. The product of enzymatic glucosylation was then chemically converted into the pentadecasaccharide hapten present in *S. flexneri* 2a-TT15, the first synthetic carbohydrate-based vaccine candidate against endemic shigellosis (Salamone et al. 2015a).

Production of Human milk oligosaccharides

L-Fuc residues are present in 50% of the HMOs. Alternatively to the use of Leloir-type α -fucosyltransferases (Petschacher and Nidetzky 2016), the fucosylation can also be achieved by α -fucosidases endowed with transfucosylating activity with yields between 5-40%, depending on the donor/acceptor ratio, although hydrolysis remains largely observed. To improve transfucosylating activity, a directed evolution strategy was applied to introduce randomly mutations on the full-length gene corresponding to the enzyme *TmaFuc* from *Thermotoga maritima* by epPCR (Osanjo et al. 2007). Five thousand clones of the α -fucosidase were first screened on solid medium in presence of the chromogenic X- α -L-Fuc substrate. From this first screening, 100 mutants were retained for further assessment of their transfucosylation activity using *pNP*- α -L-Fuc as donor substrate and *pNP*- β -D-Gal or β -D-Gal-(1,3)- β -D-Glc-O-phenyl (phenyl laminaribioside) as acceptor. From analysis of the mutations found in the best mutants, combined with structural analysis, a triple mutant

T264A/Y267F/L322P was constructed that exhibited improved transfucosylation activity yielding 59% for the production of α -L-Fuc-(1,2)- β -D-Gal-OpNP compared to 7% for the parental wild-type enzyme, in agreement with a considerable decrease of the hydrolytic activity (around a 100-fold) (Osanjo et al. 2007). The same group later employed a semi-rational approach to design an α -L-transfucosylase starting from the retaining *exo*-GH α -L-fucosidase from *Bifidobacterium longum subsp. infantis* (BiAfcB). Structural and sequence comparison of the latter with *Tma*Fuc guided the construction of several mutants containing the equivalent L321P substitution combined with neighboring mutations at donor binding site that aimed at improving the transglycosylation/hydrolysis ratio as devised by Teze and colleagues (Teze et al. 2014). The double mutant L321P/F34I yielded the best production of LNFP-II from 3-FL donor and lacto-*N*-tetraose (LNT) acceptor going from 12% for wild-type enzyme to 32%, and lowered secondary product hydrolysis. The mutant was also able to transfer the L-Fuc from 3-FL donor to three other acceptors: lacto-*N*-neotetraose (LNnT), lacto-*N*-fucopentaose I (LNFP-I) and 2'-FL, opening the way for the synthesis of more complex fucosylated HMOs (Saumonneau et al. 2015). More recently, a loop engineering strategy was applied to enhance the transfucosylation/hydrolysis ratio of a GH29 α -(1,3/4)-fucosidase from *Bifidobacterium bifidum* (BbAfcB). A 23 amino acid long α -helical loop located near the active site was replaced by the corresponding 17 amino acid long loop from the α -(1,3/4)-fucosidase from *Clostridium perfringens* to confer a better shielding from water molecules to BbAfcB. The hydrolytic activity on 3-FL was almost abolished (6,000 times lower than for wild-type enzyme) and the engineered enzyme transferred L-Fuc with a 39% yield from LNT to LNFP-II using 3-FL as donor (Zeuner et al. 2018).

Synthetic routes to access sialylated HMOs have attracted increasing attention (see review (Zeuner et al. 2014)). A *trans*-sialidase from *Trypanozoma cruzi* (*Tc*TS), able to transfer α -(2,3) sialic acid from sialoglyconjugate donors onto β -D-Gal glyconjugate acceptors, was used in various studies for the sialylation of different HMO precursors using a wide array of donors (casein glycomacropeptide, fetuin, α -Sia-(2,3')-Lac, *p*NP- α -Sia) and acceptors (Lac, LNT, galacto-oligosaccharides (GOS), LacNAc, LNB, LNnT, lacto-*N*-fucopentaose V (LNFP-V), GNB, β -D-Gal-(1,6)-D-Gal, ...) (Zeuner et al. 2014). An issue encountered with the use of *Tc*TS is related to the virulence factor of *Trypanozoma cruzi*, making it less interesting for food-related process. A highly homologous enzyme, the sialidase from *Trypanosoma rangeli* (*Tr*SA), was thus thought to be a good alternative to *Tc*TS. To confer a *trans*-sialidase activity to *Tr*SA, various engineering approaches have been used over the years. Structural comparison of *Tr*SA and *Tc*TS revealed similar topology of the active site with the exception of three amino acid residues (S120, G249 et Q284 in *Tr*SA corresponding to Y119, Y248 et P283 in *Tc*TS) assumed to be important for transferase activity (Buschiazzo 2000; Amaya et al. 2004; Paris et al. 2001). The amino acid interchanges at these three positions were not sufficient to confer a *trans*-sialidase activity (Paris et al. 2001)(Paris et al. 2005). Additional

mutations were thus introduced to recreate a favourable environment for binding the sialic acid moiety of the substrate. As a result, the variant TrSA_{5mut} containing five mutations (G249Y/Q284P/S120Y/M96V/A98P) displayed 1% of *trans*-sialidase activity compared to TcTS, transferring sialic acid from α -sialyl-(2,3)-lactose to lactose acceptor. Adding a sixth mutation in the active site led to mutant TrSA_{6mut} having a *trans*-sialidase activity enhanced up to 11% (Paris et al. 2005). In addition, Jers and colleagues (Jers et al. 2014) identified a positively charged motif in region 197-203 of TcTS, near the substrate binding cleft, that differed from TrSA. Hypothesizing its importance in *trans*-sialidase activity, a TrSA mutant incorporating this motif, Tr_{13mut}, was constructed, that contained a total of 13 mutations. This mutant showed a reduced hydrolytic activity compared to parental TrSA_{6mut} enzyme (around 80% less hydrolysis of 3-SL), but still maintained *trans*-sialidase activity. It was thus used for the sialylation of various acceptors: GOS, isomalto-oligosaccharide (IMOs), lactulose, melibiose, maltose and L-Fuc. To further improve its *trans*-sialidation activity, a computer-aided design approach was applied to TrSA_{5mut} earlier described (Pierdominici-Sottile et al. 2014). Five additional mutations were proposed (I37L, T39A, F59N, D285G, G342A) to help the stabilization of the transition state. Starting from Tr_{13mut}, the best mutant described in the literature at the time, Nyffenegger and colleagues added three of these mutations T39A, F59N and D285G, resulting in mutant Tr_{16mut} containing a total of 16 mutations and that showed a 13.6-fold improved overall transferase activity over hydrolysis ratio compared to parental TrSA_{5mut} (Nyffenegger et al. 2017).

Production of other prebiotics

Food bioactive oligosaccharides have gained interest as potent health-beneficial ingredients over the last years. They are not digested by human hydrolytic enzymes and reach the lower gastrointestinal tract where they are metabolized by beneficial intestinal microbiota, which in turn produce short-chain fatty acids providing health benefits to the host, as well as preventing the growth of harmful bacteria (Gibson and Roberfroid 1995). Several studies have suggested that prebiotic intake plays a role in the immune system, may protect against colon cancer and prevent cardiovascular diseases and metabolic syndromes. Their beneficial properties depend on the monosaccharide composition (Sanz et al. 2005), and current applications mainly concern fructo-, galacto-, xylo-, gluco-oligosaccharides (FOS, GOS, XOS) and lactulose, especially for food and beverage processing, dietary supplements, and animal feed (Mano et al. 2018). Utilization of microbial GHs (including glycosidases and transglycosylases) for a controlled regio- and stereoselective synthesis of prebiotic oligosaccharides has been extensively reviewed (Díez-Municio et al. 2014). Nonetheless, the search for new transglycosylases, the development of processes, as well as the improvement or redesign of catalysts by enzyme engineering is still relevant in order to develop new applications and/or to optimize the production yields.

For instance, an elegant engineering strategy of *Bacillus circulans* β -galactosidase was recently proposed by Tanaka and colleagues to alter its substrate specificity and enhance the production of short chain GOS (DP2 and DP3 only) from lactose substrate. For that purpose, they designed and screened a library of synthetic binding proteins (monobodies) that specifically targeted the +3 acceptor subsite. This local constraint prevented the accommodation of oligosaccharides of DP3 in the active site and their further elongation into DP4. At the same time, enzyme properties for lactose recognition and production of short chain GOS were not altered, allowing their accumulation in the medium (Tanaka et al. 2015).

Concerning XOS production, the X-ray structure resolution of an inactive mutant of the *Bacillus pumilus* IPO β -xylosidase in complex with xylobiose (DP2) enabled the identification of amino acid residues involved in xylobiose recognition. Notably, a tight binding of D-Xyl at subsite +1 was reinforced by mutating a phenylalanine (F503) into tyrosine, leading to a 20% increase of xylobiose production from xylose, compared to the wild-type enzyme. Xylobiose presents the highest prebiotic properties among xylo-oligosaccharides, and this study opened the route for similar engineering strategy within this family of enzymes (Hong et al. 2018).

Other interesting examples deal with microbial transglycosylases that use readily-available sucrose as a donor substrate to efficiently polymerize either the D-fructosyl or the D-glucosyl moiety of sucrose to produce β -D-fructan or α -D-glucan-type polysaccharides (i.e. fructansucrase or glucansucrase enzymes of GH families). By adding exogenous monosaccharides in the reaction medium with sucrose, these enzymes have the ability to transfer the D-fructosyl or D-glucosyl units onto these acceptor molecules to yield interesting dietary oligosaccharides. To lower the cost of such processes, enzyme engineering has been used to produce oligosaccharides from sole sucrose. For instance, the production of levan-type FOS (FOS with β -(2,6) glycosidic linkages) from DP2 to DP10 was developed thanks to the construction of a levansucrase-levanase fusion enzyme (levansucrase SacB from *Bacillus subtilis* and endolevanase LevB1 from *B. licheniformis*). Interesting yields of about 40% were obtained from sucrose substrate, to access 6-kestose, levanbiose and blastose among other FOS (Porrás-Domínguez et al. 2017). Concerning the production of gluco-oligosaccharides, determination of the 3D structure of the dextransucrase DSR-M from *L. citreum* NRRL B-1299, free or in complex with an isomalto-oligosaccharide (IMO) of DP4, allowed the design of variants producing short α -(1,6) chains of around 16 kg/mol (80% yield) by altering aromatic residues located in a glucan-binding domain shown to be involved in the anchoring of dextran chains (Claverie et al. 2017). Gluco-oligosaccharides rich in α -(1,6) linkages (IMO) are well-known prebiotic compounds used in Asian countries for many years. However, their prebiotic properties would be enhanced by increasing the content in rare α -(1,2) glycosidic bond. For that purpose, an engineered enzyme specialized in the α -D-(1,2) glucosylation of IMO acceptor molecules, called GBD-CD2 branching sucrose, was used for the controlled introduction of α -

(1,2)-branched D-glucosyl moieties, from 1 to 37% following the reaction conditions used (sucrose/donor ratio) (Fabre et al. 2005; Brison et al. 2010).

To finish, the amylosucrase from *Neisseria polysaccharea* (NpAS) is another α -transglycosylase that uses sucrose as D-glucosyl donor to catalyze the formation of soluble malto-oligosaccharides and insoluble amylose-like polymer. Using a computer-aided approach, semi-rational libraries of 2.7×10^4 mutants targeting as many as 23 amino acid positions of the active site were constructed and screened in the presence of sole sucrose (Vergès et al. 2017). This allowed identification of 17 mutants able to synthesize molecule which are not synthesized by the parental wild-type enzyme. Characterization of the three most promising mutants (namely 47A10, 37G4 and 39A8), containing respectively 7, 10 and 11 mutations in the active site, showed the production of a sucrose derivative, named erlose (α -D-glucosyl-(1,4)- α -D-glucosyl-(1,2)- β -D-Fructose), and panose (α -D-glucosyl-(1,6)- α -D-glucosyl-(1 \rightarrow 4)- α -D-glucose). These products, reported to have potential interest as sweeteners or prebiotic molecules, were never obtained using either natural amylosucrases previously characterized or engineered variants (Prapulla et al. 2000; Mäkeläinen et al. 2009).

Exploitation of engineered Glycoside-Phosphorylases in synthesis

Non-Leloir glycosyltransferases called Glycoside-Phosphorylases (GPs) cleave oligosaccharides in presence of inorganic phosphate, producing a glycosyl phosphate and a shorter oligosaccharide (Figure 2). This so-called phosphorolysis is reversible and can therefore be used to catalyze the stereospecific synthesis of glycosides using cheap phosphorylated glycosyl donors. When combined with a second appropriate phosphorylase, the said phosphorylated glycosyl donor can be produced *in situ* from the corresponding readily available glycosyl, making these enzymes suitable synthetic tools for the scaled-up production of carbohydrates (Nishimoto and Kitaoka 2007). GPs have been used scarcely in this purpose mainly due to a lack of diversity among the available enzymes. The known phosphorylases, their mechanism, structure and use in the synthesis of oligosaccharide have been extensively reviewed (Luley-Goedl and Nidetzky 2010; Nakai et al. 2013; O'Neill and Field 2015; Puchart 2015; Kitaoka 2015; Pergolizzi et al. 2017). Around 30 phosphorylases have been reported and were classified in the CAZy database as either GT or GH members. Furthermore, GPs can adopt either retaining or inverting mechanisms. GPs have a broad acceptor specificity as they can act on various carbohydrates (such as trehalose, kojibiose, sucrose, nigerose, maltose, ...) but their donor specificity is quite narrow (Desmet and Soetaert 2012; O'Neill and Field 2015). Several recent studies report the characterization and the engineering of GPs that could be used for glycoside synthesis (Kuhaudomlarp et al. 2018; Macdonald et al. 2018; Mukherjee et al. 2018; Awad et al. 2017).

Production of oligosaccharides

De Groeve and colleagues engineered both the donor and acceptor specificities of a cellobiose phosphorylase (CP) (reviewed in (De Groeve et al. 2011)), first discovered in extracts from the bacteria *Clostridium thermocellum* and belonging to the GH94 family. This enzyme catalyzes the reversible phosphorolysis of cellobiose into α -D-glucosyl 1-phosphate (α -D-Glc-1P). To diversify the products obtained by GPs, the donor specificity of CP from *Cellulomonas uda* was changed from cellobiose to lactose using a two-step strategy of directed evolution. First, random mutagenesis was applied by epPCR on part of the gene, targeting more specifically amino acid residues contained between T216 and V757 (out of the 822 amino acid enzyme), that include all residues located within 15Å of the catalytic site. Altogether, 10,000 mutants were generated and subsequently selected for their activity towards lactose using a growth-selective assay on minimal solid medium containing lactose as sole carbon source. Three active mutants were retained and further sequenced showing the presence of a set of 6 amino acid substitutions in all mutants. All mutations except N667T occurred at more than 15Å from the donor subsite. The variant LP1 (A397V/T508A/A512T/D557N/N667T/G681S) showed a 3-fold improved activity towards lactose and a 6-fold decreased activity towards cellobiose compared to parental wild-type enzyme. Mutant LP1 was then subjected to additional rounds of site-directed mutagenesis in order to decipher the role of each mutation by partial deconvolution. Mutations T508A present near the entrance of the active site and N667T located close to the donor site were the only mutations that conferred an activity towards lactose. The corresponding double mutant T508A/N667T named LP2 was further engineered on these two positions by site-directed saturation mutagenesis. This yielded the double mutant T508I/N667A, named LP3, which had the best activity toward lactose with a 50% increase compared to LP2, leading to a 7.5-fold increase compared to that of the parental enzyme (De Groeve et al. 2009). Mutant LP3 was ultimately used for the production of lactose in a completely new way compared to usual GalT enzymes that use expensive UDP-Gal as a donor.

CPs have a broad acceptor promiscuity and they are able to recognize several mono and disaccharides such as D-Glc, D-Man, D-GlcNAc, D-Xyl, L-Fuc, D-Ara, melibiose, gentiobiose, isomaltose, ... what is particularly valuable for glycoside synthesis (De Groeve et al. 2011). To further expand the acceptor specificity, De Groeve and colleagues identified a residue (E649) in the structure of CP from *Cellulomonas uda*, whose side-chain makes hydrogen bonding interaction with the D-Glc acceptor. Upon saturation mutagenesis of this residue, the mutant E649C was identified as being able to recognize new glucoside acceptors. Mutation E649C was then combined with mutations present in mutant LP3. Three additional amino acid positions located at the entrance of the active site were then identified for saturation mutagenesis on the basis of structure analysis. The resulting mutant (N156D, N163D, T508I,

E649G and N667A) revealed a broad acceptor specificity towards α - and β -glycosides (De Groeve et al. 2010).

Using the ISM technology, Chen and colleagues engineered the specificity of a trehalose phosphorylase for the production of β -D-Gal-1P instead of β -D-Glc-1P produced by the parental enzyme (Chen et al. 2014). Computational modelling of the lactotrehalose binding in the enzyme active site enabled identification of three amino acid residues that were further targeted by mutagenesis. Altogether, 600 variants were screened and the triple mutant L649G/A693Q/W371Y was identified with a 2,196-fold increase in activity for the release of β -D-Gal-1P. The reverse reaction allowed to produce lactotrehalose from β -D-Gal-1P and D-Glc.

Other examples of GP engineering can be found in the following references (Yamamoto et al. 2006; Nakai et al. 2010; Ye et al. 2012; Van der Borght et al. 2012; Hamura et al. 2013).

Outlook and future directions

Significant advances have been made over the last years regarding the enzyme-based *in vitro* synthesis of complex carbohydrates with the development of enzyme cascades, one-pot synthesis, and combined use of chemical and enzymatic approaches. Progress has noticeably been boosted by the advent of enzyme engineering technologies that enabled to access a whole new diversity of enzymes. This later has strongly benefited in recent years from the tremendous developments in bioinformatics (Liu and Neelamegham 2014; Sánchez-Rodríguez et al. 2014) and computer-aided design of enzymes (Khan et al. 2016) that allow nowadays to consider reprogramming of their substrate specificity toward the wanted glycosylation reaction, with sometimes the tailoring on purpose of enzymes to act on non-natural chemically modified carbohydrate precursors. Despite the impressive advances of rational engineering, semi-rational approaches leading to medium to high size library of variants still require the development of high throughput screening methods. In the field of glycoenzymes, those methods cannot always be easily implemented when looking for very specific sugar structures. They often necessitate the adjunction of analytical techniques, which are highly accurate but remain time consuming. Progress in NMR and mass spectrometry should help to tackle with more efficiency this question in the near future. The field is rapidly evolving with the democratization of these technologies, which could be also combined with post-synthesis and specific chemical labelling of the reaction products to facilitate product analysis at high throughput. In parallel, cell development for specific carbohydrate production inspired from existing biosynthetic pathways is also rapidly progressing (Chen 2018; Ruffing and Chen 2006). The incorporation of engineered enzymes gives total freedom to even more creative pathways to produce *in vivo* bioactive oligosaccharides and glycoconjugates at large scale and offers exciting perspectives.

When targeting a specific carbohydrate structure, several bio- and/or chemo-synthetic approaches can thus be employed. Of course, there is no easy way to anticipate the most appropriate one as the choice will depend on many factors including the structural complexity of the target molecules, the chemical difficulties, the enzyme and substrate availability, the environmental impact, regulation and ethical considerations, and far and foremost the economic advantage of the process knowing that production of oligosaccharides remain of high cost. All these constraints and issues will have to be considered. To develop innovative processes with the highest possible discernment, there is no doubt that strong interactions and interfacing between all these disciplines will be essential, hybrid and composite approaches constituting the future of synthetic glycobiology.

PART II: An overview on shigellosis-causing bacteria, vaccine strategies and previous projects

Generalities and epidemiological distribution of *Shigella* bacteria

Bacteria from the *Shigella* genus are Gram-negative enteroinvasive bacteria responsible for shigellosis, also called bacillary dysentery. Firstly isolated in Japan by Kiyoshi Shiga in 1898, members of the *Shigella* family are now described and classified into four groups exhibiting various serotypes: *S. dysenteriae* (group A, 13 serotypes), *S. flexneri* (group B, 15 serotypes and various emerging subserotypes), *S. boydii* (group C, 19 serotypes) and *S. sonnei* (group D, 1 serotype) (Barel and Mulard 2019; Lampel et al. 2018).

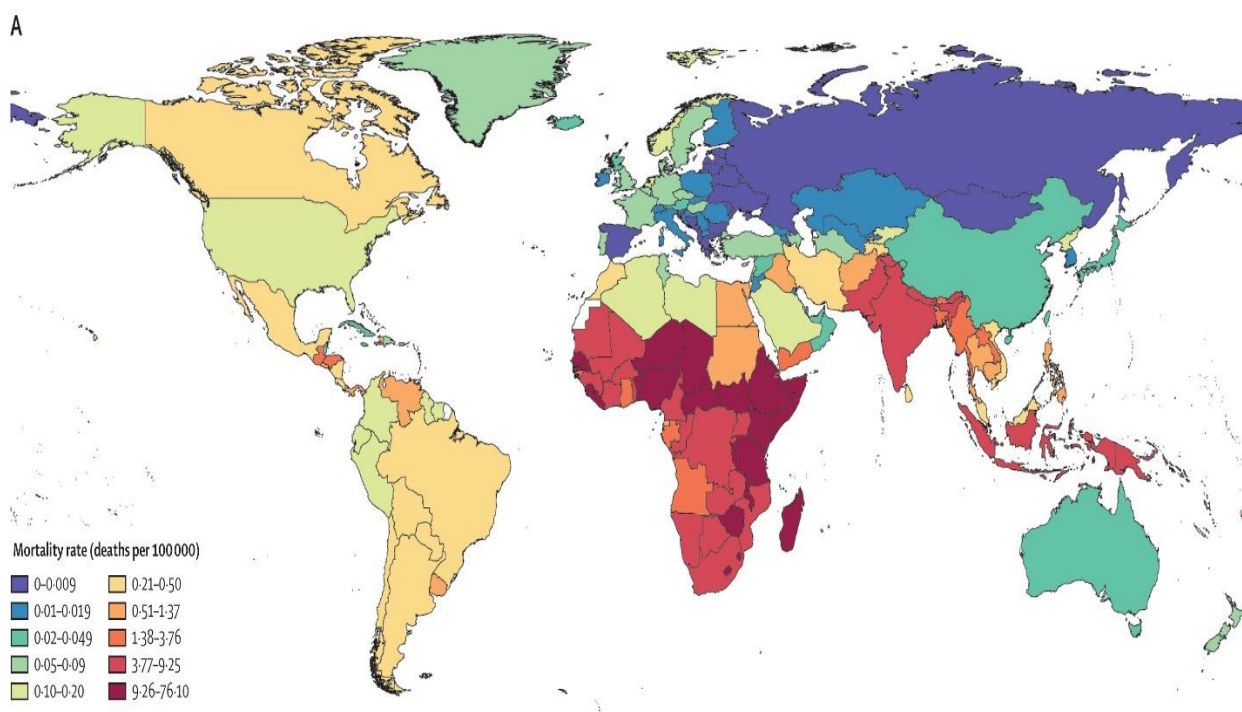


Figure 3: *Shigella* diarrhea mortality rate per 100 000 people in 2016 for all ages.

Figure from (Khalil et al. 2018) ATG=Antigua and Barbuda. VCT=Saint Vincent and the Grenadines. LCA=Saint Lucia. TTO=Trinidad and Tobago. Isl=Islands. FSM=Federated States of Micronesia. TLS=Timor-Leste.

Shigella dysenteriae 1 was responsible for burst of dysentery epidemics throughout the 20th century, in Central and East Africa, Central America, and South Asia (Levine et al. 2007), and is still a not-negligible cause of shigellosis, while *Shigella flexneri* is responsible for the endemic form of the disease. Small and targeted epidemics in developed countries are caused by *S. sonnei* (Bopp et al. 2015). Only 10 to 200 bacterium cells are necessary to cause the infection which is mainly transmitted through fecal-oral route (Dupont 1989), especially in areas where there are poor sanitation and difficult access to clean water (World Health Organization 2005). Shigellosis causes acute diarrhea, often hemorrhagic, but unlike other diarrheic diseases (e.g. caused by *Vibrio cholera* or enterotoxic *Escherichia coli*), treatment by sole re-hydration is often not effective enough. Moreover, the bacteria started to develop multi-resistance to the

first-line antibiotics (Puzari et al. 2018). The recent emerging *S. flexneri* serotype 4s is even able to resist the latest generation of antibiotics (Yang et al. 2016) making endemic shigellosis one of the predominant cause of death by infectious disease among children under age five in developing countries (World Health Organization 2005). Thus, with more than 269 million cases and 212,000 deaths (~30% among children under age 5) in 2016 (Figure 3, (Khalil et al. 2018)) and the risk of developing long-term complications (World Health Organization 2005), the development of an effective vaccine for broad protection against shigellosis-causing bacteria is of utmost importance. Despite notable improvements in sanitation over the last 15 years in the most affected areas, the need for such vaccine remains urgent (Yang et al. 2016).

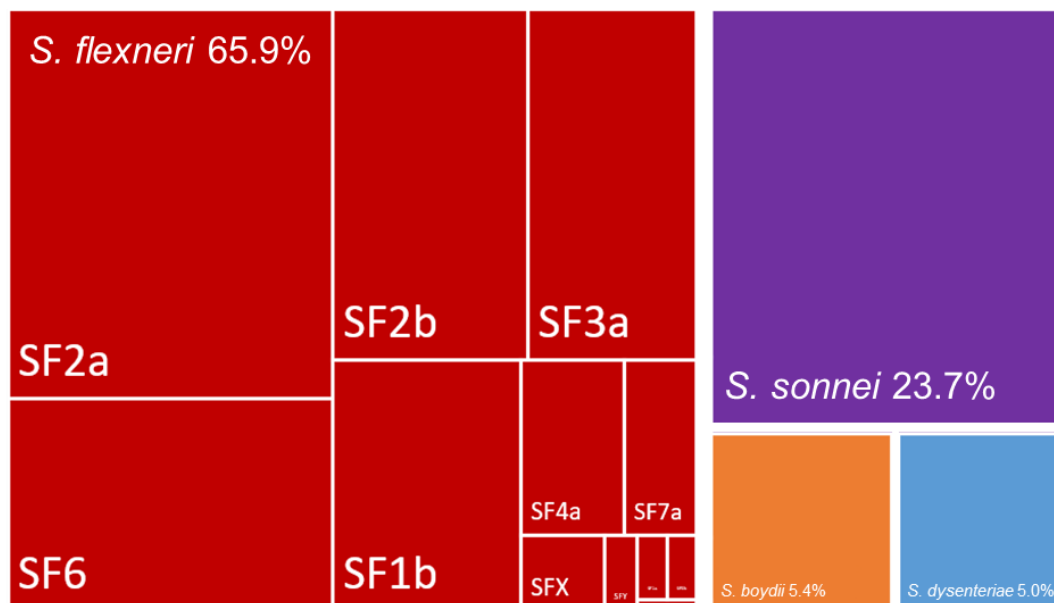


Figure 4: Species and serotype distribution of *Shigella* covering 4 sites in sub-Saharan Africa and 3 sites in South Asia, in 2014. Data from (Livio et al. 2014). Infographic reproduced with the kind authorization of Louis-Antoine Barel, Pasteur Institute

Prevalence of *Shigella* serotypes is strongly dependent on the geographical area. They are mainly from *Shigella flexneri* (SF2a/2b, SF3a, SF6 and SF1b), as well as from *S. sonnei* in developed countries and to a lesser extent from *S. boydii* and *S. dysenteriae* (Figure 4).

However, new subserotypes of *S. flexneri* have been increasingly found in certain areas such as i) SF1c (7a) in Bangladesh, Egypt, Indonesia, Pakistan and Vietnam (Qiu et al. 2012), ii) SFYv (Yb), SFxv (Xb) and SF1d in China (Shashkov et al. 2013; Luo et al. 2012), as well as iii) SF4s (Z) in China and Bangladesh (Qiu et al. 2011; Shahnaj et al. 2018), and (iv) SF4av (4e) in Australia and Russia (Perepelov et al. 2009).

Shigella flexneri polysaccharide antigens

The lipopolysaccharides (LPS) of the outer membrane of Gram-negative bacteria largely contribute to virulence, and interaction with the host (West et al. 2005; Morona and Bosch 2003). They are composed of a lipid and a polysaccharide part. The latter is defined by a core

oligosaccharide and an O-specific polysaccharide (also referred to as the O-antigen, O-Ag), itself formed by repeating units (RUs) which can vary in structure and composition depending on species and serotypes (Figure 5). All *S. flexneri* serotypes - except for serotype 6 - present a RU constituted of a linear tetrasaccharide composed of three linked L-rhamnose (A, B and C) and an N-acetyl-D-glucosamine (D) residues, also called motif ABCD: [\rightarrow 2)- α -L-Rhap-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow 3)- β -D-GlcNAc-(1 \rightarrow)] (Figure 5). All residues can be site-selectively modified by the addition of an α -D-glucopyranosyl and/or O-acetyl groups and/or phosphoethanolamine groups, as illustrated in Figure 7 (Allison and Verma 2000). These specific modifications stand for the serotype differentiation.

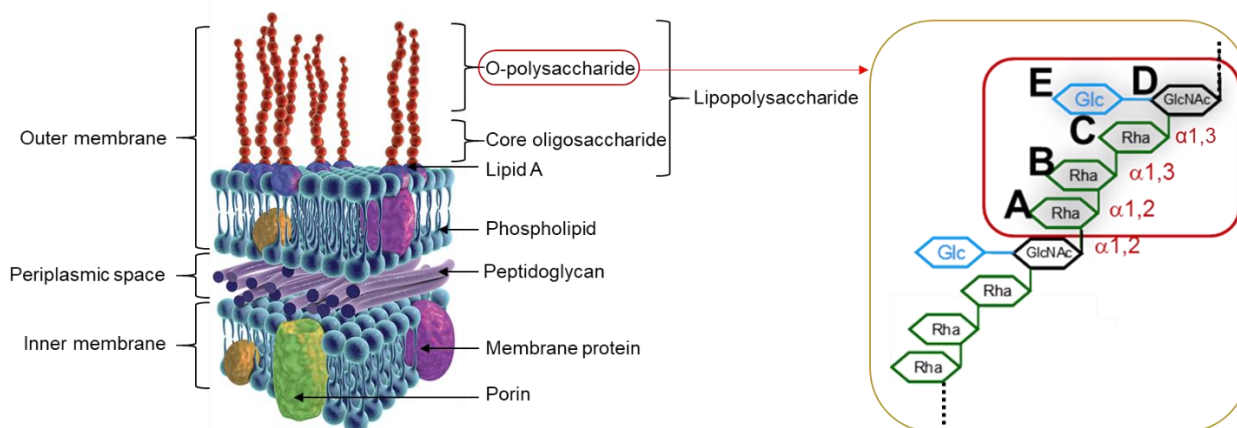


Figure 5: Illustration of the Gram negative bacteria membrane composition. From Chemical and Engineering News, <https://cen.acs.org/content/cen/articles/95/i11/Dual-therapy-first-weakens-kills.html>. Volume 95 Issue 11 | p. 5 | News of The Week. Issue Date: March 13, 2017 | Web Date: March 9, 2017. The repeating unit of the lipopolysaccharide of *Shigella flexneri* is also represented schematically (ABCDE).

The LPS biosynthesis is not fully understood yet. It can occur through two different pathways, namely Wzy polymerase-dependent and ABC transporter-dependent pathways (Kalynychn et al. 2014). Each involve enzymatic cascades. In the Wzy pathway, the repeating O-unit is assembled with dedicated glycosyltransferases in the cytoplasm from lipid-sugars anchored donors. After transport by a flipase, elongation into a polysaccharide occurs in the periplasm thanks to a membrane-bound 'Wzy' polymerase. Length is controlled by another enzyme, 'Wzz'. It is then linked to the lipid A-core oligosaccharide component by a ligase before being transported at the outer surface of the membrane using several transporters that act sequentially.

The ABC transporter-dependent pathway begins by the full assembly of the polysaccharidic part in the cytosol by specific glycosyltransferases, again using lipid-sugar anchored donors. The fully-grown polysaccharide is then transported in the periplasm thanks to the 'ABC' ATP-dependent transporter, and the mature O-antigen is linked to the lipid A-core oligosaccharide thanks to a ligase. Finally, the same transporters than for the Wzy pathway transport the LPS at the outer surface of the membrane.

Specific glucosylation of the LPS of *S. flexneri* is initiated in the cytoplasm by conversion of UDP- α -glucose into the lipid-sugar undecaprenyl- β -glucose (UndP- β -glc) by the GtrB enzyme. UndP- β -glc is then transported into the periplasm by the GtrA enzyme and a Gtr(type) glycosyltransferase (one type for each glucosylation position on the LPS) will transfer the glucose from UndP- β -glc onto the LPS which will therefore become serotype-specific. Most Gtr(type), in particular the ones involved in the specific glucosylation of prevalent serotypes 1b and 3a remain unknown (Figure 6) (Knirel et al. 2015).

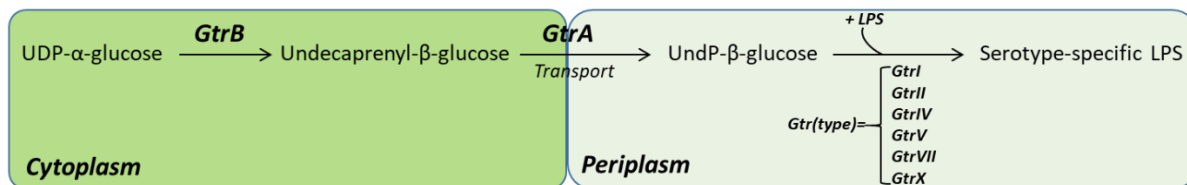


Figure 6: *S. flexneri* enzymatic machinery involved in the specific glucosylation of the LPS

Briefly, the repeating unit of *S. dysenteriae* 1 differs from *S. flexneri* by the substitution of the rhamnosyl C residue by a D-galactose unit, similarly to *S. flexneri* serotype 6. Finally, the repeating unit of *S. sonnei*, differs entirely as it is constituted of N-acetyl-L-fucosamine branched with amino and acetamido groups linked to a 2-amino-2-deoxy-L-altruronic acid (AltUA) bearing an amino group (Figure 7).

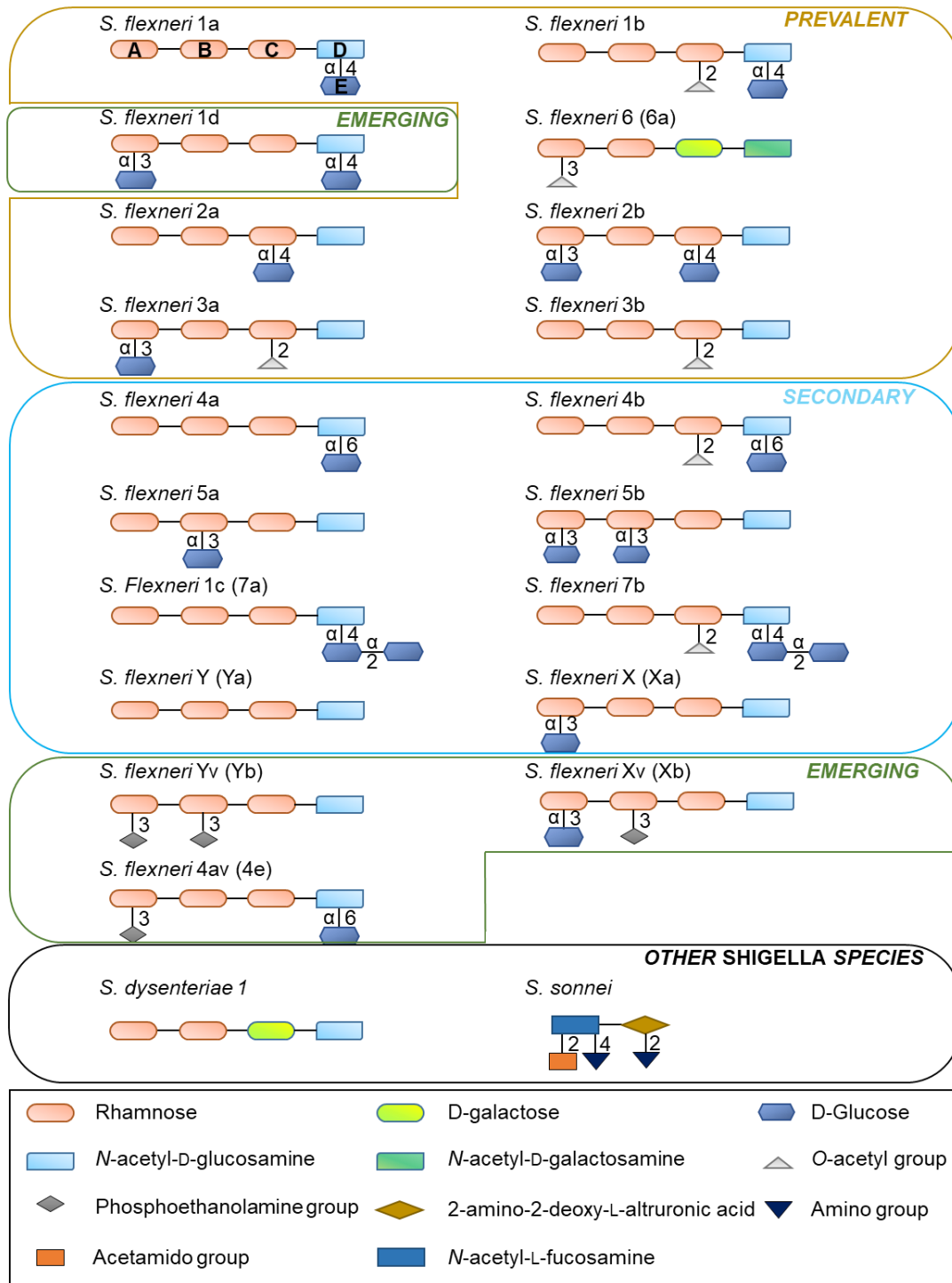


Figure 7: Illustration of the O-antigen structures of main *Shigella flexneri* strains as well as *Shigella dysenteriae* 1 and *S. sonnei* for comparison. Adapted from (Levine et al. 2007) Subserotypes presented in brackets (e.g. *S. flexneri* 1c (**7a**)) are from the newly simplified nomenclature proposed in (Shahnaij et al. 2018). Some O-antigen structures of variants are not represented (see (Knirel et al. 2015): 1a₁, 2a₂, 2b₁, 3a₁, 5a₁, X₁, Y₁, Y₂, Yv₁, 6v (6b), 7a₁, and the O-antigen structure from provisional serotype 4s (Z) (Qiu et al. 2011) is not yet characterized.

Vaccine strategies to fight shigellosis

The World Health Organization published recommendations concerning the development of target vaccine candidates against *Shigella*, and to ensure broad coverage against the shigellosis-causing bacteria, a multi-serotype approach is one of the preconized strategies, whether based on the use of a live attenuated strain or synthetic glycoconjugates. Vaccine strategies for the fight against shigellosis to date have been recently reviewed (Barel and Mulard 2019). Nowadays, the most commonly used vaccination method is based on the use of live-attenuated bacterial strains. Inactivation of the bacteria can be performed by removing the bacterial virulence factors, preventing it to infect the host or to produce toxins (Plotkin 2005; Levine et al. 2007).

The first vaccine candidate to protect against shigellosis was developed in 1971 (SmD *Shigella* vaccine, (Mel et al. 1971)). However, it was shown that the choice of the parent strain in live oral vaccines had a tremendous impact on the ability of the formulation to trigger a sufficient immune response. Later, a same mutation (*aroD*) performed in two different strains led to two vaccine candidates: SFL124, (Li et al. 1993) and SFL1070 (Kärnell et al. 1995) with different toxicity attenuation. But the lack of reproducibility led to the investigation of other attenuating mutations. For example, deletion of the genes encoding enterotoxins ShET1 and ShET2 of *Shigella* allowed to develop live oral vaccine candidates (vaccines CVD 1204 and 1208, (Kotloff et al. 2004)) which were evaluated in a phase I clinical trial. Finally, the Center for Vaccine Development of the University of Maryland (CVD) has carried out several clinical trials of live oral vaccine candidates targeting prevalent *S. flexneri* 2a serotype (Kotloff et al. 1996, 2000). In addition, this research team also proposed a design for a pentavalent live oral vaccine which should provide protection against at least *S. dysenteriae* 1, *S. sonnei* and three serotypes of *S. flexneri*: *Sf*2a, *Sf*6 (6a) and *Sf*3a. The rationale of the devised CVD vaccine relies on the fact that *S. flexneri* O-antigens representative of *Sf*2a and *Sf*3a cross-react with a total of 10 other (sub)serotypes, and elicited a broad immune response in a guinea pig (Noriega et al. 1999; Levine et al. 2007). In the meantime, serotype 6 is characterized by a different repeating unit containing a D-galactose residue and an *N*-acetyl-D-galactosamine residue in place of L-rhamnose and *N*-acetyl-D-glucosamine residues, close to the antigenic repeat of *S. dysenteriae* (Figure 7). To develop a vaccine enabling broad protection, *Sf*6 should thus be considered as it does not cross-react with other O-antigen epitopes. Adding *S. dysenteriae* and *S. sonnei* to the formulation would thus bring protection against the major shigellosis-causing bacteria (Levine et al. 2007).

However in case of shigellosis, the use of live-attenuated strains has proven costly and has been hampered by inconsistent batch-to-batch reproducibility. Extracting only the LPS part from the strain was attempted, but the use of full-size LPS is not possible owing to the toxicity of the molecule, in particular the lipid part, making the LPS detoxification step mandatory (Barel

and Mulard 2019). However, the detoxified LPS, having a low molecular weight, lacks immunogenicity (Polotsky et al. 1994).

Thus, an alternative lies in the use of glycoconjugate vaccines (Phalipon et al. 2006) containing (i) the antigenic moieties (O-Ag), which offers a better control on the process with no risk of microbial contaminations compared to the use of live-attenuated strains, coupled to (ii) a carrier protein that allows to develop a T-cell dependent, long-term immune memory thanks to its high molecular weight (Adamo 2017).

Moreover, multi-serotype glycoconjugate vaccines can be produced, made of various antigenic oligosaccharides specific for different strains (Levine et al. 2007). The choice of the linker and the carrier protein, as well as the degree of conjugation are other key determinants in the design of this kind of glycoconjugate candidates (Berti and Adamo 2018; Micoli et al. 2018a), but they will not be detailed here.

In most glycoconjugate vaccines currently released on the market, reviewed in (Micoli et al. 2018b), the oligosaccharide components are derived from polysaccharides extracted from the corresponding pathogenic strains, leading to a polydisperse formulation. One existing example is MenHibrix® (GSK), a glycoconjugate vaccine widely used and protecting against 2 serotypes of *Neisseria meningitidis*, and *Hemophilus influenza* type b (McCarthy et al. 2018), responsible for meningitis, and various infectious diseases, respectively. Thus, as the O-antigen of *Shigella spp.* was shown to be a key component for protection against the bacteria (Levine et al. 2007), their use in vaccine strategies have been considered. Therefore a parenteral (*i.e.* injection into a muscle or subcutaneous tissue as opposed to oral or nasal routes) conjugate against *Sf2a* obtained from detoxified LPS extracted from the strain coupled to a carrier protein was investigated and the corresponding vaccine candidate (pmLPS-rEPA) was evaluated in a phase III trial in Israel (Cohen et al. 1996).

However, in the latest generation of glycoconjugate vaccines, the O-Ag part tend to become fully synthetic instead of extracted from natural strains, as it offers a better product characterization with chemically defined O-Ag and a control of the oligosaccharide size, allowing to better understand and control the immune response (Adamo 2017). The first synthetic glycoconjugate vaccine (Quimi-Hib) was against *Hemophilus influenza* type b and was approved in Cuba in 2004, with 34 million doses were distributed in several countries since then (Adamo 2017). This vaccine strategy using synthetic O-Ag is considered as very promising for the fight against *Shigella spp.* (Chen and Kotloff 2016). It is important to keep in mind that a vaccine targeting *Shigella spp.* should remain a low-cost solution for developing countries (Phalipon et al. 2006). Fortunately, chemical synthesis of carbohydrates has become more economically attractive at the industrial scale with the advent of relevant technologies such as solid-phase synthesis, automated synthesis and one-pot reactions (see part I). Current investigations for the development of glycoconjugate vaccines, thoroughly reviewed in (Micoli et al. 2018b), use mostly chemical approaches for the synthesis of the oligosaccharide

component. However, chemo-enzymatic strategies can also be envisioned, as they can offer promising prospects for efficient and convergent synthetic routes, as already discussed earlier in this introduction (see part I). In that respect, the group of Chemistry of Biomolecules led by Laurence Mulard at Pasteur Institute initiated a collaboration with our team at TBI – INSA Toulouse almost 15 years ago with the goal of developing relevant chemo-enzymatic pathways for the synthesis of antigenic oligosaccharides that could enter the composition of glycoconjugate vaccine candidates against the most prevalent serotypes of *S. flexneri* (2a, 3a, 1b). This work was done within the frame of three successive research projects funded by the ANR, namely OptiGluc (2005-2008), GlucoDesign (2009-2012) and more recently CarbUniVax (2015-2019) in which is integrated the current PhD work.

Highlights on the prior work done within the frame of TBI-Pasteur Institute collaboration

One of the main roadblock when considering a fully chemical synthesis of the *Shigella* antigenic oligosaccharide resides in the regioselective 1,2-*cis* glucosylation of the L-rhamnosyl or *N*-acetyl-D-glucosaminyl residues. Despite notable improvements in chemical glucosylation (described in (Salamone et al. 2015a)), there are still drawbacks such as i) the obtention of racemic α/β stereoisomer mixtures, ii) the increasing donor structural complexity and iii) the tedious manipulation of protecting groups that lead to case-to-case consideration for each serotype-specific glucosylation pattern.

These difficulties were encountered when the Pasteur Institute developed for the first time a fully chemical pathway for the synthesis of *Sf2a* representative pentadecasaccharide (van der Put et al. 2016), mimicking the natural antigen and shown to be the minimal unit that elicits an immune response in mice (Phalipon et al. 2009). This latter was then coupled to a carrier protein (tetanus toxoid) and the corresponding vaccine candidate SF2a-TT15 was further evaluated in a phase I study (Cohen et al. 2017).

In parallel, novel chemo-enzymatic strategies to access *Sf2a*, *Sf1a/1b* haptens were explored to circumvent some of the chemical limitations, and above all this glucosylation step. These methods involved an early-stage enzymatic α -D-glucosylation of building blocks that can be mono or disaccharides (mimicking the *S. flexneri* backbone repeat) carefully designed with light protecting groups compatible with enzyme active sites as well as subsequent steps of chemical assembly and elongation (Figure 8). Indeed, upon glucosylation step, these building blocks have to be chemically converted into pentasaccharides, themselves further elongated into a full pentadecasaccharide hapten after manipulation of the protecting groups (e.g. deacetylation...).

As previously presented, *in vivo*, the enzymatic machinery from *Shigella* involved in the LPS biosynthesis are membrane-bound glycosyltransferases acting in cascade which use the

poorly available lipid-sugar donor Und- β -glc as substrate (Allison and Verma 2000). Their biochemical characterization is extremely limited as these enzymes are difficult to isolate and no crystallographic structure is available, thus limiting rational engineering efforts that would be necessary to adapt these enzymes to an exogenous protected acceptor. Moreover, no other carbohydrate-active enzyme has been reported so far in literature as being able to catalyze the transfer of a glucosyl moiety onto the LPS of *S. flexneri*. The utilization of UDP-glucose dependent glycosyltransferases (UGT or GT) would require the use of expensive nucleotide-sugar UDP-glucose as substrate, not suitable for the development of a low cost process which is a pre-requisite. Finally, lack of knowledge on glycoside phosphorylases is also an impediment that prevented their use.

Consequently, alternative enzyme candidates have to be identified or engineered. Our team at TBI is a specialist of α -retaining α -transglucosylases that use readily-available and cheap sucrose as glucosyl donor. These enzymes, which belong to the family 13 and 70 of Glycoside-Hydrolases, have also the advantage of displaying a tremendous promiscuity toward a wide range of exogenous acceptor substrates (lipids, flavonoids, carbohydrates, polyols, ...) thanks to their peculiar three-dimensional structure and very exposed active site (Daudé et al. 2014; Malbert et al. 2018). These enzymes, presented in part III of this bibliography survey, could thus be of utmost interest to attempt regioselective glucosylation of more complex molecules, such as the lightly-protected different building blocks developed in the three successive ANR projects.

During the OptiGluc project (2005-2008), the first strategy envisioned was to glucosylate rhamnose and/or N-acetyl-D-glucosamine building blocks at relevant positions (4_D , 3_A , 4_A), representative of *S. flexneri* serotypes 1b, 3a, 2a, respectively (Champion et al. 2009b) using the collection of α -transglucosylases available at TBI or engineered on purpose (Figure 8).

Briefly, the GH13 amylosucrase from *Neisseria polysaccharea* (*NpAS*), the GH70 dextranucrase- S_2 (*DSR-S₂*) and the GH70 alternansucrase (*ASR*) were shown to glucosylate the L-rhamnose with a conversion of 6.5%, 10% and 6%, respectively, using sucrose as glucosyl donor substrate. In addition to the synthesis of high molecular weight polymers of glucose, the natural reaction catalyzed by these enzymes, they also produced the α -D-glucopyranosyl-(1 \rightarrow 1)- β -L-rhamnopyranoside, which is not representative of any *S. flexneri* serotype. To prevent glucosylation at the anomeric position of the L-rhamnose, the methyl L-rhamnose (α -L-RhapOMe), carrying a blocked protecting group at C1 was further evaluated. Novel mono-glucosylated products were thus obtained using for instance the *ASR* that yielded 17% of acceptor conversion, forming α -D-Glcp-(1 \rightarrow 3)- α -L-RhapOMe and α -D-Glcp-(1 \rightarrow 4)- α -L-RhapOMe, both products of interest as being representative of *S. flexneri* serotypes 3a, 5a, and 2a, respectively.

The potential of the chemo-enzymatic approach was further exemplified with the glucosylation of N-acetyl-D-glucosamine, achieved in high yield using *DSR-S₃* (64%) while *NpAS* achieved 6% conversion. Two products were formed and further characterized as being the α -D-

glucopyranosyl-(1→4)-*N*-acetyl-D-glucosamine and the α -D-glucopyranosyl-(1→6)-*N*-acetyl-D-glucosamine, both relevant as being representative of *S. flexneri* serotypes 1a/1b and 4a/4b, respectively.

To further improve these conversion yields and extend further the range of accessible glucosylated products from α -L-RhapOMe and D-GlcpNAc, a (semi-)rational engineering strategy was applied to *NpAS*, the sole α -transglucosylase of GH13/GH70 families of known three-dimensional structure at the time of the study (Champion et al. 2009a, 2009b). A focused library of 133 single mutants was constructed targeting 7 residues of the +1 acceptor subsite in the active site located on flexible loops and identified from docking studies. Upon screening, a mutant I228Y revealed a new specificity towards α -L-RhapOMe, with remarkable 44% conversion, and forming the α -D-Glcp-(1→3)- α -L-RhapOMe product. On the same note, mutants F290D and F290K showed a high conversion of D-GlcpNAc leading to the formation of more than 90% of α -D-Glcp-(1→4)- α -D-GlcpNAc. Mutant F290K was further evaluated in the presence of a lightly protected monosaccharide carrying an O-allyl group at the anomeric position, (α -D-GlcpNAc-O-All). A conversion of 78% of this latter was found, leading to the formation of α -D-Glcp-(1→4)- α -D-GlcpNAc-OAll, but lost some of its catalytic efficiency.

To explore further the combinatorial effect of these mutations, a mutant library was constructed by pairwise combinations of mutations on vicinal residues (I228, F229, A289 and F290) (Champion et al. 2012). Such mutations occurring simultaneously on tandem residues are quite rare in Nature but could be of particular interest for the accommodation of non-natural acceptor molecules. The library screening yielded three mutants, A289P-F290L, A289P-F290C and A289P-F290I that were shown to glucosylate monosaccharide D-GlcpNAc with a 100% yield. When using monosaccharide α -D-GlcpNAc-O-All, the conversion reached up to 60% product yield. In overall, the catalytic efficiency was improved by 100-350 fold compared to that of F290K and the initial velocity transglucosylation was increased by up to 10 fold.

These studies were the first reports of enzyme engineering strategies applied to the development of transglycosylases endowed with novel or improved substrate specificity for the synthesis of lightly protected carbohydrate precursors.

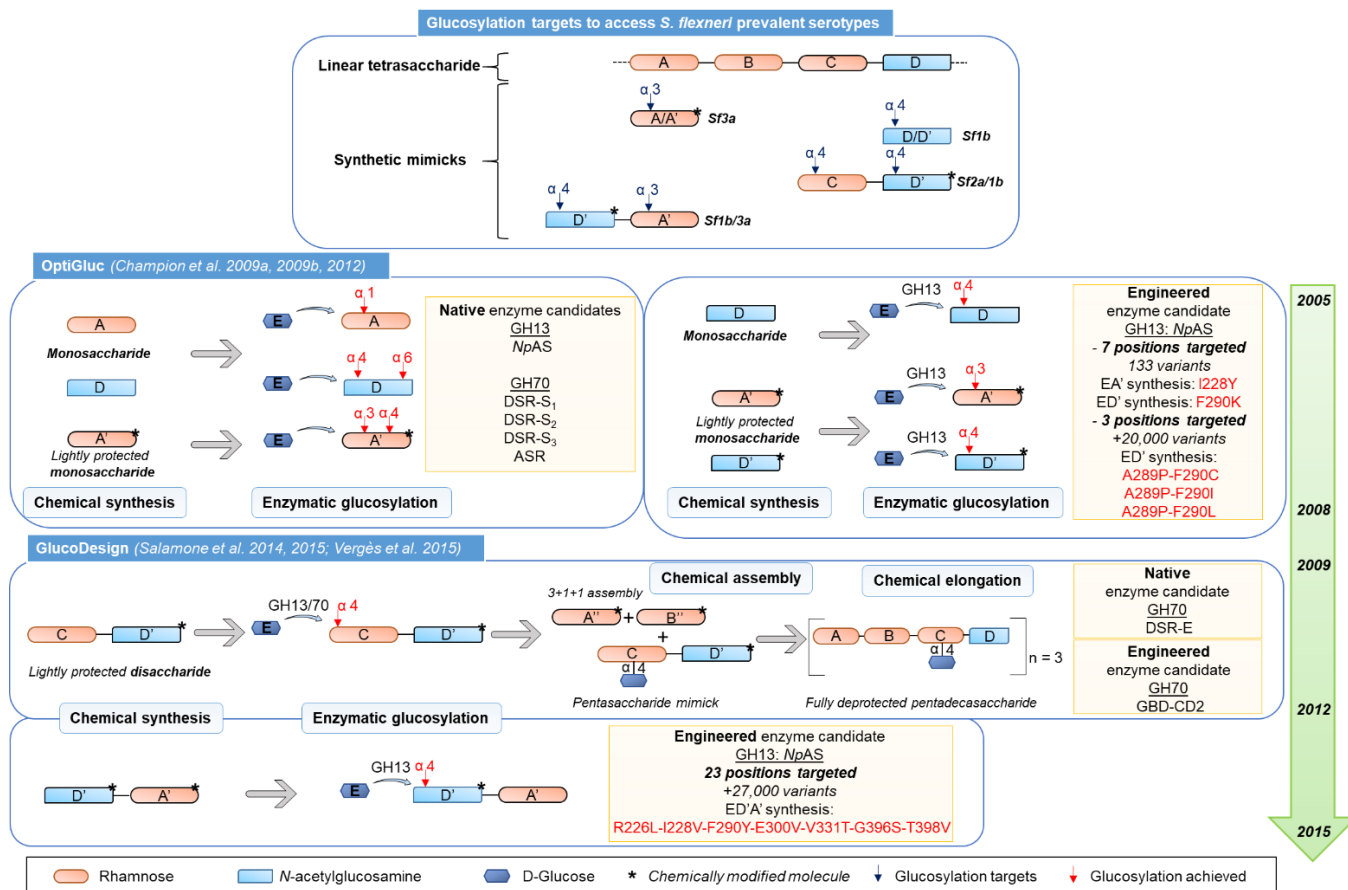
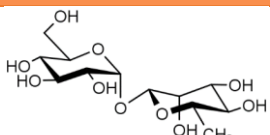
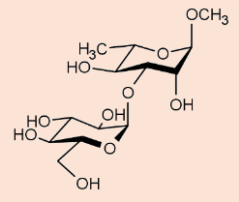
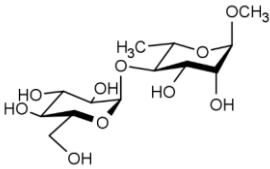
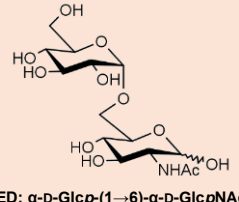
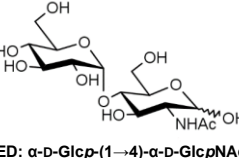
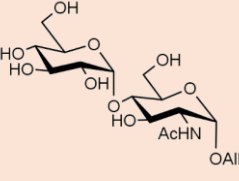
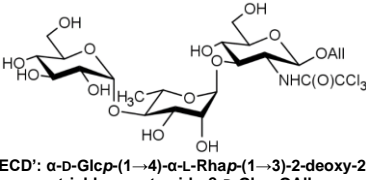
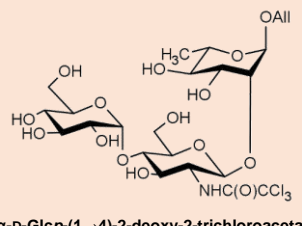


Figure 8: Summary of the previous ANR-funded projects OptiGluc and GlucoDesign. A timeframe is represented on the right. The represented acceptor molecules are either common monosaccharides (A, B, C: rhamnose and D: *N*-acetyl-D-glucosamine) or chemically modified mono, di, or tetrasaccharide(s) (A': methyl-rhamnose, D': *N*-trichloroacetyl-D-glucosamine, A'' and B'': rhamnosyl trichloroacetimidate intermediate, CD': α -L-Rhap-(1→3)-2-deoxy-2-trichloroacetamido- β -D-Glcp-OAll, D'A': 2-deoxy-2-trichloroacetamido- β -D-Glcp-(1→2)- α -L-Rhap-OAll.. The chemical nature of each saccharide is shown in Table 1. Glucosylation was achieved using different enzymes or their variants which are listed in the light yellow frame and also available in Table 1.

Based on these promising results, the aim of the GlucoDesign project (2009-2012) was to develop enzymes able to glucosylate several disaccharides constitutive of the *Sf* backbone repeating unit. The objective was then to integrate the enzymatic step at a later stage of the synthesis and generate a library of modular glyco bricks that could facilitate the assembly of hapten oligosaccharides. The first disaccharide considered for the enzymatic glucosylation was the α -L-Rhap-(1→3)-2-deoxy-2-trichloroacetamido- β -D-Glcp-OAll (named CD'). Among these enzymes, a recently identified bifunctional enzyme called DSR-E, and its truncated form ΔN_{123} -GBD-CD2 (Bozonnet et al. 2002; Fabre et al. 2005) that belongs to the subfamily of branching sucrases (see part III) was able to glucosylate successfully CD'. A collection of native transglucosylases from the GH70 family was evaluated for the first time with non-natural disaccharide acceptors (Salamone et al. 2015a).

Table 1. List of the different mono-glucosylated molecules synthesized during the OptiGluc and GlucoDesign projects. ^aMono-glucosylated product yield is the molar ratio: [Glucosylated acceptor]_{final} / [consumed sucrose]_{final} × 100

Target glucosylated product	<i>S.flexneri</i> serotype glycosylation pattern	Best enzyme / variant identified for glucosylation	Mono-glucosylated product yield (%) ^a	References
 <p>EA: α-D-Glcp-(1→1)-β-L-Rhap</p>	<i>n.a.</i>	<i>Np</i> AS (GH13)	17.5	(Champion et al. 2009b)
 <p>EA': α-D-Glcp-(1→3)-α-L-RhapOMe</p>	3a 5a X	DSR-S ₂ (GH70)	4	(Champion et al. 2009b)
		<i>Np</i> AS I228Y (GH13)	44	(Champion et al. 2009a)
 <p>EA': α-D-Glcp-(1→4)-α-L-RhapOMe</p>	2a	ASR (GH70)	10	(Champion et al. 2009b)
 <p>ED: α-D-Glcp-(1→6)-α-D-GlcpNAc</p>	4a/4b	DSR-S ₃ (GH70)	61	(Champion et al. 2009b)
 <p>ED: α-D-Glcp-(1→4)-α-D-GlcpNAc</p>	1a/1b	<i>Np</i> AS (GH13)	4	(Champion et al. 2009b)
		<i>Np</i> AS F290D/K (GH13)	≈90	(Champion et al. 2009a)
		<i>Np</i> AS A289P-F290I/C/L (GH13)	100	(Champion et al. 2012)
 <p>ED': α-D-Glcp-(1→4)-α-D-GlcpNAc-OAll</p>	1a/1b	<i>Np</i> AS F290K (GH13)	78	(Champion et al. 2009a)
		<i>Np</i> AS A289P-F290I/C/L (GH13)	40-50	(Champion et al. 2012)
 <p>ECD': α-D-Glcp-(1→4)-α-L-Rhap-(1→3)-2-deoxy-2-trichloroacetamido-β-D-Glcp-OAll</p>	2a	ΔN ₁₂₃ -GBD-CD2 (GH70)	94	(Salamone et al. 2015a)
 <p>ED'A': α-D-Glcp-(1→4)-2-deoxy-2-trichloroacetamido-β-D-Glcp-(1→2)-α-L-Rhap-OAll</p>	1a/1b	<i>Np</i> AS (GH13) R226L-I228V-F290Y-E300V-V331T-G396S-T398V	2.2	(Vergès et al. 2015)

The native DSR-E yielded 12% of acceptor conversion, while its truncated form, ΔN_{123} -GBD-CD2, was found to convert 44% of the acceptor. Upon reaction optimization, this yield was increased to 94%, forming majorly α -D-Glcp-(1 \rightarrow 4)- α -L-Rhap-(1 \rightarrow 3)-2-deoxy-2-trichloroacetamido- β -D-Glcp-OAll (ECD'). The 4_C glucosylation pattern being representative of *S. flexneri* serotype 2a. Upon g-scale enzymatic production, the ECD' trisaccharide block was used as an acceptor with a chemically synthesized protected disaccharide block [AB] and the 3+2 assembly yielded a *S. flexneri* 2a pentasaccharide mimick, which was further elongated high yield into a decasaccharide with (Salamone et al. 2015b) and then a pentadecasaccharide (Salamone et al. 2015a). This chemo-enzymatic pathway offers an advantageous alternative to the chemical route in terms of number of steps and overall yield to access Sf2a haptens, currently being used in the vaccine candidate SF2a-TT15 developed the by Pasteur Institute. Moreover, a second disaccharide was designed to target enzymatic glucosylation at position 3_{A'} and 4_{D'}, representative of serotypes Sf3a and Sf1a/1b, the 2-deoxy-2-trichloroacetamido- β -D-Glcp-(1 \rightarrow 2)- α -L-Rhap-OAll (called D'A'). As no native α -transglucosylase of the collection was found able to glucosylate this molecule, a computer-aided engineering strategy was undertaken to re-design the deep and buried active site of GH13 NpAS and adapt it to the targeted product (Vergès et al. 2015). Using a combination of molecular docking and computational protein design techniques, *in silico* mutations were explored at 23 selected amino acid positions from the active site, leading to a set of 1515 sequences, selected on the basis their affinity score and stability. At this stage, a combinatorial library of variants was further designed, taking into account amino acid frequency and pairwise residue correlations observed in the 1515 sequences. Upon screening of this library, a 7-mutation variant (R226L-I228V-F290Y-E300V-V331T-G396S-T398V) was identified to yield in low amounts (2.2%) the mono-glucosylated product ED'A', having a glucosylation at position 4_{D'}, representative of serotypes Sf 1a/1b. In spite of its low efficiency, this mutant is the first reported so far for this catalytic reaction and could be considered as a potential starting point for further directed evolution. Now that the context of this thesis, tightly related to the fight against shigellosis, was presented, the sucrose active enzymes that were previously mentioned in part I and part II, and more particularly those of family GH70 that were studied during this project will be reviewed in more details.

PART III: The α -transglucosylases from GH70 family

Classification of sucrose-active α -transglucosylases

Enzymes listed in the CAZy database are responsible for degradation, modification or creation of glycosidic bonds of various regio- and stereo-specificities (Lombard et al. 2014). They are classified in five major classes, namely the GlycosylTransferases, Glycoside Hydrolases, Polysaccharide Lyases, Carbohydrate Esterases and Auxiliary Activities, according to their sequence homology and taking into account structural characteristics of their active site and mechanistic features. This allows more accurate and reliable clustering of enzymes that could otherwise not be related in Enzyme Commission (EC) classification, which relies mostly upon substrate specificity and catalyzed reactions.

The sucrose-active α -transglucosylases are found in the so-called GH-H clan that gathers evolutionary-related members of families 13, 70 and 77 of Glycoside Hydrolases. However, most of the α -transglucosylases belong to the GH70 family, with the exception of amylosucrases from GH13 family, a class of catalysts that mainly contains enzymes involved in starch transformation. However, instead of degrading starch, amylosucrases use sucrose to synthesize an amylose-like polymer.

The focus of this work has been set exclusively on sucrose-active transglucosylases from GH70 family, described hereafter (EC 2.4.1, glycosyltransferase). As of June 2019, 620 sequences are listed in the GH70 family, for which only 66 enzymes are biochemically characterized, and 7 crystallographic structures are available (CAZy database). Thus, 89% of reported sequences are still putative α -transglucosylases, because these proteins are very large enzymes (molecular weight ranging from 150 to 350 kDa), limiting the recombinant expression in heterologous organisms such as in *E. coli* for a characterization at molecular level. These enzymes are naturally produced by lactic acid bacteria, 97% of them by *Streptococcus*, *Lactobacillus*, *Leuconostoc*, and *Weissella* species (Gangoiti et al. 2018b), and a sole genomic sequence generally contain between 3 to 6 GH70 genes. Genus repartition of GH70 is well reviewed in (Meng et al. 2016). Among the 66 characterized enzymes, 59 are so-called glucansucrases (GS), 6 are glucanotransferases and only 1 is classified as being a branching sucrose (BRS) (Figure 9).

Reactions catalyzed by GH70 enzymes

Unlike Leloir glycosyltransferases that use activated nucleotide-sugar as donor substrate to catalyze reactions of glycosylation, glucansucrases and branching sucrases use the high energy from sucrose (α -D-Glc ρ -(1,2)- β -D-Fru ρ) bond to catalyze the transfer of a glucosyl moiety onto an acceptor molecule.

Regarding glucansucrases, the first acceptor will be water or sucrose itself. Successively, the transfer will occur onto glucose or growing gluco-oligosaccharide chains, and the main final products will be polymers composed of glucosyl units linked through various types of α -osidic linkages. In Nature, dextran contains >50% α 1,6 linkages and branches linked by α 1,2, α 1,3 or α 1,4 glucosidic bonds, reuteran contains around 50% of α 1,4 (possibly more) and α 1,6 linkages, alternan presents a clear alternation of α 1,3 and α 1,6 linkages, and finally mutan presents >50% α 1,3 linkages (Monsan et al. 2010). The fructosyl moiety of sucrose is released as by-product, and small amounts of sucrose isomers can also be formed (less than 10% of glucosyl units transferred onto fructose). The *in vivo* role of these α -glucans has not been investigated in details yet. Through formation of biofilms, they could favor colonization of the environment by their producing-bacteria thanks to adhesion properties (Brooker 1976). For example, mutan is known to favor adherence of dental plaque forming bacteria at the surface of teeth, such as *Streptococcus mutans* (Tamesada et al. 2004). These biofilms could also have a protective role against environmental threats such as desiccation, cold temperatures or attack by phages or other bacteria (Flemming and Wingender 2010).

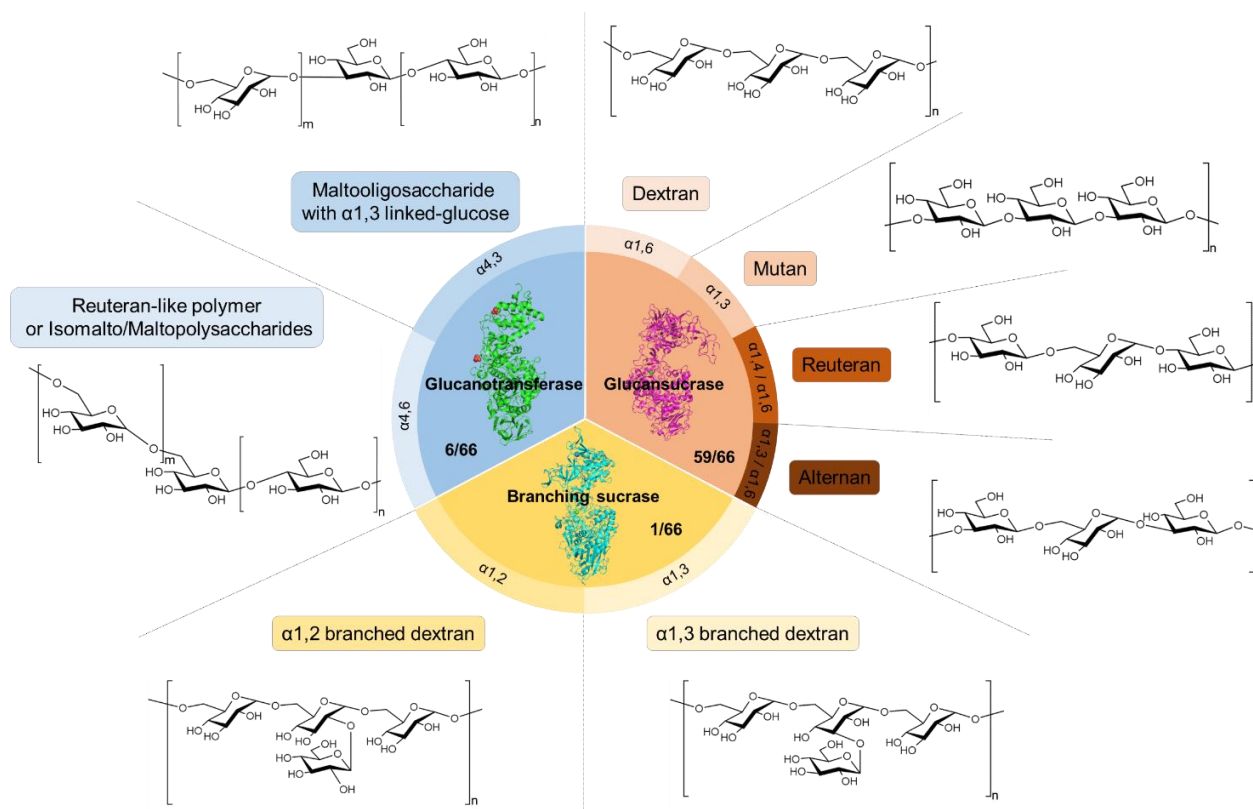


Figure 9: Overview of the products generated by the different subfamilies of the GH70 enzymes according to their specificity, in presence of sole sucrose for glucansucrases, dextran and sucrose for branching sucrases, amylose or maltooligosaccharide for glucanotransferases. Branching pattern degree for glucanotransferases and branching ratio for branching sucrases may vary according to the enzyme and the reaction conditions. Some representative three-dimensional structures are represented for each subfamily: PDB ID: 5LFC (DSR-M, dextran sucrase), 3TTQ (Δ N₁₂₃-GBD-CD2, α 1,2-branching sucrase), 5JBD (GTFB, α 4,6-glucanotransferase). The number of characterized enzymes listed in the CAZy database as of March 2019 is indicated for each subfamily.

The branching sucrases also use sucrose as glucosyl donor substrate but are not able to produce polymers. In vivo, these enzymes catalyze the formation of glucosyl branches on a linear dextran, with either α 1,2 or α 1,3 linkages, depending on their regioselectivity. The first branching sucrose discovered –and to date the only one reported in the CAZy database- was obtained by engineering a bifunctional dextran sucrose named DSR-E (Bozonnet et al. 2002). This DSR-E enzyme contains two catalytic domains, the first one (CD1) being involved in the formation of a high molar mass dextran while the second one (CD2) is involved in branching activity, *i.e.* the grafting of glucosyl units onto dextran chains with α 1,2 osidic linkages (Fabre et al. 2005). The two catalytic domains are bridged by a specific domain, named domain V, and shown for years to be involved in glucan binding (Glucan Binding Domain, GBD) (Wren et al. 1991; Moulis et al. 2006b). The truncation of the parental enzyme into the GBD-CD2 form led to an enzyme with no polymerase activity but displaying a branching activity only. Without dextran added in the reaction medium, GBD-CD2 catalyzed almost exclusively sucrose hydrolysis and formation of one sucrose isomer, leucrose. Since then, other branching sucrases were discovered in the genomic sequences of *Leuconostoc (Ln)* species (*Ln. citreum* for BRS-A, BRS-B and GBD-CD2, *Ln. fallax* for BRS-C, *Ln. mesenteroides* for BRS-E) as well as in *Lactobacillus kunkei* (BRS-D, BRS-G (also named (GTF-Z)-CD2 in (Meng et al. 2018)) and *Fructobacillus tropeoli* (BRS-F). Most of them are native enzymes possessing only one catalytic domain (e.g. BRS-A, BRS-B, BRS-C, BRS-D, BRS-E, BRS-F) while some are engineered from a bifunctional enzyme (e.g. BRS-G or GBD-CD2). All of them are present in the in-house collection of TBland are listed in Table 2.

Finally, the α 4,6 and α 4,3 glucanotransferases are disproportionating enzymes that were recently discovered and that do not use sucrose as substrate. They use starch or maltooligosaccharide as donor substrates, cleave some α 1-4 linkage at the non-reducing end of a chain and transfer whole segments of glucosyl units at the non-reducing end of another chain via the introduction of α 1,6 or α 1,3 osidic bounds depending on the enzyme specificity (Gangoiti et al. 2016; Kralj et al. 2011).

The α -glucans produced by these different GH70 enzymes differ in their structure -due to the variety of α 1,2, α 1,3, α 1,4, and α 1,6 linkages and the different branching patterns- and sizes, enabling access to various characteristics (e.g. physico-chemical property, biological activity...), of interest for food, feed, cosmetics or pharmaceutical industries (Gangoiti et al. 2018a). The characteristics of the products synthesized by these enzymes are summarized in Table 2.

Table 2. Characteristics of polymers synthesized (when available) by some characterized glucansucrases and branching sucrases (completed from (Gangoiti et al. 2018b)). ^aunpublished data, ^bto be published, ^c: (GTF-Z)-CD2 is named BRS-G in this thesis. n.a.: not applicable (branching sucrose). For branching sucrases, linkages correspond to linkages found in the branching product, when linear dextran (100% α 1,6 linkages) is used as an acceptor.

α -glucan	Organism	Enzyme	Osidic linkage specificity				Max polymer weight (g.mol ⁻¹)	References
			α 1,2	α 1,3	α 1,4	α 1,6		
Dextran	<i>Leuconostoc citreum</i> NRRL B-1299	DSR-E	5	10	3	81		(Fabre et al. 2005)
	<i>Leuconostoc citreum</i> NRRL B-1299	DSR-A		15		85		(Monchois et al. 1998a)
	<i>Leuconostoc citreum</i> NRRL B-1299	DSR-B		5		95		(Côté and Robyt 1982)
	<i>Leuconostoc citreum</i> NRRL B-1299	DSR-DP				100	10 ⁶	(Passerini et al. 2015)
	<i>Leuconostoc citreum</i> NRRL B-1299	DSR-M Δ 1		<1		99	10 ⁸	(Claverie et al. 2017)
	<i>Leuconostoc citreum</i> NRRL B-1299	DSR-M Δ 5		<1		99	10 ³	(Claverie et al. 2017)
	<i>Leuconostoc mesenteroides</i> NRRL-B512F	DSR-S vardel Δ 4N		5		95	10 ⁷	(Moulis et al. 2006a)
	<i>Oenococcus kitaharae</i> DSM 17330	DSR-OK		2		98	>10 ⁹	(Vuillemin et al. 2018)
	<i>Weissella cibaria</i> CMU	DSR-WC				100		(Kang et al. 2009)
	<i>Leuconostoc citreum</i> B/110-1-2	DSR-F		6	1	93	10 ⁶	(Fraga Vidal et al. 2011)
	<i>Weissella confusa</i> VTT E-90392	WcE392-rDSR		3		97	10 ⁷	(Kajala et al. 2015)
	<i>Lactobacillus reuteri</i> 180	GTF-180 Δ N		31		69		(van Leeuwen et al. 2008a)
	<i>Streptococcus mutans</i> GS5	GTF-D		30		70		(Hanada and Kuramitsu 1989)

Branched Dextran	<i>Leuconostoc citreum</i> NRRL B-1299	BRS-A	37			63	<i>n.a.</i>	(Passerini et al. 2015)
	<i>Leuconostoc citreum</i> NRRL B-742	BRS-B Δ1		50		50	<i>n.a.</i>	(Vuillemin et al. 2016)
	<i>Leuconostoc fallax</i> KCTC 3537	BRS-C		^a 35		^a 65	<i>n.a.</i>	(Vuillemin et al. 2016)
	<i>Lactobacillus kunkei</i> EFB6	BRS-D Δ1	^a 35			^a 65	<i>n.a.</i>	(Vuillemin et al. 2016)
	<i>Leuconostoc citreum</i> NRRL B-1299	GBD-CD2	33			67	<i>n.a.</i>	(Fabre et al. 2005)
	<i>Leuconostoc citreum</i> NRRL B-1299	ΔN ₁₂₃ -GBD-CD2	33			67	<i>n.a.</i>	(Brison et al. 2012)
	<i>Leuconostoc mesenteroides</i> KFRI-MG	BRS-E Δ1		^b 23		^b 76	<i>n.a.</i>	^b Severac et al.
	<i>Fructobacillus tropaeoli</i> F214-1	BRS-F	^a 35			^a 65	<i>n.a.</i>	Unpublished data
	<i>Lactobacillus kunkei</i> DSM 12361	(GTF-Z)-CD2 (= BRS-G) ^c		40		60	<i>n.a.</i>	(Meng et al. 2018)
	<i>Leuconostoc mesenteroides</i> BD3749	Gsy		14	23	63	<i>n.a.</i>	(Yan et al. 2018)
Mutan	<i>Streptococcus mutans</i> MT4251	GTF-SI		100				(Fujiwara et al. 1998)
	<i>Leuconostoc mesenteroides</i> NRRL B-1118	DSR-I		50		50		(Côté and Skory 2012)
Alternan	<i>Leuconostoc mesenteroides</i> NRRL B-1355	ASR		44		56	10 ⁶	(Joucla et al. 2006)
Reuteran	<i>Lactobacillus reuteri</i> 121	GTF-A			58	42		(van Leeuwen et al. 2008b)
	<i>Lactobacillus reuteri</i> ATCC 55730	GTF-O			79	21		(Kralj et al. 2005)

Mechanistic insights of the transglucosylation reaction

The sucrose-active α-transglucosylases follow the same double-displacement mechanism as α-retaining glycosidases (Koshland 1953), which involve two catalytic residues, a general acid/base (glutamic acid) and a nucleophile residue (aspartate) together with a transition state stabilizer residue (aspartate). When the sucrose donor is in a favorable position for catalysis (Vujcic-Zagar et al. 2010), a nucleophile attack of the anomeric C1 carbon of the glucosyl moiety of sucrose is performed by the carboxyl oxygen of the nucleophile aspartate

simultaneously to the protonation of the glycosidic oxygen by the acid/base glutamate. The α 1-2 linkage of sucrose is then cleaved to form a β -D-glucosyl-enzyme covalent intermediate, while releasing the fructose moiety (Figure 10).

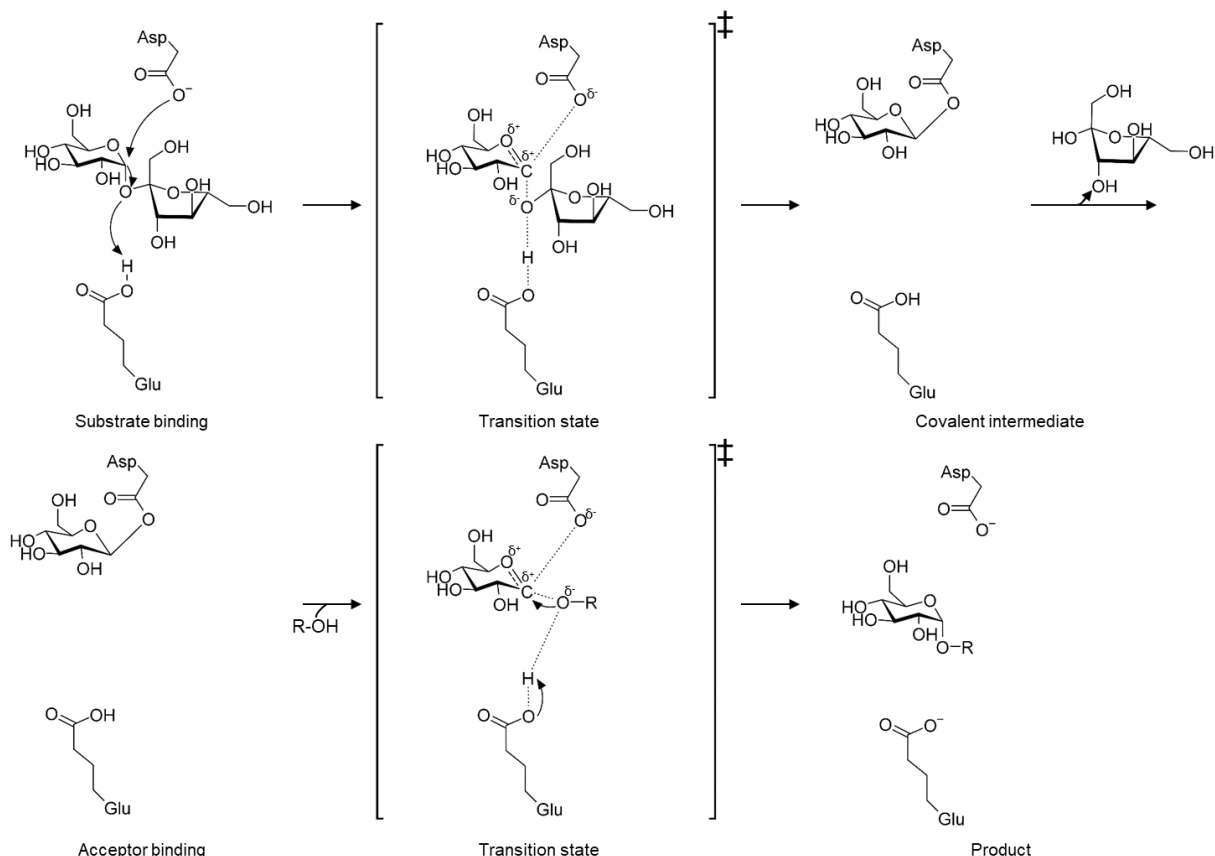


Figure 10: Representation of the Koshland double-displacement mechanism for α -retaining transglucosylases. First step (top) is the hydrolysis of the sucrose bond and second step (down) is the transglycosylation of a hydroxylated acceptor molecule.

The subsequent deglycosylation step is then dependent on the nature of the acceptor substrate. If the acceptor is water, hydrolysis of sucrose happens and free glucose is released in the medium. Then, glucose can play the role of acceptor for a second glucosyl transfer, and so on, growing gluco-oligosaccharides or α -glucan chains can be formed. However, the transfer can also occur onto acceptor molecules present in addition to sucrose (whether the natural acceptor, dextran in the case of branching sucrases; or exogenous hydroxylated acceptor if recognized by the enzyme). The acceptor is activated by protonation of the acid/base glutamate by the reactive hydroxyl of the acceptor, forming a covalent linkage between the glucosyl moiety and the acceptor with retention of the anomeric configuration. The glucosylated acceptor is then released in the medium. When several acceptors are present in the medium, they compete with each other to form various products whose distribution will depend on substrate affinities and reaction kinetics (Figure 11).

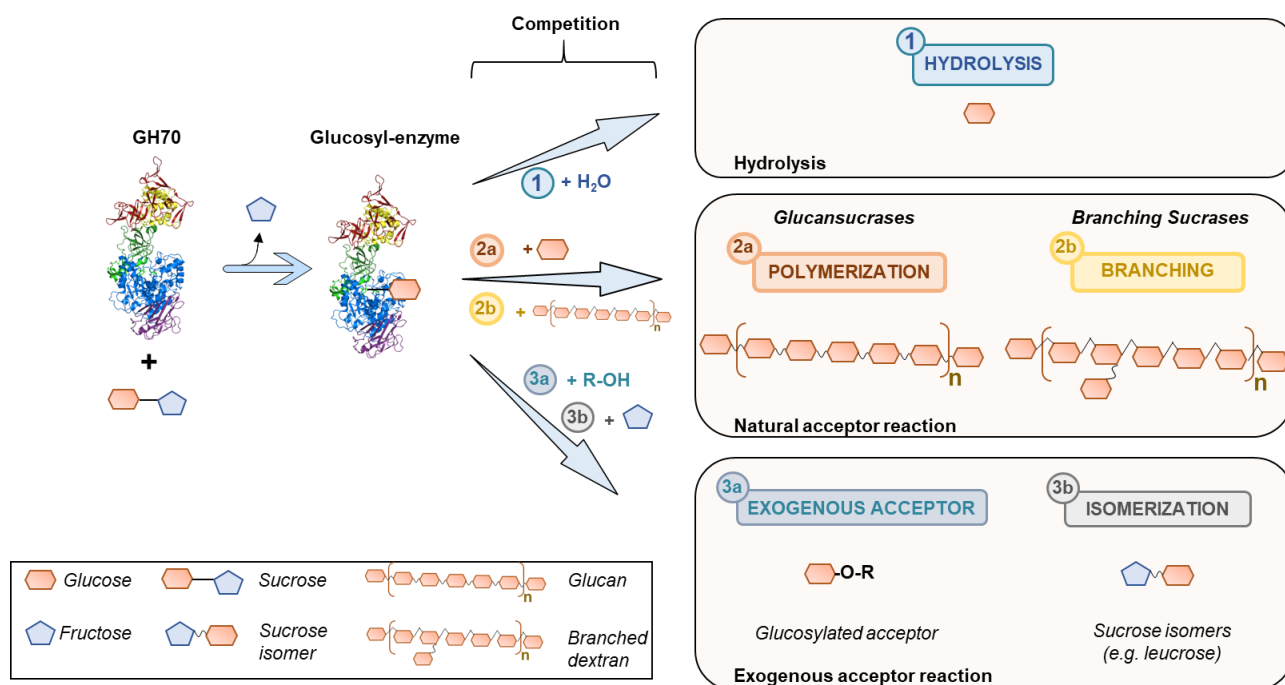


Figure 11: Schematic representation of the different reactions catalyzed by glucansucrases and branching sucrases.

General description of sequence, structure and dynamics features

The first three-dimensional structures of GH70 members were obtained within the last decade. To date, 7 crystallographic structures are available (Table 3, Figure 12A): dextransucrase GTF-180-ΔN (PDB ID: 3KLK, (Vujicic-Zagar et al. 2010)), mutansucrase GTF-SI (PDB ID: 3AIE, (Ito et al. 2011)), branching sucrose ΔN₁₂₃-GBD-CD2 (PDB ID: 3TTQ, (Brison et al. 2012)), reuteransucrase GTF-A ΔN (PDB ID: 4AMC, (Pijning et al. 2012)). More recently, the structure of the dextransucrase DSR-M Δ2 (PDB ID: 5LFC, (Claverie et al. 2017)) and the alternansucrase ASR Δ2 (PDB ID: 6HVG, (Molina et al. 2019)) were determined in our group, as well as the first structure of a glucanotransferase, GTF-B, by Dijkhuizen and colleagues (PDB ID: 5JBD, (Bai et al. 2017)). Of note, there is no crystallographic structure of a full-length GH70 enzyme so far. Available 3D-structures are derived from truncated enzymes, lacking ~200 to 900 residues in their N- and/or C-terminal extremities.

In the domain A, six amino acids are strictly conserved among members of the GH70 family in the N-terminal region of the domain: i) the three amino acids defining the catalytic triad: the nucleophile (aspartate in motif II), the acid/base (glutamate in motif III) and the transition state stabilizer (another aspartate in motif IV) (Vujicic-Zagar et al. 2010); ii) two conserved aspartate (in motif I and II, respectively), and iii) an histidine (in motif IV) (Figure 14). Confirmation of the residues forming the catalytic triad was first shown in 1997, by mutating the nucleophile, the general acid/base or the transition state stabilizer residues and verifying their impact on activity toward sucrose (Monchois et al. 1997). The overall three-dimensional structure of GH70 enzymes is a multi-domain arrangement organized in a U-shape (Figure 12B). As such, amino

acids close in their spatial localization might be very far apart on the protein primary structure. Altogether, there are five structural domains named respectively A, B, C, IV and V. Only domain C corresponds to a single polypeptide segment while domains A, B, IV and V have entangled polypeptide segments at N- and C- terminal extremities. The catalytic domain A of GH70 enzymes is organized as a $(\beta/\alpha)_8$ barrel which is an alternation of eight β -sheets ($\beta 1$ - $\beta 8$) surrounded by eight α -helices ($\alpha 1$ - $\alpha 8$). This domain comprises specific motifs, numbered motifs I to VII (Figure 13).

Table 3. List of X-ray structures determined in the GH70 family. IM = isomaltose, IM3 = isomaltotriose, IM4 = isomaltotetraose, IMM4 = 6⁴- α -D-glucosyl-maltotetraose, M3 = maltotriose, M5 = maltopentaose

Enzyme	PDB ID	Resolution (Å)	Ligands	Ligand location	Reference
GTF-180-ΔN	3KLK	1.65	<i>Apo</i>		(Vujicic-Zagar et al. 2010)
	4AYG	2.00	<i>Apo</i>		
	3HZ3	2.22	Sucrose (2)	Domain A + B	
	3KLL	2.00	Maltose (4)	Domain A + B	
GTF-SI	3AIE	2.10	<i>Apo</i>		(Ito et al. 2011)
	3AIB	3.09	Maltose	Domain A	
	3AIC	3.11	α -acarbose	Domain A	
ΔN₁₂₃-GBD-CD2	3TTQ	1.90	<i>Apo</i>		(Brison et al. 2012, 2016)
	3TTO	3.30	<i>Apo</i>		
	4TTU	2.18	α -glucose + IM	Domain V	
	4TVC	1.85	α -glucose + IM3	Domain V	
	4TVD	2.30	α -glucose + β -glucose (8)	Domain A + V	
GTF-A ΔN	4AMC	3.60	<i>Apo</i>		(Pijning et al. 2012)
DSR-M Δ2	5LFC	3.20	<i>Apo</i>		(Claverie et al. 2017, 2019)
	5NGY	3.70	IM4	Domain V	
	5O8L	3.60	Sucrose	Domain A	
	6HTV	3.90	IM4	Domain A	
ASR Δ2	6HVG	2.80	<i>Apo</i>		(Molina et al. 2019)
GTF-B ΔNΔV	5JBD	1.80	<i>Apo</i>		(Bai et al. 2017)
	5JBE	2.10	maltose + IMM4	Domain A	
	5JBF	2.19	maltose (2) + M3 + M5	Domain A + V	

The overall three-dimensional structure of GH70 enzymes is a multi-domain arrangement organized in a U-shape (Figure 12B). As such, amino acids close in their spatial localization might be very far apart on the protein primary structure. Altogether, there are five structural domains named respectively A, B, C, IV and V. Only domain C corresponds to a single polypeptide segment while domains A, B, IV and V have entangled polypeptide segments at N- and C- terminal extremities. The catalytic domain A of GH70 enzymes is organized as a

(β/α)₈ barrel which is an alternation of eight β -sheets (β 1- β 8) surrounded by eight α -helices (α 1- α 8). This domain comprises specific motifs, numbered motifs I to VII (Figure 13).

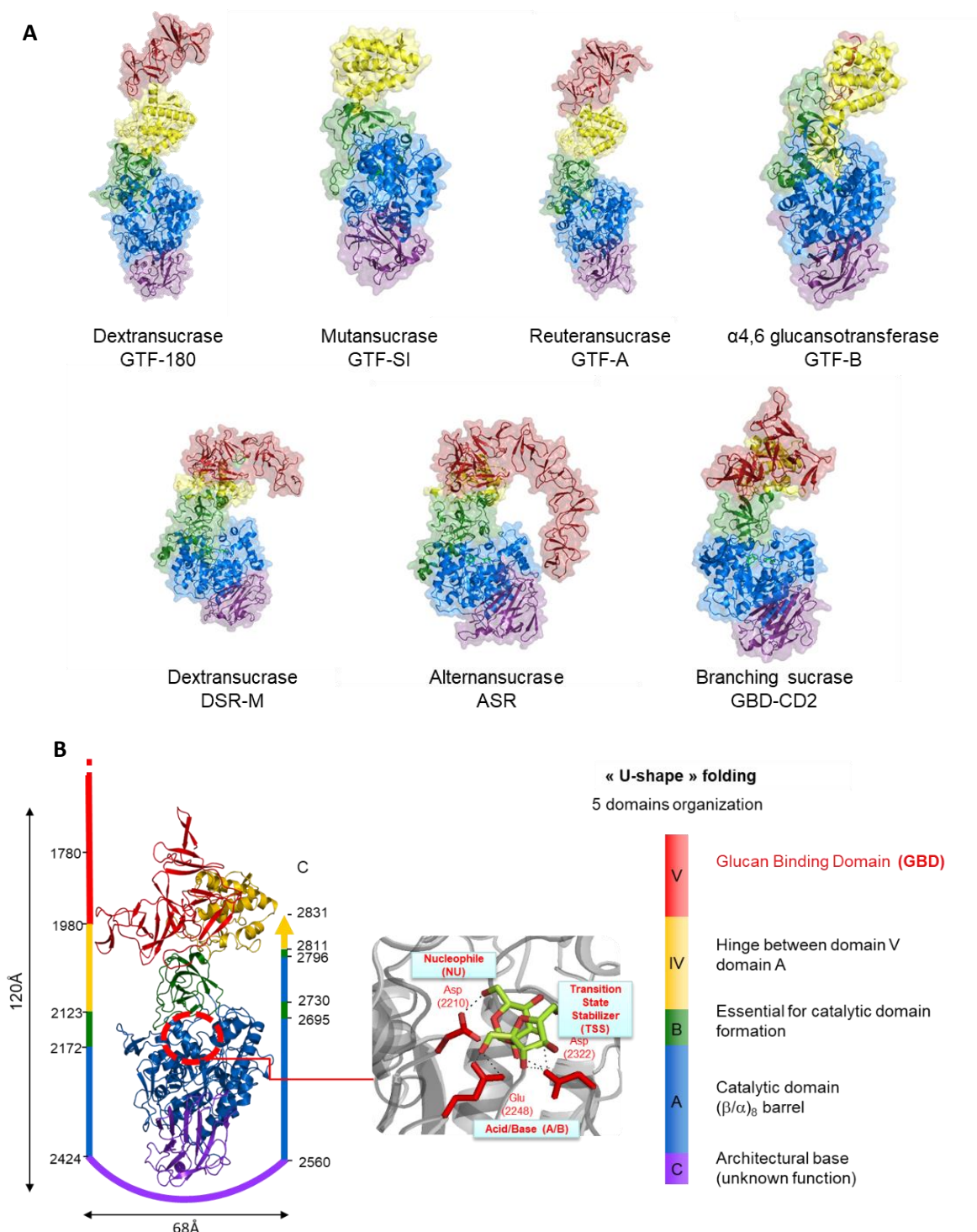


Figure 12: Structural data available on GH70 enzymes.

A/ Crystallographic structures of the 7 GH70 enzymes solved to date, here represented in their *apo* form (PDB ID and references are given in Table 2). Domains are colored as follow: domain V in red, domain IV in yellow, domain B in green, domain A in blue and domain C in purple.

B/ Overall structural organization of the truncated branching sucrose Δ N₁₂₃-GBD-CD2 (PDB ID 3TTQ) in a characteristic multi-domain “U-shape”. Contrarily to other described structures, GBD-CD2 lacks a C-terminal domain V polypeptide. A view of the catalytic triad (nucleophile D2210, acid/base E2248 and transition state stabilizer D2322) present in domain A and interacting with sucrose substrate is zoomed-in.

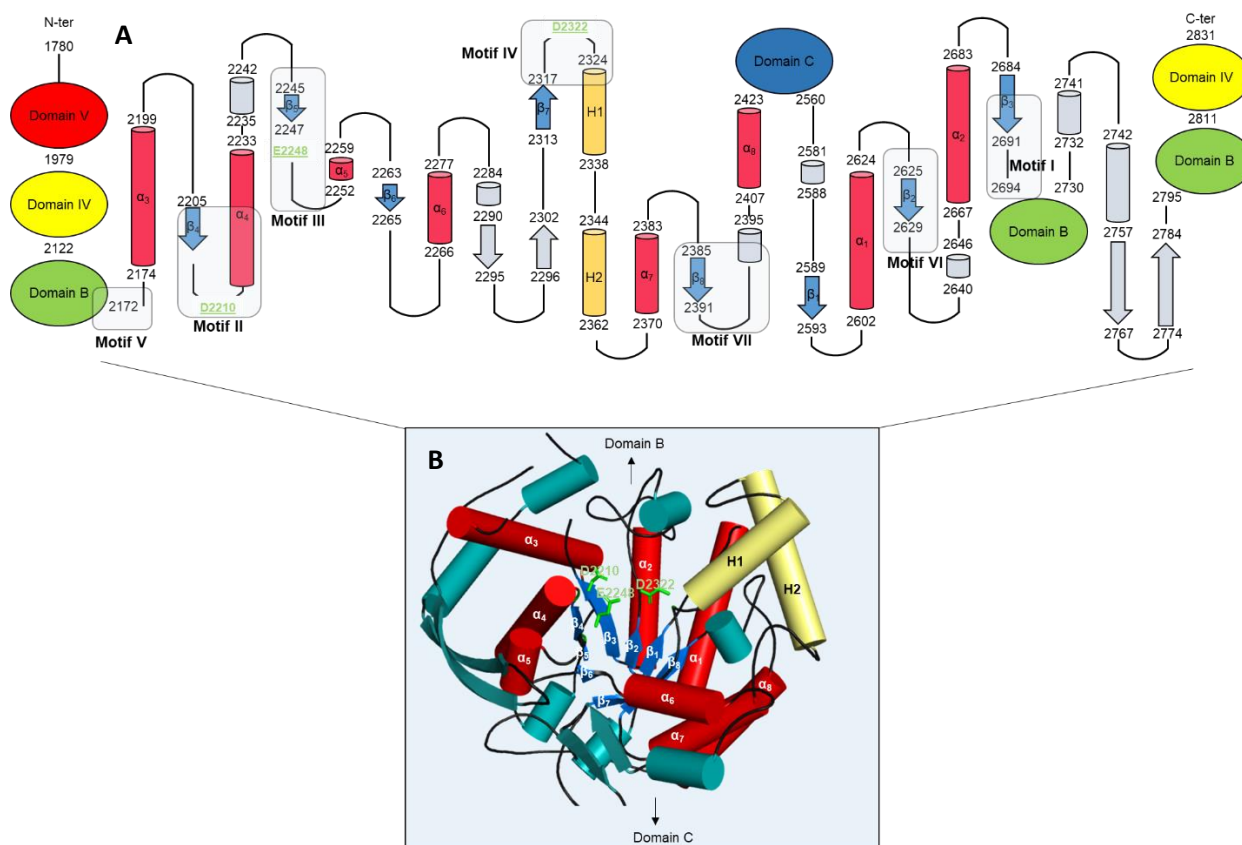


Figure 13: Topology diagram of GH70 enzymes. **A/** Schematic representation of the structural elements constituting the catalytic barrel of ΔN_{123} -GBD-CD2 (PDB ID: 3TTQ). α -helices are represented as cylinders (red for helices that are part of the $(\beta/\alpha)_8$ barrel, gold for helices H1 and H2 and gray for other helices), β -sheets as arrows (blue for sheets that are part of the $(\beta/\alpha)_8$ barrel, gray for other sheets), and loops as continuous lines. The different domains are represented in circles. The specific motifs I to VII are indicated in transparent blue boxes and the catalytic residues are represented in green.

B/ Three dimensional fold adopted by the structural elements of the catalytic barrel of ΔN_{123} -GBD-CD2 (PDB ID: 3TTQ). Helices α_1 - α_8 , H1 and H2, and sheets β_1 - β_8 are represented. At the edges of the catalytic site, the antiparallel β -sheets (2291-2294 // 2297-2301) and (2758-2766 // 2775-2783) and additional helices are represented in gray. Catalytic residues are represented in green sticks.

Subsites -1, +1 and +2 were defined according to the nomenclature proposed by Davies and colleagues (Davies et al. 1997) after the resolution of the structure of GTF-180- ΔN inactive mutant D1025N in the presence of sucrose donor (PDB ID: 3HZ3) or maltose acceptor (PDB ID: 3KLL) in the active site (Vujcic-Zagar et al. 2010). Subsite -1 fits the glucosyl moiety of sucrose, while subsite +1 fits the fructosyl moiety. When fructose is released from sucrose hydrolysis, subsite +1 and subsequent subsites +2, +3 etc. become available to fit an acceptor (cf part “Mechanistic insights of transglucosylation reaction”).

The GTF-180- ΔN D1025N:sucrose complex showed that 7 conserved residues formed subsite -1 interacting with the glucosyl moiety of sucrose: the acid/base E1063, the nucleophile D1025 (replaced with N in this mutant), the transition state stabilizer D1136 as well as H1135 (motif IV), R1023 (motif II), Q1509 (motif I) and Y1465 (domain A). The fructosyl moiety in subsite +1 interacts with less conserved residues from the catalytic site, notably E1063, W1065 (motif III), N1029 (motif II) and Q1140 (motif III). The same residues interact with sucrose in the structure of DSR-M $\Delta 2$ in complex with sucrose (PDB ID: 5O8L) (Claverie et al. 2017). Residues upfront

of the transition state stabilizer in motif IV (Figure 14), often conserved, are characteristic within a same subfamily of GH70 sharing the same linkage specificity, even if they are not sole responsible for the osidic linkages synthesized by the enzymes. For example, the triplet SEV is mostly related to dextransucrases but is also found in mutansucrases, the triplet YDA is characteristic of alternansucrase, and the triplet NNS is found in reuteransucrases. This triplet is more variable in branching sucrases (Figure 14).

	Motif II 2206	Motif III 2243	Motif IV 2317	Motif I 2688	
GBD-CD2	S I R I D A V D F I H	H I S L V E A G L D A G	I I H A H D K G V Q E K V G A	A D V V D N Q	Branching sucrose ($\alpha 1,2$)
DSR-M	S I R I D A V D N V D	H I H I L E D W S P N D	F I R A H D S E V Q T I I A K	A D F V P D Q	Dextransucrase ($\alpha 1,6$)
ASR	G I R V D A V D N V D	H L S I L E D W N G K D	F V R A H D Y D A Q D P I R K	A D W V P D Q	Alternansucrase ($\alpha 1,3/\alpha 1,6$)
GTF-180	G I R V D A V D N V D	H I N I L E D W G W D D	F V R A H D S N A Q D Q I R Q	A D W V P D Q	Dextransucrase ($\alpha 1,3/\alpha 1,3$)
GTF-A	S V R V D A P D N I D	H I N I L E D W N H A D	F V R A H D N N S Q D Q I Q N	A D W V P D Q	Reuteransucrase ($\alpha 1,4/\alpha 1,6$)
GTF-B	G E R V D A A D N I D	H L S Y N E G Y H S G A	F V T N H D Q R K N L I N R L	E D I V M N Q	Glucanotransferase ($\alpha 4,6$)
GTF-SI	S I R V D A V D N V D	H L S I L E A W S Y N D	F I R A H D S E V Q D L I R D	A D W V P D Q	Mutansucrase ($\alpha 1,3/\alpha 1,6$)
	*	*	*		
	Motif V 2165	Motif VII 2385	Motif VI 2622		
GBD-CD2	F L L A N D V D N S N P	T V P R M Y Y G D	W G I T S F E M A P Q Y		Branching sucrose ($\alpha 1,2$)
DSR-M	L L L G N D V D N S N P	T V P R V Y Y G D	W G V T S F Q M A P Q Y		Dextransucrase ($\alpha 1,6$)
ASR	F L L A N D I D N S N P	T V P R V Y Y G D	L G I T S F E L A P Q Y		Alternansucrase ($\alpha 1,3/\alpha 1,6$)
GTF-180	F L L A N D I D N S N P	S V P R V Y Y G D	W G I T S F E L A P Q Y		Dextransucrase ($\alpha 1,3/\alpha 1,3$)
GTF-A	F L L A N D I D N S N P	T I P R V Y Y G D	L G F T S F E M A P Q Y		Reuteransucrase ($\alpha 1,4/\alpha 1,6$)
GTF-B	L L V G N D I D N S N P	T V P Q V Y Y G D	L G I T D F W M A P A Y		Glucanotransferase ($\alpha 4,6$)
GTF-SI	F L L A N D V D N S N P	S V P R V Y Y G D	W G V T D F E M A P Q Y		Mutansucrase ($\alpha 1,3/\alpha 1,6$)

Figure 14: Sequence of highly conserved motifs I-VII in the sequences of GH70 available structures. Stars: catalytic residues. The alignment was created using ENDscript 2 (Robert and Gouet 2014). Residues numbers are numbered according to GBD-CD2 sequence.

The domain B is involved in the formation of domain A. It is generally formed of 5 to 6 antiparallel, highly-twisted β -sheets (5 in GBD-CD2) (Ito et al. 2011). Mutagenesis studies targeting some of the loops (loops B1 and B2) in domain B of GTF180 showed its important role in linkage specificity of this enzyme. For example the residue Leu940 when mutated into a tryptophan, altered the formation of both $\alpha 1,3$ and $\alpha 1,6$ linkages towards the formation of $\alpha 1,6$ linkages exclusively (Meng et al. 2014).

Domain C constitutes the architectural basis of the U-shaped enzyme and is the only domain formed of a continuous polypeptide fragment. It is mainly composed of β -sheets (10 in GBD-CD2) and both its N- and C- terminal fragments are quite conserved in GH70 family. Its function is still unknown but it was hypothesized that this domain could play a role in enzyme stability (unpublished data, PhD thesis of (Molina 2019)).

The domain V, also called Glucan binding domain (GBD), outlines the extremities of the enzymes. The capacity of glucan binding of this domain have been demonstrated in numerous studies (Abo et al. 1991; Monchois et al. 1998b; Argüello-Morales et al. 2000). Recently, and thanks to the first 3D structures obtained in complex with oligosaccharides mimicking the natural products, sugar binding pockets and amino acids in interaction with glucosyl units of isomaltooligosaccharides were identified in the domain V of the branching sucrose GBD-CD2,

the dextransucrase DSR-M and the alternansucrase ASR (Brison et al. 2016; Claverie et al. 2017; Molina et al. 2019). They always involve a tyrosine as well as a lysine and a glutamine residue, and if one of this amino acid is mutated into alanine –in one or several sugar pockets- the resulting size of the polymer (or the degree of branching) is drastically reduced (Claverie et al., 2017). Moreover, the structure of DSR-M and ASR clearly shows that there is a conformation where the domain V points towards the active site (Claverie et al. 2017; Molina et al. 2019). This further indicates that domain V could bind growing glucan chains and help keep them close to the active site for a semi-processive elongation, as previously proposed (Konishi et al. 1999; Monchois et al. 1998b, 1999; Moulis et al. 2006b; Osorio et al. 2019).

Domain IV, located between domain B and V, is unique to GH70 enzymes. Its exact function remains unknown, but it was suggested that it could play the role of a hinge helping to bring the flexible domain V close to domain A and moving it away as the dextran chain grows, easing the elongation (Ito et al. 2011). Two conformations -with domain V close or further from the active site- have been solved for the enzyme GTF-180- Δ N (PDB ID: 3KLLK and 4AYG, respectively). The enzyme behavior was also studied in solution by small-angle X-ray scattering (SAXS), showing the flexibility of the domain V and the bend between domains IV and V, thus supporting this hypothesis (Vujicic-Zagar et al. 2010; Pijning et al. 2014).



Figure 15: A large scale collective motion from bend normal mode (Figure from (Ben Imeddourene et al. 2018)). View of the normal mode of GBD-CD2 corresponding to the bend mode. Black arrows are directed towards the motion direction and their length is proportional to the amplitude of the movement. The orange sphere shows the catalytic site.

This hinge effect was also confirmed by modelling studies. Recently, Ben Imeddourene *et al.* performed long molecular dynamics simulations (1 μ s) using the branching sucrose GBD-CD2, showing the different motions the enzyme undergoes. In particular, correlated motions of the domains A/C and V could be observed in the bend normal mode (Figure 15) and was indicative

of the domain V moving towards domain A (Ben Imeddourene et al. 2018). Interestingly, the domain B was proposed, for the first time, to act as a flexible structural hinge region between domains A/C and IV/V.

Simulations showed that the domain V was the most flexible domain of the protein, which correlated with crystallographic data obtained in previous studies and with other enzymes (Pijning et al. 2014). Domain V was moving toward and away from the catalytic site, with a motion centered on a region from residues 1956 to 2004 in domains IV and V (Figure 15). This study also showed the flexibility of loops surrounding the active site, which could be related to the acceptor promiscuity of the enzyme, and that sucrose binding stabilized some of these loops (Ben Imeddourene et al. 2018).

Finally, and very recently, another study explored the enzyme-glucan interaction in the domain V of the GTF-SI mutansucrase (Osorio et al. 2019). MD simulations performed with GTF-SI (GTF) and GTF-SI deleted of the domain V (GTF-V) were compared. When domain V was present, it was highly flexible and caused conformational changes in residues from domain IV, B and A, although most residues involved in the catalysis remained unchanged. They concluded that the presence of the domain V should not have an effect on the catalytic parameters of the enzyme.

The authors also modelled a growing chain of glucan, trying to predict what could be the binding mode in the active site and along the enzyme. They showed that the glucan chain could be stabilized in domain V, the latter being responsible for orientation of the glucan chain, and it also interacted with domains A, B, IV. No interactions between domains A, B and IV and the growing glucan chain were observed in GTF-V model, and it just protruded from domain B (Figure 16), further indicating the importance of domain V in stabilization of the growing chain.

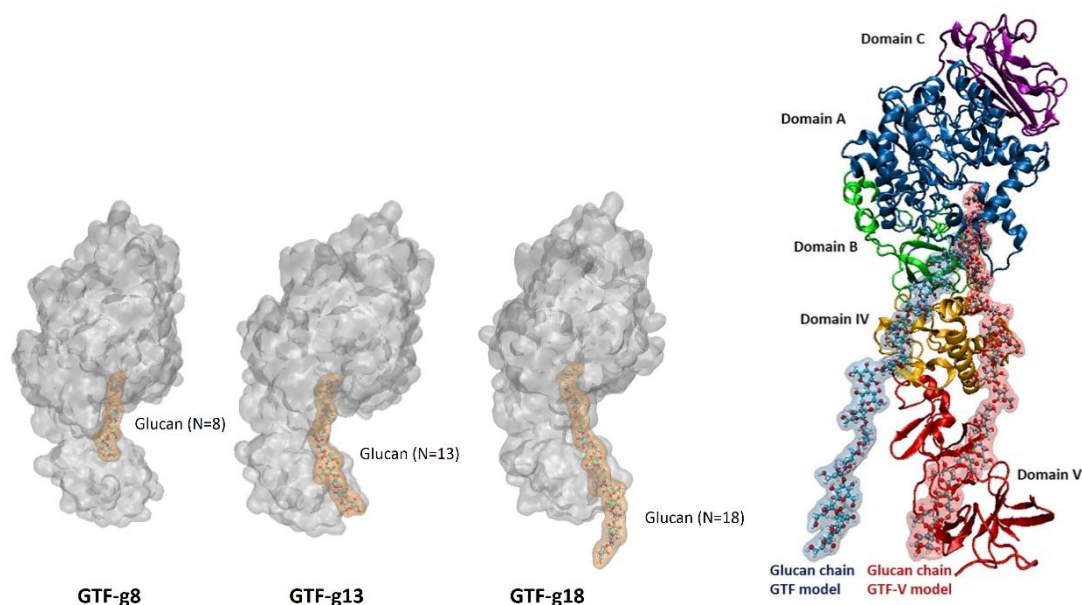


Figure 16: Orientation of the growing glucan chain from domain V (from Osorio *et al.* 2019, Figures 4 and S3 in the publication). Models representing the enzyme-glucan complexes with a growing glucan chain ($n = 8, 13, 18$ and 23 , from left to right). The glucan chain from GTF-V model protrudes from domain B without interacting with domains B and IV.

Latest evolutionary evidences on GH70 enzymes

Based on structural information recently acquired, it has been proposed that GH70 glucansucrases derived from GH13 α -amylases by successive: i) insertion of the specific domain IV, ii) circular permutation of the $(\alpha/\beta)_8$ barrel in the catalytic domain and insertion of the domain V, and iii) loop changes around the active site. The evolutionary intermediates would correspond to subfamilies GtfB-like and GtfC/D-like 4,6- α -glucanotransferases (Gangoiti et al. 2016; Bai et al. 2017) (Figure 17). The successive steps of evolution from GH13 α -amylases to GH70 glucansucrase (which are present in oral-bacteria) are proposed to be diet-driven, as humans shifted from essentially starch-based food to increased sucrose intake (Bai et al. 2017). As such, the subsite -1 that fits the glucosyl moiety of sucrose was enclosed, preventing longer donor chains to be used and restricting the donor to sucrose.

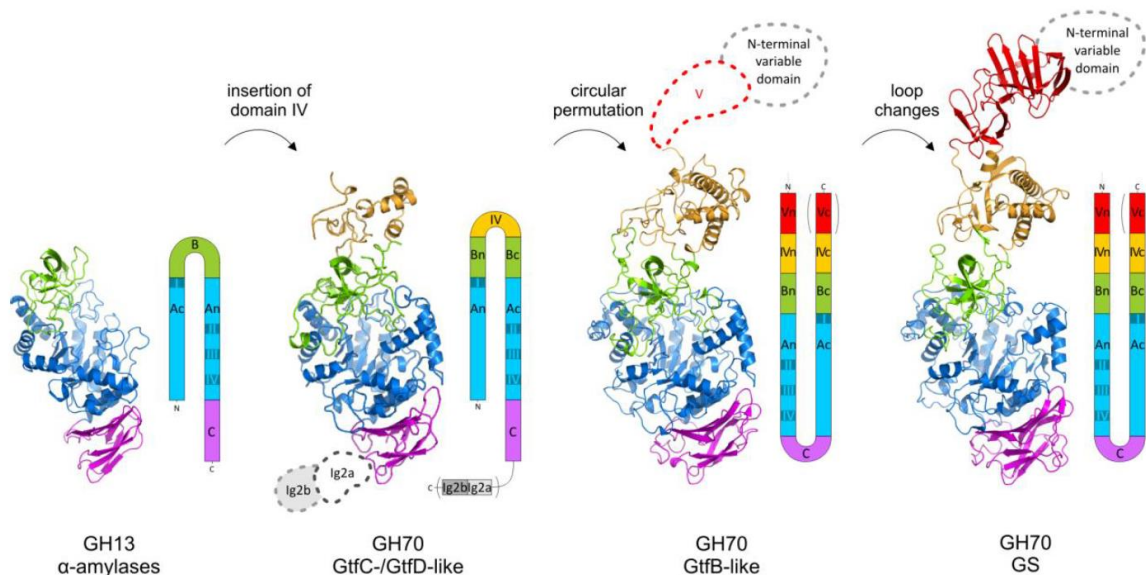


Figure 17: X-ray structures and modelled structures of representative members of GH13 and GH70 families, with their overall domain organization (Figure from (Gangoiti et al. 2018b)).

From left to right: crystal structure of α -amylase from *Bacillus licheniformis* (PDB ID: 1BLI; (Machius et al. 1998)); GtfC-type 4,6- α -glucanotransferase from *Exiguobacterium sibiricum* 255-15 (homology model, (Gangoiti et al. 2016)); crystal structure of GtfB- Δ N Δ V, the 4,6- α -glucanotransferase from *Lactobacillus reuteri* 121 (PDB ID: 5JBD; (Bai et al. 2017)); crystal structure of Gtf180- Δ N, the dextranucrase from *Lactobacillus reuteri* 180 (PDB ID: 3KLK; (Vujicic-Zagar et al. 2010)). The positions of domains absent in the model or crystallized constructs are indicated in dotted lines. A schematic U-shaped domain organization is shown next to each model/structure. This highlights the circular permutation that led to a different order of the conserved motifs (II-III-IV-I instead of I-II-III-IV) in GH70 GtfB-like enzymes and GSs, but not in GH70 GtfC- and GtfD-like enzymes and GH13 α -amylases. Some of the proposed evolutionary events related to domain organization and enzyme specificity are indicated. In α -amylase-like enzymes, insertion of domain IV may have led to a GtfC/GtfD-type domain organization; gene duplication events resulted in circular permutation of the catalytic domain; changes in the loop arrangement around the active site then resulted in a shift of substrate specificity from starch to sucrose (glucanotransferase to glucansucrase).

It is not yet known how exactly bifunctional dextranucrases (displaying two distinct catalytic domains that catalyze both polymerase and branching sucrose activities) and branching sucrases appeared in this evolutionary landscape. It was proposed that branching sucrases emerged after separation of 4,6- α -glucanotransferases and glucansucrases (Meng et al.

2016). Attempts to replace specific residues of branching sucrases for conversion into polymerases have proven difficult. For example, the GBD-CD2 residues A2249, G2250 and F2214 were replaced by GTF-180 equivalents, yielding mutants A2249W, G2250W, A2249D-G2250W and F2214N. The resulting mutant enzymes did not gain any polymerase activity and had a reduced hydrolytic activity on sucrose (Brison et al. 2012). However, mutations targeting motif II (7 amino acids substitutions), motif IV (5 amino acids substitutions), or both (12 amino acids substitutions) (Figure 14), were introduced and corresponded to residues present in the dextransucrase DSR-S. When one motif was changed, the resulting enzyme was found inactive but the double-motif mutant although having a highly decreased hydrolytic activity, formed a product which could be stained using Schiff coloration, sign that the product was most likely a polymer (Fabre et al. 2005).

Recently, the dextransucrase “Gsy” from *Ln. mesenteroides* BD3749 presenting features of both glucansucrases and branching sucrases was reported (Yan et al. 2018). It produced an α 1,6 linked glucan polymer with α 1,3 and α 1,4 branches. This enzyme was proposed to be an evolutionary intermediate between glucansucrases and branching sucrases. After sequence alignment of the core motifs of Gsy with GS or BRS, specific residues not found in GS nor BRS were identified (Gsy numbering): a tyrosine at position 592 in motif II (asparagine in GS and phenylalanine in BRS), a methionine at position 693 in motif IV (phenylalanine in GS and isoleucine in BRS), an arginine at position 695 in motif IV (histidine in BRS) and an arginine at position 699 in motif IV (serine in GS and lysine in BRS). It was proposed that these sites had a critical role in specificity of the enzymes and that their change could lead to a conversion of BRS into GS but this remains to be experimentally verified.

A quick look at structure-function relationships of the branching sucrose GBD-CD2

Several studies aimed at investigating in detail the structure-function relationships of GH70 enzymes, in particular molecular determinants responsible for the activity, substrate binding, product profile and size, or linkage specificity. To date, GBD-CD2 is the sole member of the branching sucrose subfamily for which a 3D structure is available. As this subfamily of enzymes was mainly targeted in this thesis work, more details about this enzyme will be briefly presented hereafter.

The active site of GBD-CD2 is constituted of a large open pocket that form a gorge (Brison et al. 2012). By sequence alignments, residues R2208, D2210, E2248, H2321, D2322, Y2650, and Q2694 were identified as equivalent to the conserved GTF180 residues that interact with the glucosyl moiety of sucrose in this enzyme. More recently, Ben Imeddourene *et al.* identified by modelling methods the main amino acids in interaction with sucrose, highlighting residues

K2323, N2648 and F2136 (Figure 18) (Ben Imeddourene et al. 2018). Interestingly, residue K2323 was found to interact with sucrose -as proposed by Brison *et al.* in 2012- but with the glucosyl unit and not the fructosyl unit of sucrose, while F2136 formed a hydrophobic stacking interaction.

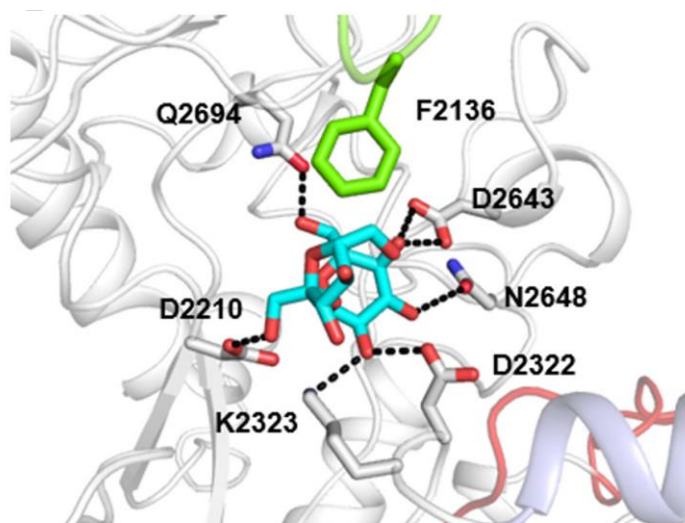


Figure 18: Main interactions between residues of the active site of GBD-CD2 and sucrose. Figure from Ben Imeddourene et al. 2018.

Clearly, less residues interact with sucrose compared to GTF180, what could explain the differences observed in sucrose affinity between GBD-CD2 and GTF-180, with GBD-CD2 having a 32-fold higher K_m .

Based on the 3D structure, several mutants of the branching sucrase GBD-CD2 were constructed (see “Latest evolutionary evidences on GH70 enzymes” paragraph). Targeted residues were non-conserved compared to GTF180: A2249, G2250, and F2214. In particular, mutant F2214N lost its ability to branch dextran while the other mutants lost sucrose hydrolysis activity (Brison et al. 2012). Three specific loops (D2292-I2299, G2731-S2796 and R2157-F2163) of the active site of GBD-CD2 were proposed as structural features responsible for branching activity with α 1,2 specificity, but the absence of crystallographic structures in complex with acceptors bound in the active site prevents confirmation of this hypothesis.

Structure in complex with glucose molecules allowed identification of several binding sites (Brison et al. 2016), notably in grooves expanding from subsite -1. Residues E2265, S2266 and E2270 from so-called binding sites A2 and A3 present in the active site, also conserved in the α 1,2 branching sucrase BRS-A (Passerini et al. 2015), were proposed as putative residues responsible for the α 1,2 specificity (Brison et al. 2016), but with no biochemical confirmation. Rearrangement of loop 7 of the active site (H2319-S2369, connecting β 7-sheet and α 7-helix (Figure 13), and containing H1-H2 subdomain) was observed when comparing structure in *apo* form and in complex with isomaltotriose. This rearrangement opened a cleft in the active site that could be responsible for accommodation of the acceptor in a conformation allowing α 1,2 branching. However, complex with longer oligosaccharides would be required to confirm this

hypothesis (Brison et al. 2016). Finally, this study also allowed the identification of the sugar binding pockets of domain V, as previously described.

The study of structure-function relationships of GBD-CD2 did not yet allow to fully understand what molecular determinants are responsible for the α 1,2 specificity. However, they helped to better understand the behavior of the enzyme toward its natural acceptor, and identify critical residues in subsite acceptors. This allowed to guide a structure-based engineering of this enzyme in a later study, notably for the glucosylation of flavonoids (see in below section, (Malbert et al. 2018)). Comparison with other structures of α 1,2 branching enzymes are still necessary, as well as determination of three-dimensional crystallographic structures of α 1,3 branching sucrases that could help to pinpoint residues responsible for their specificity.

Examples of engineered GH70 enzymes for glucodiversification

The use of engineered enzymes for bioactive oligosaccharide synthesis has already been largely addressed in the first part of the bibliographic survey (see part I). We will thus focus in this section on the use of engineered glucansucrases and branching sucrases of family GH70 to achieve glucodiversification of various molecules (hetero-oligosaccharides and gluco-conjugates) (Table 4). The Table 4 briefly lists some of the exogenous molecules glucosylated by native or engineered glucansucrases.

For instance, the DSRB742 dextransucrase was first evolved by X-ray mediated random mutagenesis from wild-type DSRN, as this method was previously found to yield interesting random mutants. The goal was to improve glucosylation of exogenous disaccharides, which could be used as food sweeteners and stabilizers, prebiotics or immunity-promoting agents. The generated enzyme DSRB742 contained 4 mutations. The same 4 mutations were also introduced individually on the gene of the wild-type enzyme by site-directed mutagenesis, generating mono-mutants DSRN1 to DSRN4. After evaluation of the 4 different mutants, mutant DSRN3 (K395T) showed a ~2-fold improvement of its glucosylation rate toward maltose, gentiobiose and salicilin glucosylation compared to the wild-type (Nam et al. 2008). Similarly, GTF-180 Δ N mutated on positions Q1140, W1065 and N1029 allowed for change in the product ratio when using lactose as an acceptor, generating hetero-oligosaccharide mixtures called GL34 and containing 5 compounds (F1-F5) of DP3-DP4 (Pham et al. 2017, 2018), which could have prebiotic effects (Pham et al. 2019). These residues were identified by studying the structure of GTF-180 in complex with maltose. Residue Q1140 is present at subsite -1, W1065 perform hydrophobic stacking interaction with glucosyl units of maltose at +1 and +2 subsites, and N1029 is involved in hydrogen bonds with maltose. The three residues are in close proximity to acceptor subsites and could be involved more generally in interactions with other acceptors like lactose. Therefore, 23 single mutants containing mutations at one of these 3 positions were used (Meng et al. 2015, 2017). In particular, mutants Q1140W, Q1140N

and N1029T showed a better production of some of the glucosylated lactose products (F2 and F4, or F3, respectively) from around 1.5 to 2-fold better than wild type enzyme. Meanwhile, N1029G and W1065M mutants synthesized new products, compared to the parental enzyme (Pham et al. 2018).

Residue Q1140 from GTF-180 was also targeted in studies aiming at glucosylating stevioside RebA, used as a sweetener, and the most promising results were obtained with mutant Q1140E, for which a better conversion of RebA was observed by TLC compared to the wild-type (Devlamynck et al. 2019). After optimizing reaction conditions, mutant Q1140E was able to reach a conversion of almost 100%, producing glucosylated RebA that was further tested for its organoleptic properties and had a good sweetness with less aftertaste than with aglycon RebA. This study further pointed out the critical role of residue Q1140, close to the transition state stabilizer D1136, for recognition of exogenous acceptor.

Other mutants of this enzyme were constructed with mutations targeting L938, L981 and N1029 in order to glucosylate molecules of industrial interest. Residues L938 and L981 were identified as being close to subsite +1 in the structure of GTF180 in complex with maltose (Meng et al. 2015). Compared to the wild-type, mutant L981A had improved monoglucosylation yields of catechol (~1.6-fold), resorcinol (~3-fold), hydroquinone (~7-fold), butanol (~9.8-fold), hexanol (~4.8-fold) and octanol (~4.8-fold) (Devlamynck et al. 2016) .

Due to the recent discovery of the branching sucrase sub-family, mutants were mainly built for structure-function relationship studies. The sole example found in the literature concerns the engineering of GBD-CD2 for the glucosylation of a range of potentially antioxidant flavonoid molecules (Malbert et al. 2018). A library of 22 mutants was constructed after analysis of the structure docked with flavonoid molecules. Residues present in first and second shells of the active site, in +1, +2 and +3 acceptor subsites, were targeted (W2135, F2136, F2163 and L2166). Various flavonoid molecules (quercetin, luteolin, morin, naringenin, apigenin, and chrysin) that shared a common backbone characteristic of flavonoids but with different hydroxylation patterns were then tested in glucosylation assays. For each molecule, glucosylation at different positions was observed depending on the mutants considered. Apigenin and chrysin were poorly glucosylated and mutants could improve yield from only 2% for the parental enzyme to 5%. More promising results were obtained for quercetin and luteolin, well glucosylated by several mutants, in particular W2135C-F2136I and W2135S-F2136L for quercetin (from ~58% yield for the parental enzyme to ~90% yield), and W2135C-F2136I and W2135G for luteolin quercetin (from ~45% yield for the parental enzyme to ~100% yield). These mutants also produced new glucosylated forms of the molecules (QG3a and LG3a respectively). Morin and naringenin glucosylation were also improved in particular with mutants W2135G and W2135S-F2136L (from ~20% yield for the parental enzyme to 66 and 60%, respectively), and they also allowed to access new glucosylated positions (MG2c and NG2b, respectively).

Table 4: List of some GH70 enzymes that were used in literature for the glucosylation of various molecules. ^a ΔN_{123} -GBD-CD2 is engineered from a bifunctional glucansucrase, wild type notation is therefore not applicable. ^bFor the library used in Malbert et al. 2018, best mutants were dependent on the molecule used as acceptor.

Template enzyme	Wild-type / Variant	Acceptor targeted for glucosylation	Applications	Reference
Dextransucrase from <i>Lm</i> B1299CB	Wild type	Quercetin	Food (additive), Cosmetics	(Moon et al. 2007)
Dextransucrase from <i>Lm</i> B1299CB	Wild type	Epigallocatechin gallate	Food, Pharmaceutical, Cosmetics	(Moon et al. 2006)
Dextransucrase from <i>L/EG001</i>	Wild type	Puerarin	Food, Pharmaceutical	(Ko et al. 2012)
Dextransucrase from <i>Lm</i> B1299CB	Wild type	Ampelopsin	Pharmaceutical, Cosmetics	(Woo et al. 2012)
Dextransucrase from <i>Lm</i> B-512 FMCM	Wild type	Gallic acid	Cosmetics	(Nam et al. 2017)
GTF-180	Wild type	Lactose	Food (prebiotics)	(Pham et al. 2017)
DSR-N	K395T	Maltose, gentibiose, salicilin	Food, Pharmaceutical	(Nam et al. 2008)
GTF-180	Q1140W, Q1140N, N1029G	Lactose	Prebiotics	(Pham et al. 2018)
	Q1140E	Steviosides	Food (sweetener)	(Devlamynck et al. 2019)
	L981A	Catechol, resorcinol, hydroquinone, butanol, hexanol, octanol	Heavy industry	(Devlamynck et al. 2016)
ΔN_{123} -GBD-CD2	Parental enzyme ^a	α -L-Rhap-(1 \rightarrow 3)-2-deoxy-2-trichloroacetamido- β -D-Glcp-OAll	Pharmaceutical	(Salamone et al. 2015a)
ΔN_{123} -GBD-CD2	W2135X-F2136X ^b F2163X-L2166X ^b	Quercetin, luteolin, morin, naringenin, apigenin, chrysin	Food (additive), Cosmetics	(Malbert et al. 2018)

The number of studies describing the use of engineered glucansucrases and branching sucrases for glucodiversification emerged only in the last few years but one can expect this number to increase thanks to the glucosylation potential of these enzymes which use a cheap substrate and hopefully to new members of this family that will be discovered as well as to the better understanding of their mode of action.

Chapters II and III of this thesis are additions to the existing literature and will describe the use of engineered branching sucrases for generation of hetero-oligosaccharides, *i.e.* glucosylated tetrasaccharides.

RESULTS

CHAPTER II

Summary: *This chapter presents the chemo-enzymatic strategy devised to access antigenic oligosaccharides representative of Shigella flexneri serotypes, the chemical route developed to produce a lightly protected tetrasaccharide **ABC'D'** and the use of native branching sucrases from the GH70 family for glucosylation of the tetrasaccharide target molecule.*

Main achievements: *The native branching sucrases were able to glucosylate the tetrasaccharide, forming two distinct mono-glucosylated products (**P1** and **P2**) as well as poly-glucosylated products. The main mono-glucosylated product **P1** was characterized and the resulting pentasaccharide corresponded to Shigella flexneri serotypes 4a/4b.*

Contributions to the work: *The design and chemical synthesis of the tetrasaccharide was done by our colleagues at Pasteur Institute (L. Mulard, G. le Héget and Z. Hu).*

I was responsible for the production and purification of the enzymes, screening of their activity on sucrose, set-up and optimization of the reaction conditions, development of the analytical methods, scale-up of the reaction, and purification of the product of interest.

*High resolution mass spectrometry for analysis of the reaction media was performed by H. Barbier. Structural elucidation of the product **P1** was performed by NMR by A. Ben Imeddourene.*

A convergent chemoenzymatic strategy to deliver a diversity of *Shigella flexneri* serotype-specific O-antigen segments from a unique lightly protected tetrasaccharide core

Introduction

Shigella are Gram negative enteroinvasive bacteria frequently associated with diarrhoeal disease. They are responsible for shigellosis, or bacillary dysentery, and represent a significant cause of mortality and morbidity worldwide (Kotloff et al. 2018). Recent estimates emphasize the importance of shigellosis burden in the pediatric populations living in low and middle income countries (Liu et al. 2016; Platts-Mills et al. 2015), They also point to *Shigella* as being the second leading cause of diarrhoeal mortality among all ages (Khalil et al. 2018). Disease burden combined to increasing antibiotic-resistance call for a vaccine ensuring protection against *Shigella* (Barry et al. 2013; Tagliabue and Rappuoli 2018), and in particular against the numerous *S. flexneri* serotypes identified amongst the prevalent isolates found in *Shigella* infection (Livio et al. 2014). *Shigella* lipopolysaccharides (LPSs), and in particular the O-antigen (O-Ag) moieties, are prime targets for the host's protective immunity against infection (Cohen et al. 1991, 1988) and serum IgG antibodies to *Shigella* LPS have emerged as a correlate of protection against shigellosis (Cohen et al. 2019). Over the past decades, conjugates derived from *Shigella* surface polysaccharides have been extensively investigated as parenteral vaccines (Barel and Mulard 2019). Synthetic carbohydrate conjugates encompassing well-defined oligosaccharides acting as functional O-Ag mimics are actively explored as an attractive option (Pozsgay et al. 1999; Phalipon et al. 2009; Said Hassane et al. 2009; Bélot et al. 2005). Developed by L.A. Mulard and collaborators against *S. flexneri* serotype 2a (Phalipon et al. 2009; van der Put et al. 2016; Bélot et al. 2005), the most advanced vaccine candidate in this category – SF2a-TT15 – was found safe and immunogenic in young naïve adults in a phase I clinical trial (Cohen et al. 2017). To answer the need in the field (Livio et al. 2014), investigation towards broadening *S. flexneri* serotype coverage is ongoing (Boutet et al. 2016). For that purpose, synthetic routes for straightforward access to well-defined O-Ag fragments are essential (Hu et al. 2017; Hargreaves et al. 2014; Chassagne et al. 2013; Gauthier et al. 2014). However, the multiple *S. flexneri* types and subtypes identified so far have revealed a tremendous O-Ag structural diversity built from a unique core (Knirel et al. 2015). Indeed, while most *S. flexneri* O-Ags share the same linear backbone composed of 1,2-*trans* linked L-rhamnosyl (A, B, C) and *N*-acetyl-D-glucosamine (D) residues, the repeating units of the different serotype-specific O-Ags differ by their α -D-glucosylation (E) pattern, as well as by their *O*-acetyl and phosphoethanolamine substitutions (Figure 1). One main synthetic roadblock resides in the regio- and stereo-specific control of the 1,2-*cis* glucosylation step as concluded from the numerous oligosaccharides representative of *S. flexneri* O-Ag glucosylation patterns that were synthesized chemically. Early attempts at the chemical

grafting of the α -D-glucosyl residue on protected di-, tri-, and tetrasaccharide acceptors met issues (Mulard et al. 2000; Wessel and Bundle 1985), resulting in poor yields of the α -linked glucosylated products. As a general trend in subsequent work, the glucosyl residue was introduced first to form a disaccharide building block featuring the branching point (Hargreaves et al. 2014; Mulard et al. 2002; Boutet et al. 2008; B elot et al. 2004).

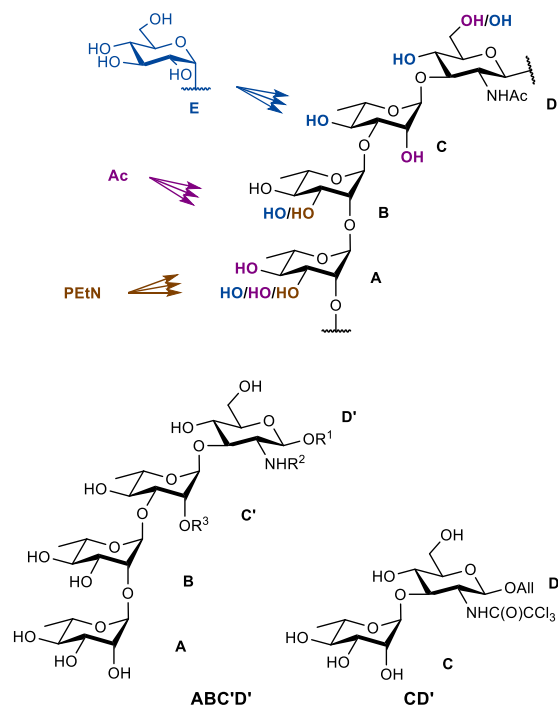


Figure 1: *Top:* The linear ABCD tetrasaccharide featuring the core of the O-Ag repeating units characterizing all *S. flexneri* serotypes and subtypes, except *S. flexneri* type 6. Type-specific substitutions include α -D-glucopyranosyl residue (E, blue), stoichiometric and/or non-stoichiometric O-acetyl groups (Ac, purple) and phosphoethanolamine residue (PEtN, brown). *Bottom left:* ABC'D', the lightly protected ABCD-like tetrasaccharide showing the generic protection pattern. All hydroxyl groups are identified as potential sites of enzymatic glucosylation with modification at 3_A, 3_B, 4_C, 4_D and 6_D featuring *S. flexneri* type-specific branching points. Rⁿ: protecting groups. *Bottom right:* the CD' disaccharide previously used as acceptor in the chemoenzymatic synthesis of a *S. flexneri* 2a 10mer and 15mer oligosaccharides (Salamone et al. 2015a, 2015b).

The chemoenzymatic synthesis of complex oligosaccharides has found wide interest over the past decades (Schmaltz et al. 2011; Crout and Vic 1998; Palcic 2011; Nycholat et al. 2013). In recent years, the use of synthetic core intermediates and their modification upon action of a range of glycosyltransferases has contributed to tremendous progress in the field. In this context, achievement from the Boons' group on the synthesis of libraries of asymmetrical multi-antennary N-glycans based on a scaffold compatible with customized modifications (Wang et al. 2013) has paved the way to novel developments. Obviously, the feasibility of protecting group-controlled enzymatic glycosylation by use of partially protected acceptors represents a powerful strategy to guide glycodiversification from a core precursor (Li et al. 2016; Gagarinov et al. 2017, 2019). As part of this ongoing trend supporting innovation in the field of chemoenzymatic synthesis, we have explored in recent years the potential of chemical synthesis and site-selective enzymatic α -D-glucosylation working in synergy to access various

S. flexneri serotype-specific carbohydrate precursors (Champion et al. 2009a; Urbach et al. 2014; Vergès et al. 2015) as a way to overcome the poor α/β stereoselectivity of the chemical glucosylation process (Nigudkar and Demchenko 2015). In particular, the successful chemoenzymatic synthesis of the 15mer oligosaccharide hapten involved in SF2a-TT15 from a chemically synthesized disaccharide precursor (Salamone et al. 2015a). However and of special concern, these elegant and promising strategies still required intensive chemistry efforts as distinct synthetic routes have to be developed depending on the targeted *S. flexneri* serotype.

With the aim of alleviating efforts and offering novel and easy-to-implement synthetic alternatives, we came up with the idea of developing a highly divergent chemoenzymatic strategy opening the way to a large panel of *S. flexneri* type-specific oligosaccharides. Taking advantage of the unique structure of the ABCD tetrasaccharide defining the backbone repeating unit of most *S. flexneri* O-Ags (Figure 1), the strategy is centered on the use of a common non-natural ABCD-like tetrasaccharide precursor. The latter is designed to be compatible with a site-selective enzymatic 1,2-*cis* glucosylation and subsequent chain elongation, allowing to deliver a portfolio of well-defined modular building blocks that could then be easily assembled into *S. flexneri* O-Ag mimics relevant for vaccine design.

Herein, we present the design and chemical synthesis of ABC'D', an original lightly protected ABCD-like tetrasaccharide (Figure 1), and its subsequent glucosylation using highly promiscuous sucrose-active α -transglucosylases found in the Glycoside Hydrolase family 70 (Moulis et al. 2016). This unprecedented concept, which avoids the use of poorly accessible glucosyl transferases, in particular those predicted to encompass multiple transmembrane regions such as non-Leloir glucosyltransferases found in *S. flexneri* (Allison and Verma 2000), and complex nucleotide- or polyprenol phosphate-activated donors, is exemplified in this study to produce a pentasaccharide specific for *S. flexneri* serotypes 4a/4b. Extension of this challenging albeit promising chemoenzymatic strategy in order to provide a large panel of type-specific oligosaccharides representative of *S. flexneri* O-Ag diversity from one single synthetic tetrasaccharide core is in progress.

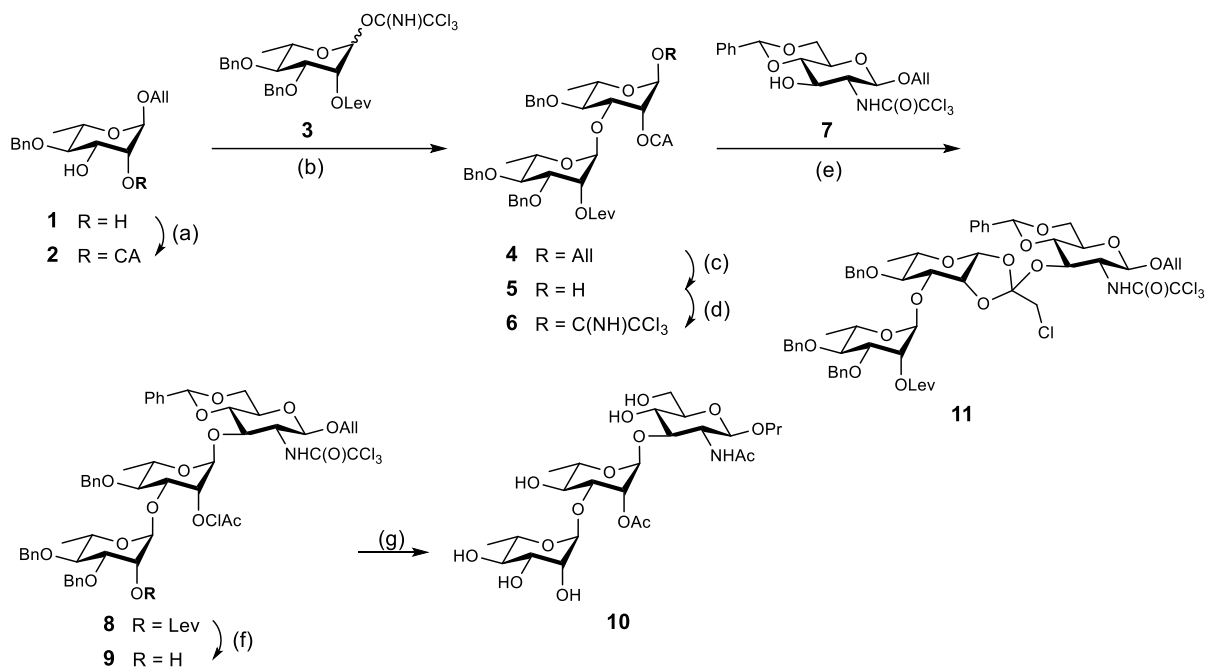
Design and chemical synthesis of tetrasaccharide **12 (ABC'D')**

Built from the ABCD backbone repeating unit common to all except one *S. flexneri* O-Ags identified to date, tetrasaccharide **ABC'D'** was designed for compatibility with enzymatic α -D-glucosylation at all possible naturally occurring glucosylation sites encountered on *S. flexneri* O-Ags (Figure 1) and with inherent O-Ag O-acetylation in addition to feasibility of subsequent chemical chain elongation of the products of enzymatic glucosylation at both ends. Biocatalyst compatibility imposed minimal protection. On that basis, the selected **ABC'D'** (tetrasaccharide **12**) was synthesized as an allyl glycoside encompassing a 2_D-trichloroacetamide (TCA) moiety

to ensure anchimeric assistance during chain elongation and recovery of the 2_D-acetamide upon final Pd/C-mediated hydrodechlorination (Hu et al. 2017). This protecting group combination was found appropriate to achieve the chemoenzymatic conversion of a CD' disaccharide into a *S. flexneri* 2a 15mer hapten of vaccine interest (Salamone et al. 2015a). Owing to the 2_C-O-acetylation found in the O-Ag of some of the most prevalent *S. flexneri* serotypes, for example *S. flexneri* 1b and 3a (Knirel et al. 2015), tetrasaccharide **12** also bears a chloroacetyl group (CA) at position 2_C. Advantageously, the CA moiety provides a direct access to a 2_C-acetate as demonstrated herein on the model trisaccharide (**10**) or more commonly to a 2_C-OH, as required.

Trisaccharide **9** features a BCD segment closely resembling the selected **ABC'D'** acceptor. Besides benzyl ethers at O-4_C, O-3_B and O-4_B, it bears all protecting groups making the glucosyl acceptor **12** unique *per se*. It was synthesized from two easily accessible crystalline precursors, the rhamnoside diol **1** (Boutet et al. 2008; Gigg et al. 1985), itself evolved into the trichloroacetimidate **3** (Boutet and Mulard 2008), and the glucosaminide acceptor **7** (Guen et al. 2017) (Scheme 1). Regioselective chloroacetylation at the axial 2-OH of diol **1** by means of the acid-mediated opening of the orthoester formed by reaction with 2-chloro-1,1,1-trimethoxyethane was high yielding (Boutet and Mulard 2008; Oscarson and Tedebark 1996). Glycosylation of acceptor **2** with rhamnosyl donor **3** gave disaccharide **4**, then converted to trichloroacetimidate **6** by means of hemiacetal **5**. TMSOTf-mediated glycosylation of acceptor **7** with donor **6** gave the fully protected trisaccharide **8** together with orthoester **11** (27%) as evidenced from the signal typical of an orthoester quaternary carbon at 120.6 ppm in the ¹³C NMR spectrum of side-product **11**. The engagement of chloroacetate in the formation of intersaccharidic orthoesters is not without precedent (Fürstner et al. 2002), and has occasionally been exploited to differentiate between ester protecting groups (Wang et al. 2009). It may be overcome by using harsher glycosylation conditions (Abronina et al. 2011). Irrespective of this outcome, trisaccharide **8** was delevulinoylated into the target alcohol **9**, which was then converted to monoacetate **10** upon Pd/C concomitant debenzoylation, allyl to propyl reduction and hydrodechlorination of the TCA and CA moieties into the corresponding acetamide and acetyl group, respectively.

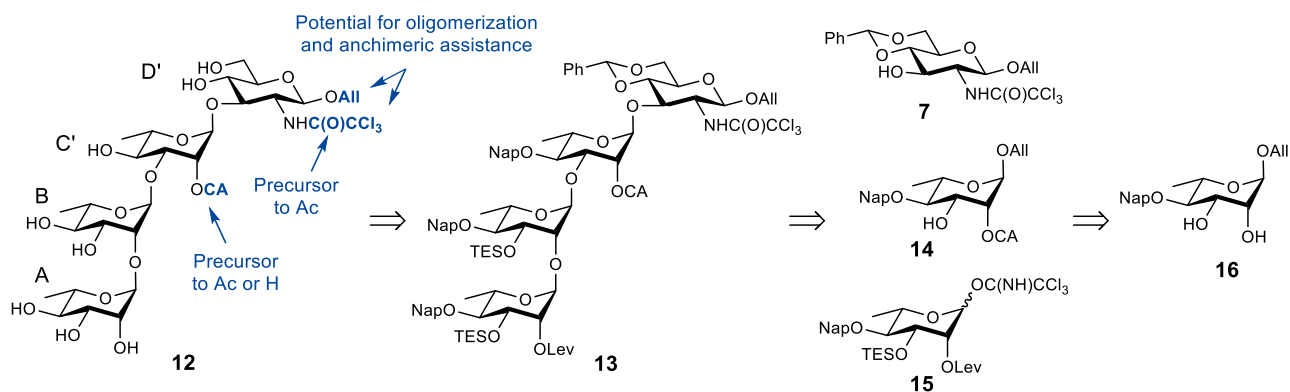
Scheme 1: Synthesis of Trisaccharide 9 and Pd/C-mediated Hydrodechlorination Thereof to Give BCD 10.¹



¹Reaction conditions: (a) (i) (MeO)₃CCH₂Cl, PTSA, MeCN, (b) 80% aq. AcOH, 95%; (b) TMSOTf, Et₂O, -20 °C, 91%; (c) (i) [Ir], THF, (ii) I₂, THF/H₂O, 83%; (d) Cl₃CCN, DBU, DCE, 96%; (e) TMSOTf, DCM, -15 °C, 62%; (f) H₂NNH₂, AcOH, pyridine, 86%; (g) H₂, Pd/C, EtOH:DCE, 47%. [Ir] = [Ir(THF)₂(PMePh₂)₂]PF₆.

Having achieved the proof of concept supporting CA as a suitable precursor to either a 2_C-O-acetyl or a 2_C-hydroxyl group, we turned to the chemical synthesis of the targeted **ABC'D'**. A highly convergent strategy was established, exploiting the utility of crystalline diol **16** as precursor to rhamnosides A, B and C and the iterative use of donor **15** for the introduction of rhamnosides B and A. It also ensured a minimal number of deprotection steps, using hydroxyl masking groups orthogonal to allyl, TCA and CA (Scheme 2). A strategy still barely adopted in oligosaccharide synthesis, although it has met some success for the synthesis of demanding glycolipids bearing unsaturated lipid chains (Swarts and Guo 2010; Ding et al. 2017; Lee et al. 2016) and is the subject of increasing interest (Li and Liu 2014), was favored. It involves a panel of protecting groups compatible with acid-mediated removal. Therein, the selected combination entails a 4_D,6_D-O-benzylidene acetal, a 2-naphtylmethyl ether (Nap) (Li and Liu 2014) to mask the 4-OH of the three rhamnosides and a triethylsilyl ether (TES) (Smith 1999) for its ease of introduction at O-3 of diol **16**. Tunable acid-sensitive ester protecting groups recently developed (Li and Liu 2014) do not fulfill orthogonality criteria with the above selection. Emphasis was therefore on convergence and the remaining hydroxyl group in donor **15** was masked by means of a levulinoyl ester to ensure anchimeric assistance during glycosylation and subsequent selective cleavage in the presence of the 2_C-CA (Ho and Wong 1975).

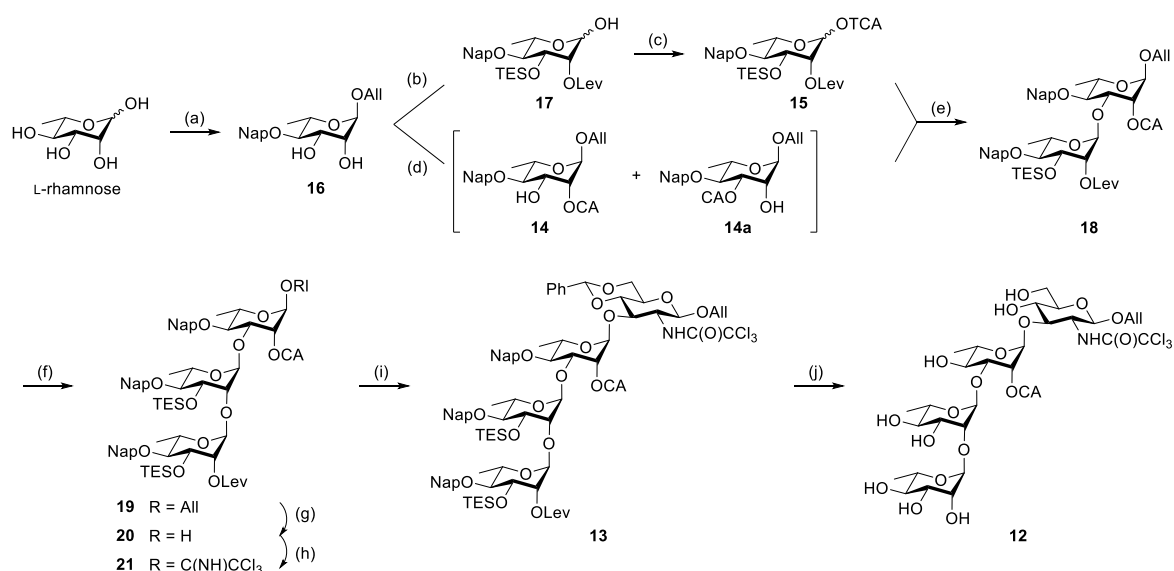
Scheme 2. The Tetrasaccharide **12** Showing Key Intermediate and Building Blocks.



Crystalline diol **16** was obtained in four steps from L-rhamnose according to a protocol adapted from the well-established synthesis of diol **1** (Scheme 3). Regioselective silylation at OH-3 with triethylsilyl chloride in the presence of diisopropylethylamine and subsequent levulinoylation at OH-2 gave the fully protected allyl rhamnoside intermediate. Anomeric deallylation of the later was best performed by means of conventional metal-catalysed allyl isomerisation and iodine-mediated hydrolysis of the propen-1-yl intermediate in a buffered medium. By avoiding intermediate purification, this three-step process was achieved in 80% overall yield to give hemiacetal **17**, the key intermediate to the crystalline trichloroacetimidate **15**, then isolated with an excellent yield. The established optimized conditions enabled the synthesis of donor **15** on a 30 g scale. Alternatively, diol **16** was converted to the 2-CA acceptor **14** as described to obtain the 4-O-benzyl analog **2**. The crude acceptor, isolated as a 95:5 mixture of the 2-CA and 3-CA regioisomers, was engaged in a TMSOTf-mediated glycosylation with donor **15** to give the crystalline disaccharide **18**. Hydrazine mediated-delevulinoylation unmasked the 2_A-OH to give the corresponding acceptor, which was directly engaged in glycosylation with donor **15** to avoid TES loss and achieve trisaccharide **19**. The two-step deallylation process optimized to obtain hemiacetal **17** was adapted to the later, albeit changing iodine for *N*-iodosuccinimide to ensure mild hydrolysis conditions compatible with protecting group integrity. The resulting hemiacetal **20** was evolved into the *N*-phenyl trifluoroacetimidate donor **21**, which reacted with an excess of glucosaminide acceptor **7** in toluene containing catalytic TBSOTf. To our satisfaction, under these conditions, improved in comparison to those used for obtaining trisaccharide **8**, the reaction proceeded smoothly and no orthoester formation was observed. Tetrasaccharide **13** was isolated in a very good yield and engaged in a two-step controlled deprotection process. Conventional delevulinoylation unmasked the 2_A-OH providing an intermediate, which was not isolated. Instead, the obtained alcohol was treated in TFA/toluene (9:1, v/v), under conditions inspired from the work of Liu et al, (Li and Liu 2014) to achieve the concomitant chemoselective removal of the TES, Bzl and Nap protecting groups, albeit in a modest 36% yield over two steps post extensive RP-HPLC purification. In agreement with previous reports (Lee et al. 2016; Li and Liu 2014), all protecting groups on residues A, B and D' were readily cleaved. In contrast, the isolated 4_C-Nap ether was found more resistant to these established conditions. Prolonged reaction time at room temperature were necessary to

achieve full unmasking of the 4_C-OH, possibly at the extent of the rhamnosidic linkage integrity. Indeed, to our knowledge this work exemplifies for the first time the applicability of TFA-mediated acid treatment for the removal of multiple protecting groups, including an isolated Nap ether, on a complex oligosaccharide, which contains several acid-sensitive glycosidic linkages. As expected, the CA moiety survived these conditions. In contrast, it was found sensitive to the original RP-HPLC/lyophilization conditions (MeCN/H₂O), which resulted in the repeated isolation of minute amounts of dechloroacetylated tetrasaccharide **22**. Since the CA ester had to comply with conditions most suitable for GH-mediated enzymatic glucosylation, these observations encouraged further investigation on the range of pH compatible with tetrasaccharide **12** integrity. A follow up of the kinetics at different pH revealed that an optimal pH range between 4 to 5 (data not shown).

Scheme 3. Chemical Synthesis of the Tetrasaccharide **12**.¹



¹Reaction conditions: (a) (i) AcCl, AlI₃, reflux; (ii) Me₂C(OMe)₂, PTSA, MeCN, (iii) NaH, NapBr, DMF, -5 °C to rt, (iv) 80% aq AcOH, 80 °C, 87%; (b) (i) TESCl, DIPEA, MeCN, (ii) DCC, DMAP, LevOH, DCM, (iii) [Ir], THF then I₂/NaHCO₃, THF/H₂O 5:1, 80% (α/β 4:1); (c) Cl₃CCN, DBU, DCE, 91% (α/β 19:1); (d) (i) (MeO)₃CCH₂Cl, PTSA, MeCN, (ii) 90% aq. TFA, quant. (**14/14a**: 19:1); (e) TBSOTf, 4 Å MS, PhMe, -40 °C, 82%; (f) (i) NH₂NH₂·H₂O, Pyridine/AcOH 3:2, (ii) **15**, TBSOTf, 4 Å MS, PhMe, -20 °C, 83%; (f) (i) [Ir], THF, (ii) NIS, THF/H₂O 5:1, 80%; (g) NPTFA-Cl, K₂CO₃, Acetone, 91%; (h) **7**, TBSOTf, 4 Å MS, PhMe, -40 °C, 84%; (i) (i) NH₂NH₂·H₂O, Pyridine/AcOH 3:2; (ii) TFA/PhMe 9:1, 36%.

Enzymatic glucosylation of tetrasaccharide **12** (ABC'D')

Branching sucrases are GH70 enzymes produced by some strains of lactic acid bacteria that use a cheap and abundant substrate, sucrose, to glucosylate hydroxylated molecules. (Fabre et al. 2005; Passerini et al. 2015; Vuillemin et al. 2016). Naturally, they branch glucosyl moieties via α-1,2 or α-1,3-linkages onto linear chains of α-1,6 linked glucosyl units called dextrans. But these enzymes have also demonstrated a remarkable promiscuity toward exogenous acceptors such as carbohydrates, flavonoids or polyols for instance (Daudé et al. 2014; Malbert et al. 2018), making them candidates of choice to attempt the glucosylation of

tetrasaccharide **12**. One of these enzymes, namely ΔN_{123} -GBD-CD2, is that used for the efficient glucosylation of the lightly protected CD' (Figure 1) disaccharide acceptor to produce a *S. flexneri* 2a oligosaccharide precursor. (Salamone et al. 2015a). Furthermore, the X-ray structure of this branching sucrose revealed an exposed active site (Brison et al. 2012, 2016) -much less buried than that of an amylosucrase from GH13 family used in our prior studies (Champion et al. 2009a, 2012; Vergès et al. 2015)- and thus more compatible with larger protected oligosaccharide accommodation. All these reasons prompted us to test a set of six recently described sucrose-utilizing branching sucrases (Vuillemin et al. 2016) for tetrasaccharide **12** glucosylation. Three of them (BRS-A, ΔN_{123} -GBD-CD2, BRS-D) branch dextrans through α -1,2 osidic linkages whereas the three others (BRS-B Δ 1, BRS-C, BRS-E) add glucosyl units only via α -1,3 linkages.

After an optimization of reaction conditions (for which the procedure is presented in Complementary Results of this chapter), the enzymes were screened using 1 M of sucrose as glucosyl donor and 50 mM of tetrasaccharide **12** as acceptor, in 50 μ L reaction volume at pH 5.1. This pH was chosen to be a compromise between enzyme activity (optimal pH at 5.75) and acceptor stability, as tetrasaccharide **12** was shown to be more stable at pH between 4 and 5 (See Complementary Results of this chapter, Figure C3). HPAEC-PAD analyses confirmed that sucrose depletion was quasi-total in all reactions (data not shown) after 16h of reaction and RP-HPLC-UV-MS analyses revealed that all enzymes could transfer glucosyl units issued from sucrose onto acceptor **12** ($t_R = 22.5$ min). Two main products, called **P1** ($t_R = 21.65$ min) and **P2** ($t_R = 21.9$ min), were shown to have a molecular mass m/z of 1057.24, corresponding to NH_4^+ adducts of mono-glucosylated forms of tetrasaccharide **12** (Figure 2A). All enzymes also produced several poly-glucosylated products containing up to 6 glucosyl units as inferred from mass spectrometry analysis (Figure 3). HPLC-UV detection further indicates that enzymes such as ΔN_{123} -GBD-CD2 and BRS-D are the more prone to synthesize longer oligosaccharides.

Tetrasaccharide conversion was determined both from acceptor glucosylation (determined from the peak area of the glucosylated products detected by RP-HPLC-UV-MS) or from acceptor consumption (Table S1). In the first case, values ranging from 2.5 to 22.8 % were obtained depending on the enzyme (Figure 3B). For the reactions performed with the α -1,3 branching sucrases (BRS-B Δ 1, BRS-C and BRS-E), comparable conversion values were calculated from acceptor **12** consumption. For the others, they were either close or higher than those estimated from the product analysis. This is particularly marked for the reactions catalyzed by BRS-D and ΔN_{123} -GBD-CD2 and could be attributed to the higher amounts of poly-glucosylated formed with these enzymes, which are difficult to quantify. However, BRS-E and BRS-B Δ 1 clearly stand as the most efficient enzymes for the tetrasaccharide conversion, between 2.3 and 9.3 times higher than those obtained with the other branching sucrases.

Moreover, these two enzymes are much more specific for the synthesis of the mono-glucosylated product **P1** while the other enzymes are more versatile and produce both P1, P2 and poly-glucosylated products.

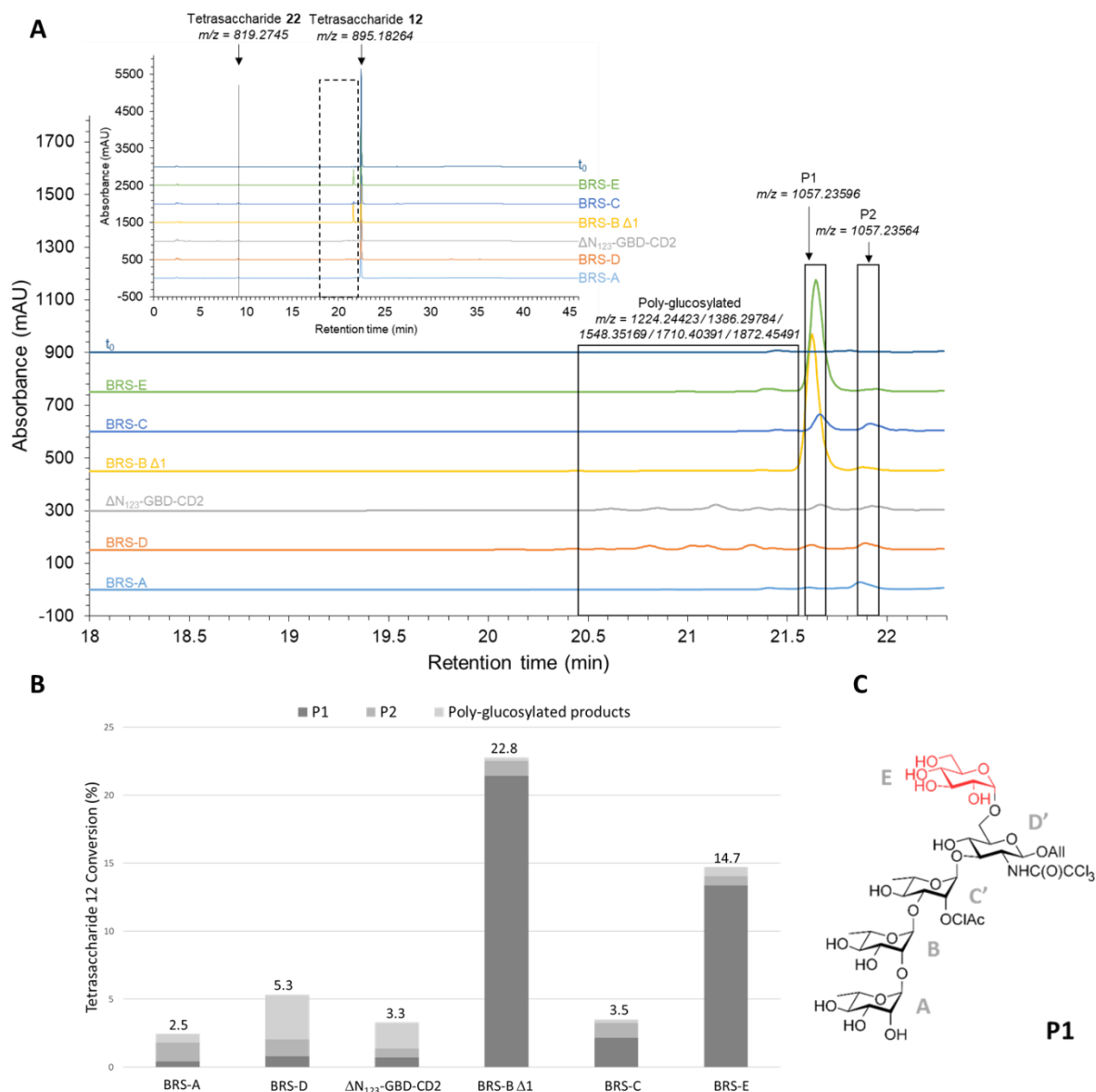


Figure 2: **A/** Chromatograms of the RP-HPLC-UV-MS analysis for the enzymatic glucosylation of tetrasaccharide **12** by six selected GH70 branching sucrases after 16 h of reaction in presence of sucrose (1 M) and tetrasaccharide **12** (50 mM) at pH 5.1. The region between 18.0 and 22.3 minutes corresponding to the frame in dash line has been enlarged for clarity purpose. The initial time (t_0) is indicated as reference. **P1** ($t_R = 21.65$ min), **P2** ($t_R = 21.9$ min) and poly-glucosylated products ($t_R = 20.6 / 20.8 / 21.0 / 21.1 / 21.15 / 21.2 / 21.3 / 21.45$ min) are products of the enzymatic glucosylation. A shift in molecular mass by 162 Da, characteristic of monoglucosylation, was observed for **P1** and **P2**. For poly-glucosylated products, incremental shifts of 162 Da were observed up to a shift of 1008 Da corresponding to 6 glucosyl units and reaching the sensitivity limit of the spectrometer. Tetrasaccharide **22** ($t_R = 22.5$ min), the dechloroacetylated form of **12**, was detected in the reaction mixture after 16 h of reaction, albeit only in non-interfering small amounts (< 2%); m/z are indicated below each product. **B/** Tetrasaccharide **12** conversion by the six enzymes, calculated as the ratio between the total area of P1, P2 and polyglucosylated products, and the area of tetrasaccharide **12** at the beginning of reaction. The relative proportions of P1, P2 and polyglucosylated products are represented in bar charts. **C/** Structure of **P1** characterized by NMR spectroscopy.

To produce **P1** at larger scale, a reaction in 1 mL volume was performed using a preparation of purified BRS-B $\Delta 1$ and an optimized acceptor **12**:sucrose ratio of 1:20 (carbohydrate total dry mass = 386 g/L). In these conditions, the conversion rate of tetrasaccharide **12** into pentasaccharide **P1** was 31% (Figure S1). NMR characterization of **P1** was based firstly on the comparison of the HSQC spectra (Figure 3A and B) of tetrasaccharide **12** and pentasaccharide **P1** as glucosylation was anticipated to result in shifted resonances at the glucosylated position and adjacent ones. Obviously, C-6_{D'} was shifted toward higher frequency, from 61.78 ppm to 65.87 ppm, while the adjacent 5_{D'} carbon was shifted to lower frequency, from 76.23 ppm to 74.51 ppm. In addition, the ¹H and ¹³C resonances corresponding to the neighboring 4_{D'} position slightly shifted, by around 0.1 ppm each. All data converged, revealing that glucosylation occurred on the D' moiety of tetrasaccharide **12**, yielding the ABC'[E α (1 \rightarrow 6)]D' pentasaccharide (Figures 2C, 3 and S3). Glucosylation at OH-6_{D'} is representative of the repeating units corresponding to *S. flexneri* serotypes 4a and 4b O-Ags (Perepelov et al., 2012) This finding demonstrates the potential of the devised synthetic strategy, opening new routes for the glucodiversification of tetrasaccharide **12** to access additional pentasaccharides representative of *S. flexneri* serotype-specific O-Ag repeating units. Unfortunately, pentasaccharide **P2** was produced in trace amounts, thus limiting its further characterization. Nonetheless, this original contribution sets the ground-work for further engineering of sucrose-utilizing branching sucrases to improve their catalytic performances and product yields, or expand their α -D-glucosylation regioselectivity, on the chemically synthesized core ABC'D' tetrasaccharide (**12**).

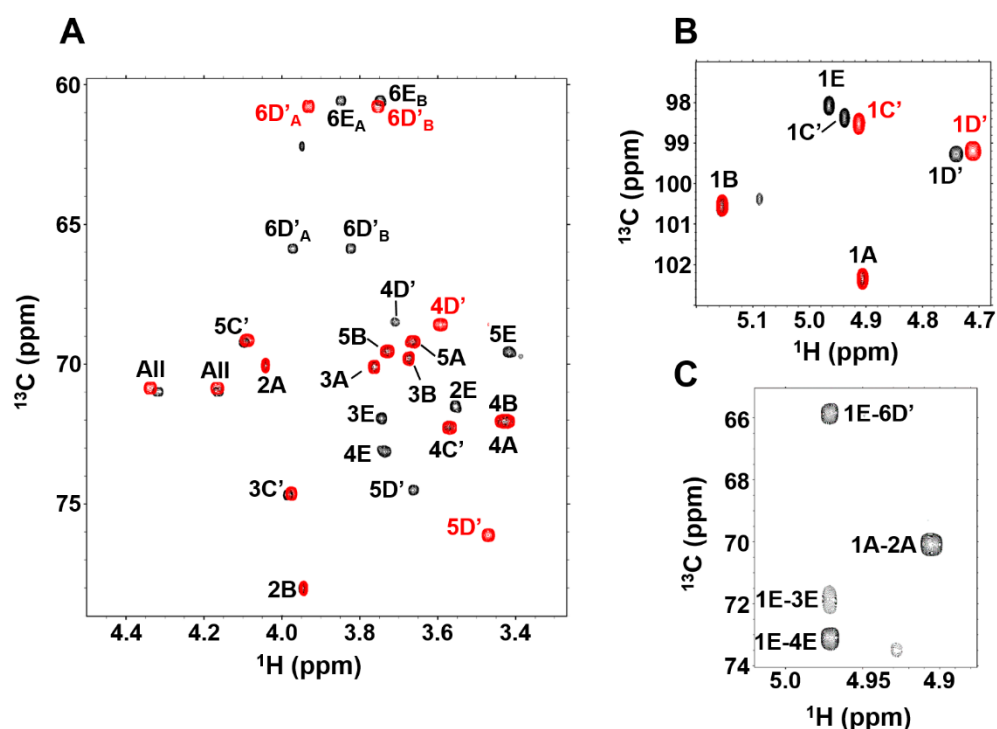


Figure 3: **A/** Superimposed ¹H-¹³C HSQC spectra of: **A/** the C2, C3, C4 and C5 resonances, and **B/** the anomeric region of pentasaccharide **P1** (black) and tetrasaccharide **12** (red). **C/** The ¹H-¹³C HMBC illustrates the correlation between the E and D' units of **P1**. All spectra were acquired at 950 MHz.

Conclusion

The concept validation exemplified herein provides a novel access to a large diversity of *S. flexneri* type-specific pentasaccharides from a simple non-natural lightly protected tetrasaccharide scaffold (**12**), representative of the backbone repeating units of most *S. flexneri* O-Ags. Designed for optimal manipulation compatible with chain elongation at both ends post-enzymatic glucosylation, tetrasaccharide **12** was obtained in 13 chemical steps from two easily available crystalline monosaccharides. Going beyond the synthesis of a panel of *S. flexneri* O-Ag fragments representing multiple different *Shigella* serotypes, this attractive diverging chemoenzymatic strategy offers new opportunities to provide libraries of complex microbial oligosaccharides derived from a common core, albeit differing by their branching pattern, whether in the form of glucosyl residues (Mann and Morenga 2015) or other substitutions. Glycan diversity characterizing the cell surface of numerous microbial species, many of which of vaccinal interest such as *Shigella*, calls for innovative ways toward expanding our arsenal of structurally well-defined oligosaccharides to decipher critical biological phenomena or for use in glycoarrays enabling improved diagnosis. Bacteria such as *Salmonella enterica* (Liu et al. 2014) or *Escherichia coli* (Stenutz et al. 2006) and fungi such as *Cryptococcus neoformans* (Cherniak and Sundstrom 1994) are other examples of families of pathogens featuring closely resembling polysaccharide antigens differentiating types within species. While the proof-of-concept is demonstrated herein for a small subset of available α -transglucosidases, impressive developments in glycoenzyme engineering technologies, that enable the tailoring on purpose of naturally available biocatalysts toward improved and/or novel substrate specificity for the high yielding conversion of a non-natural substrate into the targeted product (Benkoulouche et al. 2019) add to the exciting potential foreseen for the proposed approach.

Acknowledgements

This work was funded by the French National Research Agency (ANR Project CARBUNIVAX, ANR-15-CE07-0019-01), which included postdoctoral fellowships to Z.H. and A.B.I. as well as PhD fellowships to L.A.B (Médicament, Toxicologie, Chimie, Imageries (MTCI) doctoral school) and M.B. (Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB) doctoral school). Y.L.G. and G.L.H. are thankful to the Médicament, Toxicologie, Chimie, Environnement (MTCE) and MTCI doctoral schools, for their ministerial PhD fellowships, respectively. Financial support from the TGIR- RMN-THC Fr3050 CNRS (Gif-sur-Yvette) for the NMR experiments is gratefully acknowledged as well as N. Morellet for helpful advice on NMR experiments. The authors also thank D. Guieysse (TBI, Toulouse) for his help in NMR studies. The authors are grateful to N. Monties and E. Severac for providing their help in the analytical method development and product analysis, and purification by LC. The authors thank Floriant Bellvert and Hanna Barbier for granting access to HRMS facility of the MetaToul platform of the Toulouse Biotechnology Institute (Toulouse, France), and performing the

product analysis. The Authors also thank Frédéric Bonhomme (UMR3523 CNRS, Institut Pasteur) for providing the HRMS analyses.

Supporting Materials & Methods

Chemical synthesis

General procedures. Nuclear magnetic resonance (NMR) spectra were recorded at 303 K on a Bruker Avance spectrometer at 400 MHz (^1H) and 100 MHz (^{13}C) equipped with a BBO probe or at 298 K on a Bruker Avance spectrometer at 800 MHz (^1H) and 200 MHz (^{13}C) equipped with a QCI CryoProbe®. Spectra were recorded in deuterated chloroform (CDCl_3) or in deuterated water (D_2O). Elucidations of chemical structures was based on ^1H , COSY, DEPT-135, HSQC, decoupled HSQC, ^{13}C , decoupled ^{13}C and HMBC experiments. Signals are reported as m (multiplet), s (singlet), d (doublet), t (triplet), dd (doublet of doublet), q (quadruplet), dt (doublet of triplet), dq (doublet of quadruplet), ddd (doublet of doublet of doublet). Signals can also be described with a prefix b (broad). Coupling constants are reported in hertz (Hz). Chemical shifts are reported in ppm (δ) relative to the residual solvent peak. Sugar residues are lettered according to the lettering of the repeating unit of the *S. flexneri* O-Ag (**A**, **B**, **C**, **D**, **E**) and identified by a subscript in the listing of signal assignments. HRMS spectra were recorded on a WATERS QTOF Micromass instrument in the positive-ion electrospray ionization (ESI⁺) mode. Solutions were prepared using 1:1 MeCN/H₂O containing 0.1% formic acid, 7:3 MeCN/aq. MeOH containing 10 mM CH₃COONH₄. Analytical RP-HPLC was performed using an Akzo Nobel Kromasil® column, C₁₈, 4.6 x 150 mm, 3.5 μm , 100 Å, eluting at 1.0 mL.min⁻¹ with MeCN in TFA 0.08% 0 to 40% for 20 min, 15 °C. Detection was by UV at $\lambda = 210, 230$ or 254 nm, and ELSD.

Synthesis of trisaccharide 10

Allyl 4-O-benzyl-2-O-chloroacetyl- α -L-rhamnopyranoside (2). Diol **1** (1.07 g, 3.65 mmol) was dissolved in anhyd. MeCN (750 mL) then 2-chloro-1,1,1-trimethoxyethane (1.0 mL, 7.42 mmol, 2.0 equiv.) and PTSA (170 mg, 0.89 mmol, 0.2 equiv.) were added. After stirring for 40 min at rt, a TLC control (Tol/EtOAc 8:2) showed the total conversion of the starting diol **1** (R_f 0.11) into a less polar product (R_f 0.80). 80% aq. AcOH (3 mL) was then added and the solution was stirred for 25 min at rt. A TLC control (Toluene/EtOAc 8:2) showed the appearance of a more polar product (R_f 0.53). The mixture was diluted with H₂O (150 mL) then extracted with EtOAc (300 mL). The organic phase was washed with H₂O (100 mL) and brine (150 mL) then dried over anhyd. Na₂SO₄ and concentrated to dryness to give alcohol **2** as a colorless oil. Acceptor **2** had ^1H NMR (CDCl_3) δ 7.41-7.26 (m, 5H, H_{Ar}), 5.89 (m, 1H, CH=CH₂), 5.30 (m, 1H, $J_{\text{trans}} = 17.3$ Hz, $J_{\text{gem}} = 1.6$ Hz, CH=CH₂), 5.24-5.21 (m, 2H, H-2, CH=CH₂), 4.82 (d, 1H, $J_{1,2} = 1.2$ Hz, H-1), 4.81 (d, 1H, H_{Bn}), 4.75 (d, 1H, $J = 11.0$ Hz, H_{Bn}), 4.20-4.15 (m, 4H, H_{All}, H-3,

CH₂Cl), 4.00 (m, 1H, H_{All}), 3.80 (dq, 1H, H-5), 3.37 (pt, $J_{3,4} = J_{4,5} = 9.4$ Hz, H-4), 2.05 (d, 1H, $J_{3,OH} = 4.4$ Hz, OH), 1.38 (d, 3H, $J_{5,6} = 6.3$ Hz, H-6). ¹³C NMR (CDCl₃) δ 167.1 (CO_{ClAc}), 138.1 (C_{IVAr}), 133.3 (CH=CH₂), 128.7, 128.0, 127.9 (3C, C_{Ar}), 117.7 (CH=CH₂), 96.1 (C-1, ¹J_{C,H} = 171.0 Hz), 81.5 (C-4), 75.3 (C_{Bn}), 74.4 (C-2), 70.1 (C-3), 68.2 (C_{All}), 67.6 (C-5), 40.8 (CH₂Cl), 18.0 (C-6). HRMS (ESI⁺): *m/z* 393.1069 (calcd for C₁₈H₂₃ClO₆Na [M+Na]⁺ *m/z* 393.1081)

Allyl 3,4-di-O-benzyl-2-O-levulinoyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-4-O-benzyl-2-O-chloroacetyl- α -L-rhamnopyranoside (4). The crude material from above was taken into anhyd. Et₂O (20 mL) and donor **3** (2.64 g, 4.51 mmol, 1.3 equiv.) was added. The suspension was stirred at rt for 25 min then cooled down to -15 °C and TMSOTf was added (40 μ L, 0.22 mmol, 0.06 equiv.). After stirring for 25 min at this temperature, a TLC control (Toluene/EtOAc 8:2) showed the total conversion of the intermediate alcohol **2** (Rf 0.38) and the presence of a less polar product (Rf 0.45). The mixture was neutralized by addition of Et₃N, and the suspension was filtered over a pad of Celite®. Volatiles were removed under reduced pressure and the residue was purified by flash chromatography (Cyclohexane/Acetone 9:1 to 0:1) to give disaccharide **4** (2.63 g, 91%) as a colorless oil. The coupling product had ¹H NMR (CDCl₃) δ 7.38-7.25 (m, 15H, H_{Ar}), 5.88 (m, 1H, CH=CH₂), 5.41 (dd, 1H, $J_{1,2} = 1.9$ Hz, $J_{2,3} = 3.3$ Hz, H-2_B), 5.29 (m, 1H, $J_{trans} = 17.2$ Hz, $J_{gem} = 1.6$ Hz, CH=CH₂), 5.23-5.20 (m, 2H, H-2_C, CH=CH₂), 5.04 (d, 1H, H-1_B), 4.91 (d, 1H, $J = 11.3$ Hz, H_{Bn}), 4.81 (d, 1H, $J = 10.9$ Hz, H_{Bn}), 4.80 (d_{po}, 1H, $J_{1,2} = 1.8$ Hz, H-1_C), 4.65 (d, 1H, H_{Bn}), 4.61 (d_{po}, 1H, $J = 11.3$ Hz, H_{Bn}), 4.59 (d_{po}, 1H, H_{Bn}), 4.46 (d, 1H, $J = 11.3$ Hz, H_{Bn}), 4.20 (dd, 1H, $J_{2,3} = 3.2$ Hz, H-3_C), 4.15 (m, 1H, H_{All}), 4.11 (s, 2H, CH₂Cl), 3.99 (m, 1H, H_{All}), 3.87 (dd, 1H, H-3_B), 3.82-3.72 (m, 2H, H-5_B, H-5_C), 3.45 (pt, $J_{3,4} = J_{4,5} = 9.5$ Hz, H-4_C), 3.41 (pt, $J_{3,4} = J_{4,5} = 9.2$ Hz, H-4_B), 2.75-2.63 (m, 4H, CH_{2Lev}), 2.17 (s, 3H, CH_{3Lev}), 1.30 (d, 3H, $J_{5,6} = 6.3$ Hz, H-6_B), 1.29 (d, 3H, $J_{5,6} = 6.4$ Hz, H-6_C). ¹³C NMR (CDCl₃) δ 206.0 (CO_{Lev}), 171.8 (CO_{2Lev}), 166.7 (CO_{ClAc}), 138.5, 138.0, 137.9 (3C, C_{IVAr}), 133.3 (CH=CH₂), 129.0-127.6 (15C, C_{Ar}), 117.6 (CH=CH₂), 99.6 (C-1_B, ¹J_{C,H} = 172.8 Hz), 97.6 (C-1_C, ¹J_{C,H} = 171.0 Hz), 80.2 (C-4_B), 79.7 (C-4_C), 77.4 (C-3_B), 77.0 (C-3_C), 75.5, 75.0 (2C, C_{Bn}), 74.0 (C-2_C), 71.6 (C_{Bn}), 69.3 (C-2_B), 68.7 (C-5_B), 68.2 (C_{All}), 67.9 (C-5_C), 40.8 (CH₂Cl), 38.0 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.2 (CH_{2Lev}), 17.9 (2C, C-6_B, C-6_C). HRMS (ESI⁺): *m/z* 817.3044 (calcd for C₄₃H₅₁ClO₁₂Na [M+Na]⁺ *m/z* 817.2967)

3,4-Di-O-benzyl-2-O-levulinoyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-4-O-benzyl-2-O-chloroacetyl- α/β -L-rhamnopyranose (5). [Ir] (51 mg, 59 μ mol, 0.02 equiv.) was dissolved in anhyd. THF (10 mL) and the solution was degassed repeatedly then stirred under a satd. H₂ atmosphere for 25 min. The resulting yellow solution was degassed repeatedly before being poured into a solution of disaccharide **4** (1.97 g, 2.48 mmol) in anhyd. THF (35 mL). The mixture was stirred under an Ar atmosphere at rt for 17 h. A TLC control (Toluene/EtOAc 8:2) indicated the total conversion of the starting allyl glycoside (Rf 0.60) into a less polar intermediate (Rf 0.70). A solution of I₂ (1.25 g, 4.63 mmol, 2.0 equiv.) in THF/H₂O (5:1, 36 mL)

was added. After stirring for 2 h at rt, a TLC control (Toluene/EtOAc 8:2) showed the conversion of the intermediate compound into a more polar product (Rf 0.35). The reaction was quenched with 10% aq. NaHSO₃ (100 mL). The mixture was concentrated under reduced pressure, and extracted with DCM (3 × 250 mL). The combined organic phases were washed with brine (150 mL), dried over anhyd. Na₂SO₄ and passed through a phase separator filter. The residue was purified by flash chromatography (Toluene/EtOAc 8:2 to 5:5) to give hemiacetal **5** (1.56 g, 83%) as a pale yellow oil (α/β 9:1). The α -anomer had ¹H NMR (CDCl₃) δ 7.39-7.25 (m, 15H, H_{Ar}), 5.41 (dd, 1H, $J_{1,2} = 1.8$ Hz, $J_{2,3} = 3.2$ Hz, H-2_B), 5.23 (dd, 1H, $J_{1,2} = 1.9$ Hz, $J_{2,3} = 3.2$ Hz, H-2_C), 5.18 (dd, 1H, $J_{1,OH} = 4.0$ Hz, H-1_C), 5.05 (d, 1H, H-1_B), 4.90 (d, 1H, $J = 11.2$ Hz, H_{Bn}), 4.81 (d, 1H, $J = 10.9$ Hz, H_{Bn}), 4.65 (d, 1H, H_{Bn}), 4.60 (d_{po}, 1H, H_{Bn}), 4.59 (d_{po}, 1H, H_{Bn}), 4.46 (d, 1H, $J = 11.4$ Hz, H_{Bn}), 4.24 (dd, 1H, H-3_C), 4.11 (s, 2H, CH₂Cl), 3.99 (dq, 1H, H-5_C), 3.86 (dd, 1H, H-3_B), 3.76 (dq, 1H, H-5_B), 3.46 (pt, $J_{3,4} = J_{4,5} = 9.5$ Hz, H-4_C), 3.41 (pt, $J_{3,4} = J_{4,5} = 9.4$ Hz, H-4_B), 2.75-2.63 (m, 5H, OH, CH_{2Lev}), 2.17 (s, 3H, CH_{3Lev}), 1.30-1.28 (m, 6H, H-6_B, H-6_C). ¹³C NMR (CDCl₃) δ 206.2 (CO_{Lev}), 171.8 (CO_{2Lev}), 166.8 (CO_{ClAc}), 138.4, 138.0, 137.9 (3C, C_{IVAr}), 128.5-127.6 (15C, C_{Ar}), 99.4 (C-1_B, $^1J_{C,H} = 171.8$ Hz), 91.3 (C-1_C, $^1J_{C,H} = 172.2$ Hz), 80.3 (C-4_B), 79.7 (C-4_C), 77.4 (C-3_B), 76.2 (C-3_C), 75.4, 75.1 (2C, C_{Bn}), 74.3 (C-2_C), 71.6 (C_{Bn}), 69.3 (C-2_B), 68.7 (C-5_B), 67.8 (C-5_C), 40.8 (CH₂Cl), 38.0 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.1 (CH_{2Lev}), 17.9 (2C, C-6_B, C-6_C).

3,4-Di-O-benzyl-2-O-levulinoyl- α -L-rhamnopyranosyl-(1→3)-4-O-benzyl-2-O-

chloroacetyl- α/β -L-rhamnopyranosyl trichloroacetimidate (6). Hemiacetal **5** (1.39 g, 1.84 mmol) was dissolved in anhyd. DCE (15 mL). CCl₃CN (0.55 mL, 5.49 mmol, 3.0 equiv.) and DBU (83 μ L, 0.56 mmol, 0.3 equiv.) were added at rt. After stirring for 1 h under Ar at rt. Follow up by TLC (Tol/EtOAc 8:2 + 1% Et₃N) showed the conversion of the starting hemiacetal **5** (Rf 0.26) into a less polar product (Rf 0.64). The mixture was concentrated and directly purified by flash chromatography (Tol/EtOAc 95:5 to 0:1 + 1% Et₃N) to give donor **6** (1.59 g, 96 %, α/β 95:5) as a brownish oil. Trichloroacetimidate **6** had ¹H NMR (CDCl₃) δ 8.73 (bs, 1H, NH), 7.40-7.26 (m, 15H, H_{Ar}), 6.23 (d, 1H, $J_{1,2} = 1.8$ Hz, H-1_C), 5.44 (dd, 1H, $J_{1,2} = 1.8$ Hz, $J_{2,3} = 3.2$ Hz, H-2_B), 5.38 (dd, 1H, $J_{2,3} = 3.3$ Hz, H-2_C), 5.10 (d, 1H, H-1_B), 4.91 (d, 1H, $J = 11.3$ Hz, H_{Bn}), 4.82 (d, 1H, $J = 10.9$ Hz, H_{Bn}), 4.67-4.60 (m, 3H, H_{Bn}), 4.51 (d, 1H, $J = 11.3$ Hz, H_{Bn}), 4.28 (dd, 1H, H-3_C), 4.13 (d, 2H, CH₂Cl), 3.97 (dq, 1H, H-5_C), 3.87 (dd, 1H, H-3_B), 3.79 (dq, 1H, H-5_B), 3.57 (pt, $J_{3,4} = J_{4,5} = 9.5$ Hz, H-4_C), 3.43 (pt, $J_{3,4} = J_{4,5} = 9.4$ Hz, H-4_B), 2.73-2.65 (m, 4H, CH_{2Lev}), 2.17 (s, 3H, CH_{3Lev}), 1.34 (d, 3H, $J_{5,6} = 6.3$ Hz, H-6_C), 1.27 (d, 3H, $J_{5,6} = 6.2$ Hz, H-6_B). HRMS (ESI⁺): m/z 915.2114 (calcd for C₄₂H₄₇Cl₄O₁₂NH₄ [M+NH₄]⁺ m/z 915.2196).

Allyl 3,4-di-O-benzyl-2-O-levulinoyl- α -L-rhamnopyranosyl-(1→3)-4-O-benzyl-2-O-chloroacetyl- α -L-rhamnopyranosyl-(1→3)-6-O-benzyl-2-deoxy-2-trichloroacetamido- β -D-glucopyranoside (8) and 3,4-di-O-benzyl-2-O-levulinyl- α -L-rhamnopyranosyl-(1→3)-4-O-benzyl- α -L-rhamnopyranose 1,2-(6-O-benzyl-2-deoxy-2-trichloroacetamido- β -D-

glucopyranosid)-3-yl) orthochloroacetate (11). Alcohol **7** (630 mg, 1.39 mmol) and donor **6** (1.45 mg, 1.61 mmol, 1.2 equiv.) were dissolved in DCM (15 mL), and activated 4 Å MS (1.0 g) was added. The suspension was stirred for 20 min at rt, then cooled to -15 °C, and TMSOTf (25 µL, 138 µmol, 0.1 equiv.) was added. After stirring for 45 min at -20 °C, a TLC control (Tol/EtOAc 8:2) showed the disappearance of acceptor **7** (Rf 0.24) and the presence of two less polar products (Rf 0.4 and 0.49). The mixture was neutralized by addition of Et₃N, and the suspension was filtered over a pad of Celite®. Volatiles were removed under reduced pressure and the crude material was purified by flash chromatography (Toluene/EtOAc 95:5 to 0:1) to provide in order of elution first the orthoester **11** (441 mg, 27%) then trisaccharide **8** (1.02 g, 62%) both as a white foam. The former had ¹H NMR (CDCl₃) δ 7.64-7.61 (m, 2H, H_{Ar}), 7.41-7.18 (m, 18H, H_{Ar}), 7.07 (d, 1H, *J*_{2,NH} = 7.2 Hz, NH), 5.89 (m, 1H, CH=CH₂), 5.50 (s, 1H, H_{Bzl}), 5.39 (dd, 1H, *J*_{1,2} = 1.4 Hz, *J*_{2,3} = 3.3 Hz, H-2_B), 5.30 (m, 1H, *J*_{trans} = 17.2 Hz, *J*_{gem} = 1.5 Hz, CH=CH₂), 5.23 (m, 1H, *J*_{cis} = 10.3 Hz, CH=CH₂), 5.07 (d, 1H, *J*_{1,2} = 8.2 Hz, H-1_D), 4.93 (bs, 1H, H-1_B), 4.87 (d, 1H, *J* = 10.7 Hz, H_{Bn}), 4.84-4.79 (m, 4H, H-1_C, H-3_D, 2H_{Bn}), 4.63 (d, 1H, *J* = 10.8 Hz, H_{Bn}), 4.61-4.58 (m, 2H, H-2_C, H_{Bn}), 4.33 (m, 1H, H_{All}), 4.16-4.04 (m, 5H, H_{All}, H-6_{aD}, H-5_B, 2H_{Bn}), 3.86 (dd, 1H, H-3_B), 3.70 (s, 2H, CH₂Cl), 3.68 (dd_{po}, 1H, H-6_{bD}), 3.63 (dd_{po}, 1H, *J*_{2,3} = 4.2 Hz, H-3_C), 3.58 (m, 1H, H-5_D), 3.51 (pt, 1H, *J*_{3,4} = *J*_{4,5} = 8.9 Hz, H-4_D), 3.45 (pt, 2H, *J*_{3,4} = *J*_{4,5} = 9.3 Hz, H-4_B, H-4_C), 3.39 (ddd, 1H, H-2_D), 3.11 (dq, 1H, H-5_C), 2.78-2.64 (m, 4H, CH_{2Lev}), 2.18 (s, 3H, CH_{3Lev}), 1.48 (d, 3H, *J*_{5,6} = 6.2 Hz, H-6_B), 1.24 (d, 3H, *J*_{5,6} = 6.2 Hz, H-6_C). ¹³C NMR (CDCl₃) δ 206.2 (CO_{Lev}), 171.8 (CO_{2Lev}), 162.0 (NHCO), 138.7, 138.4, 137.7, 137.5 (4C, C_{IVAr}), 133.2 (CH=CH₂), 129.4-126.7 (18C, C_{Ar}), 120.6 (CO₂CH₂Cl), 118.5 (CH=CH₂), 101.6 (C_{Bzl}), 100.2 (C-1_B, ¹*J*_{C,H} = 171.5 Hz), 98.4 (C-1_D, ¹*J*_{C,H} = 165.9 Hz), 97.3 (C-1_C, ¹*J*_{C,H} = 179.5 Hz), 92.4 (CCl₃), 81.3 (C-4_D), 79.9 (C-4_B), 79.2 (C-2_C), 78.6 (C-4_C), 78.5 (C-5_B), 77.8 (C-3_B), 75.7, 75.3, 71.2 (3C, C_{Bn}), 70.9 (C_{All}), 70.7 (C-5_C), 69.9 (C-3_D), 69.5 (C-2_B), 68.6 (2C, C-5_B, C-6_D), 65.8 (C-5_D), 58.8 (C-2_D), 46.6 (CH₂Cl), 38.1 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.2 (CH_{2Lev}), 18.0 (C-6_B), 17.9 (C-6_C). HRMS (ESI⁺): *m/z* 1212.2871 (calcd for C₅₈H₆₅Cl₄O₁₇Na [M+Na]⁺ *m/z* 1210.2904)

Trisaccharide **8** had ¹H NMR (CDCl₃) δ 7.49-7.46 (m, 2H, H_{Ar}), 7.38-7.23 (m, 18H, H_{Ar}), 6.96 (d, 1H, *J*_{2,NH} = 7.4 Hz, NH), 5.88 (m, 1H, CH=CH₂), 5.55 (s, 1H, H_{Bzl}), 5.37 (dd, 1H, *J*_{1,2} = 1.9 Hz, *J*_{2,3} = 3.1 Hz, H-2_B), 5.30 (m, 1H, *J*_{trans} = 17.3 Hz, *J*_{gem} = 1.6 Hz, CH=CH₂), 5.24-5.21 (m, 2H, H-2_C, CH=CH₂), 5.07 (d, 1H, *J*_{1,2} = 8.2 Hz, H-1_D), 5.02 (d, 1H, H-1_B), 4.89 (d, 1H, *J* = 11.2 Hz, H_{Bn}), 4.85 (d, 1H, *J*_{1,2} = 1.5 Hz, H-1_C), 4.70 (d, 1H, *J* = 11.1 Hz, H_{Bn}), 4.63 (d, 1H, H_{Bn}), 4.57 (d_{po}, 1H, *J* = 11.3 Hz, H_{Bn}), 4.54 (m, 1H, H-3_D), 4.47 (d, 1H, H_{Bn}), 4.43 (d_{po}, 1H, H_{Bn}), 4.39 (dd_o, 1H, H-6_{aD}), 4.35 (m, 1H, H_{All}), 4.14-4.07 (m, 2H, H-3_C, H_{All}), 4.02 (s, 2H, CH₂Cl), 3.98 (dq_{po}, 1H, H-5_C), 3.82 (dd_{po}, 1H, H-3_B), 3.80-3.71 (m, 2H, H-6_{bD}, H-5_B), 3.63-3.55 (m, 2H, H-4_D, H-5_D), 3.45 (ddd, 1H, H-2_D), 3.39 (pt, 1H, *J*_{3,4} = *J*_{4,5} = 9.4 Hz, H-4_B), 3.30 (pt, 1H, *J*_{3,4} = *J*_{4,5} = 9.6 Hz, H-4_C), 2.72-2.62 (m, 4H, CH_{2Lev}), 2.17 (s, 3H, CH_{3Lev}), 1.28 (d, 3H, *J*_{5,6} = 6.1 Hz, H-6_B), 0.74 (d, 3H, *J*_{5,6} = 6.1 Hz, H-6_C). ¹³C NMR (CDCl₃) δ 206.2 (CO_{Lev}), 171.8 (CO_{2Lev}), 166.3 (CO_{ClAc}), 162.2 (NHCO), 138.6, 138.1, 138.0, 136.9 (4C, C_{IVAr}), 133.2 (CH=CH₂), 129.1-126.5

(18C, C_{Ar}), 118.4 (CH=CH₂), 102.1 (C_{Bzl}), 99.3 (C-1_B, ¹J_{C,H} = 171.8 Hz), 98.4 (C-1_D, ¹J_{C,H} = 164.1 Hz), 97.2 (C-1_C, ¹J_{C,H} = 172.7 Hz), 92.2 (CCl₃), 80.2 (C-4_B), 80.1 (C-4_D), 79.7 (C-4_C), 77.4 (C-3_B), 76.3 (C-3_C), 75.1, 74.9 (2C, C_{Bn}), 74.2 (C-3_D), 73.2 (C-2_C), 71.5 (C_{Bn}), 70.8 (C_{All}), 69.3 (C-2_B), 68.7 (2C, C-5_B, C-6_D), 68.1 (C-5_C), 66.2 (C-5_D), 59.9 (C-2_D), 40.6 (CH₂Cl), 38.0 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.1 (CH_{2Lev}), 18.0 (C-6_B), 17.1 (C-6_C). HRMS (ESI⁺): *m/z* 1210.2820 (calcd for C₅₈H₆₅Cl₄O₁₇Na [M+Na]⁺ *m/z* 1210.2904)

Allyl 3,4-di-O-benzyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-4-O-benzyl-2-O-chloroacetyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-6-O-benzyl-2-deoxy-2-trichloroacetamido- β -D-

glucopyranoside (9). To a solution of trisaccharide **8** (430 mg, 0.36 mmol) in Pyridine (6 mL) stirred at rt under an Ar atmosphere were added AcOH (4 mL) and NH₂NH₂·H₂O (100 μ L, 2.06 mmol, 5.7 equiv.). After stirring for 35 min at rt. A TLC control (Toluene/Acetone 9:1) showed the conversion of the starting material (R_f 0.34) into a more polar product (R_f 0.22). Following addition of H₂O (100 mL) and DCM (50 mL), the two layers were separated and the aq. one was extracted with DCM (2 x 50 mL). The combined organic extracts were washed with brine (100 mL), dried over anhyd. Na₂SO₄ and concentrated to dryness. The residue was purified by flash chromatography (Toluene/EtOAc 85:15 to 0:1) to give alcohol **9** (342 g, 86%) as a white solid. The later had ¹H NMR (CDCl₃) δ 7.50-7.47 (m, 2H, H_{Ar}), 7.39-7.23 (m, 18H, H_{Ar}), 7.04 (d, 1H, *J*_{2,NH} = 7.4 Hz, NH), 5.88 (m, 1H, CH=CH₂), 5.55 (s, 1H, H_{Bzl}), 5.30 (m, 1H, *J*_{trans} = 17.2 Hz, *J*_{gem} = 1.5 Hz, CH=CH₂), 5.24-5.21 (m, 2H, H-2_C, CH=CH₂), 5.07 (d_o, 1H, H-1_B), 5.06 (d_{po}, 1H, *J*_{1,2} = 7.9 Hz, H-1_D), 4.86 (d, 1H, *J*_{1,2} = 1.6 Hz, H-1_C), 4.86 (d, 1H, *J* = 11.2 Hz, H_{Bn}), 4.66 (d, 1H, H_{Bn}), 4.64 (d_{po}, 1H, *J* = 11.8 Hz, H_{Bn}), 4.60 (bs, 2H, H_{Bn}), 4.58 (d_{po}, 1H, H_{Bn}), 4.54 (m, 1H, H-3_D), 4.49 (d, 1H, *J* = 11.0 Hz, H_{Bn}), 4.39 (dd_{po}, 1H, *J*_{5,6a} = 4.6 Hz, H-6a_D), 4.36 (m, 1H, H_{All}), 4.13-4.08 (m, 2H, H-3_C, H_{All}), 4.03 (s, 2H, CH₂Cl), 3.99 (dq_{po}, 1H, H-5_C), 3.91 (dd, 1H, *J*_{1,2} = 1.7 Hz, *J*_{2,3} = 3.1 Hz, H-2_B), 3.81 (pt, 1H, H-6b_D), 3.75-3.69 (m, 2H, H-3_B, H-5_B), 3.64-3.55 (m, 2H, H-4_D, H-5_D), 3.48 (ddd_{po}, 1H, H-2_D), 3.44 (pt, 1H, *J*_{3,4} = *J*_{4,5} = 9.4 Hz, H-4_B), 3.30 (pt, 1H, *J*_{3,4} = *J*_{4,5} = 9.6 Hz, H-4_C), 2.36 (bs, 1H, OH), 1.28 (d, 3H, *J*_{5,6} = 6.3 Hz, H-6_B), 0.77 (d, 3H, *J*_{5,6} = 6.2 Hz, H-6_C). ¹³C NMR (CDCl₃) δ 166.3 (CO_{ClAc}), 162.2 (NHCO), 138.5, 138.1, 138.0, 137.0 (4C, C_{IVAr}), 133.2 (CH=CH₂), 129.1-126.4 (18C, C_{Ar}), 118.4 (CH=CH₂), 102.1 (C_{Bzl}), 101.2 (C-1_B, ¹J_{C,H} = 170.8 Hz), 98.3 (C-1_D, ¹J_{C,H} = 165.5 Hz), 97.2 (C-1_C, ¹J_{C,H} = 172.1 Hz), 92.2 (CCl₃), 80.4 (C-4_C), 80.1 (C-4_D), 79.7 (C-4_B), 79.4 (C-3_B), 76.4 (C-3_C), 75.1, 75.0 (2C, C_{Bn}), 74.1 (C-3_D), 73.7 (C-2_C), 72.1 (C_{Bn}), 70.8 (C_{All}), 69.0 (C-2_B), 68.7 (C-6_D), 68.4 (C-5_B), 68.1 (C-5_C), 66.2 (C-5_D), 60.0 (C-2_D), 40.6 (CH₂Cl), 17.9 (C-6_B), 17.1 (C-6_C). HRMS (ESI⁺): *m/z* 1112.2651 (calcd for C₅₃H₅₉Cl₄O₁₅Na [M+Na]⁺ *m/z* 1112.2537).

Propyl α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-acetyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-acetamido-2-deoxy- β -D-glucopyranoside (10). To a stirred solution of trisaccharide **9** (76.9 mg, 70 μ mol) in 96% aq. EtOH (6 mL) and DCE (150 μ L), were added Pd/C (75.6 mg). The suspension was stirred under H₂ atmosphere for 24 h at rt. After this time, a solution of

Na₂HPO₄ (0.5 M, 500 µL) was added and the suspension was stirred for another 3 d at rt. MS analysis of the reaction mixture revealed a molecular weight of corresponding to that of the target tetrasaccharide. The reaction mixture was passed through a 0.2 µm filter. After freeze-drying, purification of the residue by RP-LC (215 nm, 0-40% linear gradient of MeCN/H₂O over 27 CV at a flow rate of 30 mL•min⁻¹), gave trisaccharide **10** (28.7 mg, 47%) as a fluffy solid after repeated freeze-drying. Propyl glycoside **10** had ¹H NMR (D₂O) δ 5.00-4.96 (m, 2H, H-2_C, H-1_B), 4.86 (bs, 1H, H-1_C), 4.50 (d, 1H, *J*_{1,2} = 8.5 Hz, H-1_D), 4.08 (dq, 1H, H-5_C), 4.01 (bd, 1H, H-2_B), 3.96-3.89 (m, 2H, H-6_{aD}, H-3_C), 3.85-3.69 (m, 3H, OCH_{2Pr}, H-2_D, H-6_{bD}), 3.65 (dd, 1H, *J*_{2,3} = 2.7 Hz, *J*_{3,4} = 9.4 Hz, H-3_B), 3.60-3.40 (m, 7H, H-4_C, OCH_{2Pr}, H-3_D, H-4_D, H-5_B, H-4_B), 2.16 (s, 3H, CH_{3Ac}), 2.05 (s, 3H, CH_{3NHAc}), 1.54 (psex, 2H, *J* = 6.9 Hz, CH_{2Pr}), 1.25-1.22 (m, 6H, H-6_B, H-6_C), 0.86 (t, 3H, CH_{3Pr}). ¹³C NMR (D₂O) δ 177.1 (NHCO), 175.8 (CO_{Ac}), 105.4 (C-1_B, ¹*J*_{C,H} = 171.3 Hz), 103.3(C-1_D, ¹*J*_{C,H} = 163.5 Hz), 101.1 (C-1_C, ¹*J*_{C,H} = 173.4 Hz), 85.1 (C-3_D), 79.8 (C-3_C), 78.7 (C-5_D), 75.1 (C-2_C), 74.9 (OCH_{2Pr}), 74.4 (C-4_B), 73.9 (C-4_C), 72.8 (C-3_B), 72.7 (C-2_B), 72.0 (C-5_B), 71.6 (C-5_C), 71.1 (C-4_D), 63.5 (C-6_D), 57.9 (C-2_D), 25.0 (CH_{3NHAc}), 24.7 (CH_{2Pr}), 22.9 (CH_{3Ac}), 19.2, 19.0 (2C, C-6_C, C-6_B), 12.3 (CH_{3Pr}). HRMS (ESI⁺): *m/z* 620.2496 (calcd for C₂₅H₄₃NO₁₅Na [M+Na]⁺ *m/z* 620.2531); *m/z* 598.2674 (calcd for C₂₅H₄₄NO₁₅ [M+H]⁺ *m/z* 598.2711). RP-HPLC (215 nm): *R*_t = 9.82 min.

Chemical synthesis of tetrasaccharide **12**

Allyl 4-O-(2-naphthylmethyl)-α-L-rhamnopyranoside (16). Acetyl chloride (50 mL, 693 mmol, 2.5 equiv.) was added dropwise to AlIOH (610 mL) at 0 °C, the solution was stirred for 25 min, and then L-rhamnose monohydrate (50 g, 277 mmol) was added. The mixture was heated for 2.5 h at 70 °C then for 15 h at 40 °C. Follow up by TLC (DCM/MeOH 8:2) indicated the total conversion of the starting hemiacetal (*R*_f = 0.2) into a less polar product (*R*_f = 0.7). The bath temperature was cooled to 0 °C and the solution was neutralized by addition of NaHCO_{3(s)} (103 g). The suspension was filtered over a pad of Celite® and solvents were evaporated and co-evaporated three times with Toluene. The crude material was dissolved in anhyd. acetone (300 mL) then 2,2-dimethoxypropane (100 mL, 0.81 mol, 3.0 equiv.) and PTSA (3.04 g, 16 mmol, 0.05 equiv.) were successively added. After stirring for 3 h at rt, TLC (DCM/MeOH 9:1) indicated the total conversion of the intermediate allyl glycoside (*R*_f = 0.3) into a less polar product (*R*_f = 0.6). The solution was neutralized by adding Et₃N (4 mL) and solvents were evaporated under reduced pressure. The residue was dissolved in DCM (600 mL) and washed with H₂O (3 x 300 mL) and brine (200 mL). The organic layer was dried by passing through phase separator filter and concentrated to dryness. The crude residue was dissolved in DMF (800 mL) under an Ar atmosphere, the bath temperature was cooled to -5 °C and NaH (60% oil dispersion, 29.1 g, 0.73 mol, 2.4 equiv.) was added portion wise to this suspension. The mixture was stirred for 2 h at rt under Ar and then 2-(bromomethyl)naphthalene (73.5 g, 0.33 mol, 1.2 equiv.) was added portion wise at -5 °C. After stirring the RM for 2 h under Ar at rt, a

TLC control (cHex/EtOAc 7:3) indicated the total conversion of the intermediate alcohol ($R_f = 0.3$) into a less polar product ($R_f = 0.67$). The reaction was quenched at 0 °C by addition of MeOH (50 mL). Solvents were eliminated under reduced pressure and volatiles were co-evaporated with Tol. The residue was taken in EtOAc (400 mL) and washed with H₂O (3 x 300 mL) and brine (150 mL). The organic phase was dried over anhyd. Na₂SO₄, filtered, and concentrated to dryness. The crude residue was dissolved in 80% aq. AcOH (500 mL) and the solution was stirred for 20 h at 80 °C. A TLC control (Cyclohexane/EtOAc 5:5) indicated the total conversion of the intermediate ($R_f = 1.0$) into a more polar product ($R_f = 0.2$). Solvents were removed under vacuum and traces of AcOH were eliminated by co-evaporation with Toluene (3 x 400 mL). The residue was taken in MeCN (500 mL) and washed with cold Cyclohexane (2 x 300 mL). The MeCN layer was evaporated to dryness and recrystallization in hot Cyclohexane was performed to afford the expected diol **16** (82.5 g, 87%) as a white to off-white crystal. ¹H NMR (400 MHz, CDCl₃) δ 7.87 - 7.78 (m, 4H, H_{ArNap}), 7.52 - 7.44 (m, 3H, H_{ArNap}), 5.89 (dddd, 1H, $J = 17.2, 10.4, 6.0, 5.2$ Hz, CH=CH_{2All}), 5.28 (dq_{app}, 1H, $J = 17.2, 1.5$ Hz, CH=CH_{2All}), 5.19 (dq_{app}, 1H, $J = 10.4, 1.5$ Hz, CH=CH_{2All}), 4.95 - 4.86 (m, 2H, H_{ArNap}), 4.81 (d, $J = 1.4$ Hz, 1H, H-1), 4.17 (ddt, 1H, $J = 12.9, 5.1, 1.5$ Hz, 1H, CH₂-CH=All), 4.02 - 3.93 (m, 3H, CH₂-CH=All, H-2, H-3), 3.79 (dq, 1H, $J = 9.2, 6.3$ Hz, H-5), 3.41 (t_{app}, 1H, $J = 9.2$ Hz, H-4), 2.45 (brs, 2H, OH-2, OH-3), 1.38 (d, 3H, $J = 6.3$ Hz, H-6). ¹³C NMR (100 MHz, CDCl₃) δ 133.8 (CH=CH_{2All}), 128.4 (C_{IVAr}), 128.0 (C_{IVAr}), 127.7 (C_{IVAr}), 126.7-125.8 (7C, CH_{Ar}), 117.4 (CH=CH_{2All}), 98.5 (C-1, ¹J_{C,H} = 170.1 Hz), 75.1 (CH_{2Nap}), 71.6, 71.3 (2C, C-2, C-3), 68.0 (CH₂-CH=All), 67.3 (C-5), 18.1 (C-6). HRMS (ESI⁺): m/z 362.1985 (calcd for C₁₆H₂₂O₅NH₄⁺ [M+NH₄]⁺ m/z 362.1967); m/z 367.1576 (calcd for C₁₆H₂₂O₅Na⁺ [M+Na]⁺ m/z 367.1521). mp 347.95 K

Allyl 4-O-(2-naphthylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranoside (S1). To a solution of diol **16** (20.0 g, 58.1 mmol) in anhyd. MeCN (800 mL) stirred at 0 °C under an Ar atmosphere were successively added dropwise TES-Cl (11.70 mL, 69.7 mmol, 1.2 equiv.) and DIPEA (15.2 mL, 87.1 mmol, 1.5 equiv.). After stirring the RM for 2 h at rt under Ar, TLC (Cyclohexane/EtOAc 8:2) showed complete consumption of the starting material ($R_f = 0.13$) and the presence of a main product ($R_f = 0.70$). MeOH (1.2 mL, 29.0 mmol, 0.5 equiv.) was added, the mixture was stirred for 15 min, and volatiles were evaporated under reduced pressure. Toluene was added and a precipitate appeared. The suspension was filtered and the filtrate was concentrated to dryness. The residue was purified by flash chromatography (Toluene/EtOAc 95:5 to 9:1) to give alcohol **S1** (19.7 g, 74%) as a yellow oil. The later had ¹H NMR (CDCl₃) δ 7.90-7.79 (m, 4H, H_{ArNap}), 7.53-7.45 (m, 3H, H_{ArNap}), 6.00-5.87 (m, 1H, CH=CH₂), 5.36-5.26 (dd, 1H, CH=CH₂), 5.25-5.18 (dd, 1H, CH=CH₂), 5.09-5.01 (d, 1H, CH_{2Nap}), 4.87 (s, 1H, H-1), 4.84-4.76 (d, 1H, CH_{2Nap}), 4.25-4.15 (ddt_{app}, 1H, CH_{2All}), 4.15-4.09 (dd, 1H, $J_{3,2} = 3.5$ Hz, H-3) 4.05-3.98 (ddt_{app}, 1H, CH_{2All}), 3.86 (bd, 1H, H-2), 3.84-3.73 (m, 1H, H-5), 3.44 (t, 1H, $J_{4,3} = J_{4,5} = 9.2$ Hz, H-4), 2.68 (bs, 1H, OH), 1.35-1.27 (d, 3H, H-6), 1.07-0.98 (t, 9H, CH_{3TES}), 0.78-0.68 (m, 6H, $J_{6,5} = 6.3$ Hz, CH_{2TES}). ¹³C NMR (CDCl₃) δ 136.0 (C_{IVNap}),

134.0 (CH=CH₂), 133.3 (C_{IVNap}), 133.0 (C_{IVNap}), 128.0-125.8 (C_{ArNap}), 117.0 (CH=CH₂), 98.0 (C-1, ¹J_{C,H} = 169.0 Hz), 81.3 (C-4), 75.5 (CH_{2Nap}), 73.3 (C-3), 72.2 (C-2), 67.8 (CH_{2All}), 67.4 (C-5), 17.9 (C-6), 6.8 (CH_{3TES}), 5.1 (CH_{2TES}). HRMS (ESI⁺): *m/z* 481.2386 (calcd for C₂₆H₃₈O₅SiNa [M+Na]⁺ *m/z* 481.2394).

Allyl 2-O-levulinoyl-4-O-(2-naphthylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranoside (S2). Alcohol **S1** (19.7 g, 43.0 mmol) was dissolved in anhyd. DCM (144 mL) and stirred at rt. DCC (17.7 g, 85.9 mmol, 2.0 equiv.), DMAP (4.2 g, 34.4 mmol, 0.8 equiv.) and levulinic acid (13.2 mL, 128.9 mmol, 3.0 equiv.) were successively added. After stirring for 2 h at rt, a TLC control (Cyclohexane/EtOAc 7:3) showed complete consumption of the starting material (R_f 0.6) and the presence of a more polar product (R_f 0.53). The reaction mixture was concentrated under reduced pressure. The crude material was taken in EtOAc (100 mL) and the resulting suspension was filtered on a pad of Celite®. H₂O (30 mL) was added to the filtrate and the organic layer was washed successively with 10% aq. CuSO₄ (30 mL), H₂O (30 mL), satd aq. NaHCO₃ (30 mL) and brine (30 mL). The organic layer was dried by stirring over anhyd. Na₂SO₄, filtered and concentrated under vacuum. The residue was purified by flash chromatography (Tol/EtOAc 8:2) to give the protected monosaccharide **S2** (17.7 g, 74%) as a yellow to brown oil. The later had ¹H NMR (CDCl₃) δ 7.90-7.79 (m, 4H, H_{ArNap}), 7.54-7.46 (m, 3H, H_{ArNap}), 5.97-5.84 (m, 1H, CH=CH₂), 5.33-5.26 (dd, 1H, CH=CH₂), 5.25-5.17 (dd, 1H, CH=CH₂), 5.11-5.06 (m, 2H, CH_{2Nap}, H-2), 4.84-4.78 (d, 1H, CH_{2Nap}), 4.72 (bd, 1H, H-1), 4.26-4.21 (dd, 1H, J_{3,2} = 3.6 Hz, J_{3,4} = 9.3 Hz, H-3), 4.20-4.13 (ddt_{app}, 1H, CH_{2All}), 4.01-3.95 (ddt_{app}, 1H, CH_{2All}), 3.82-3.73 (m, 1H, H-5), 3.42 (t, 1H, J_{4,5} = 9.4 Hz, H-4), 2.82-2.70 (m, 4H, CH_{2Lev}), 2.22 (s, 3H, CH_{3Lev}), 1.31 (d, 3H, H-6), 0.99 (t, 9H, J_{6,5} = 6.3 Hz, CH_{3TES}), 0.68 (q, 6H, CH_{2TES}). ¹³C NMR (CDCl₃) δ 206.1 (CO_{Lev}), 172.1 (CO_{2Lev}), 136.1 (C_{IVNap}), 133.7 (CH=CH₂), 133.3 (C_{IVNap}), 133.0 (C_{IVNap}), 128.0-125.8 (C_{ArNap}), 117.0 (CH=CH₂), 96.6 (C-1, ¹J_{C,H} = 170.6 Hz), 81.4 (C-4), 75.5 (CH_{2Nap}), 73.5 (C-2), 71.1 (C-3), 68.0 (CH_{2All}), 67.8 (C-5), 38.0 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.2 (CH_{2Lev}), 18.0 (C-6), 6.8 (CH_{3TES}), 4.9 (CH_{2TES}). HRMS (ESI⁺): *m/z* 479.2754 (calcd for C₃₁H₄₄O₇SiNa [M+Na]⁺ *m/z* 479.2791)

2-O-Levulinoyl-4-O-(2-naphthylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranose (17). *Route 1:* [Ir] (358 mg, 0.636 mmol, 0.02 equiv.) was dissolved in anhyd. THF (30 mL) and the solution was degassed repeatedly then stirred under a satd. H₂ atmosphere for 30 min. The resulting yellow solution was degassed repeatedly before being poured into a solution of rhamnoside **S2** (17.7 g, 31.8 mmol) in anhyd. THF (317 mL). The mixture was stirred under an Ar atmosphere at rt for 2 h. A TLC control (Cyclohexane/EtOAc 7:3) showed complete conversion of the starting material (R_f = 0.53) into a less polar intermediate (R_f = 0.56). A solution of I₂ (16.1 g, 63.6 mmol, 2.0 equiv.) and NaHCO₃ (21.4 g, 254.4 mmol, 8.0 equiv.) in 1:5 H₂O/THF (389 mL) was then poured at 0 °C. After stirring for 2 h at this temperature, a TLC control (Cyclohexane/EtOAc 7:3) showed complete conversion of the intermediate (R_f = 0.56) into a

more polar product ($R_f = 0.26$). 10% aq. $\text{Na}_2\text{S}_2\text{O}_5$ (100 mL) was added. THF was evaporated under reduced pressure and DCM (300 mL) was added. The aq. layer was extracted with DCM (2 x 200 mL) and the combined organic phases were washed with satd aq. NaHCO_3 (100 mL) and brine (100 mL). The organic layer was dried over anhyd. Na_2SO_4 , filtered and concentrated under vacuum. The residue was purified by flash chromatography (cHex/EtOAc 8:2 to 6:4 + 1% Et_3N) to give hemiacetal **17** (14.6 g, 89%, α/β 4:1) as an off-white to yellow oil.

Route 2: To a solution of diol **16** (20.0 g, 58.07 mmol) in anhyd. MeCN (580 mL) stirred at rt were successively added dropwise TES-Cl (10.23 mL, 60.97 mmol, 1.05 equiv.) and DIPEA (15.17 mL, 87.11 mmol, 1.5 equiv.). After stirring the RM for 2 h at rt, a TLC control (Cyclohexane/EtOAc 8:2) showed complete consumption of the starting material (R_f 0.13) and the presence of a main product (R_f 0.70). MeOH (2.35 mL, 58.07 mmol, 1 equiv.) was added, the mixture was stirred for 15 min, and volatiles were evaporated under reduced pressure. Toluene was added and a precipitate appeared. The suspension was filtered and the filtrate was concentrated to dryness. The crude residue was dissolved in anhyd. DCM (194 mL) and stirred at rt. Then were successively added DCC (23.96 g, 116.14 mmol, 2.0 equiv.), DMAP (5.68 g, 46.46 mmol, 0.8 equiv.) and levulinic acid (17.84 mL, 174.21 mmol, 3.0 equiv.). After stirring for 15 h at rt, TLC (Cyclohexane/EtOAc 7:3) showed complete consumption of the starting material (R_f 0.60) and the presence of a more polar compound (R_f 0.53). The reaction mixture was concentrated under reduced pressure. The crude material was taken in EtOAc (200 mL) and the resulting suspension was filtered on a pad of Celite®. H_2O (150 mL) was added to the filtrate and the organic layer was washed successively with 10% aq. CuSO_4 (2 x 150 mL), H_2O (2 x 150 mL), satd aq. NaHCO_3 (150 mL) and brine (150 mL). The organic layer was dried over anhyd. Na_2SO_4 , filtered and concentrated under vacuum. [Ir] (982 mg, 1.16 mmol, 0.02 equiv.) was dissolved in anhyd. THF (50 mL) and the solution was degassed repeatedly then stirred under a satd. H_2 atmosphere for 1 h. The resulting yellow solution was degassed repeatedly before being poured into a solution of the crude residue in anhyd. THF (580 mL). The mixture was stirred under an Ar atmosphere at rt for 3 h. A TLC control (Cyclohexane/EtOAc 7:3) showed complete conversion of the starting material (R_f 0.53) into a less polar product (R_f 0.56). A mixture of I_2 (29.5 g, 116.1 mmol, 2 equiv.) and NaHCO_3 (39.0 g, 464.6 mmol, 8 equiv.) in 1:5 $\text{H}_2\text{O}/\text{THF}$ (711 mL) was added to the mixture stirred at 0 °C. After stirring for 2 h at this temperature, a TLC control (Cyclohexane/EtOAc 7:3) showed complete conversion of the intermediate into a more polar product (R_f 0.26). Satd aq. $\text{Na}_2\text{S}_2\text{O}_5$ (500 mL) was added. THF was evaporated under reduced pressure and DCM (500 mL) was added. The aq. layer was extracted with DCM (2 x 250 mL) and the combined organic phases were washed with satd aq. NaHCO_3 (500 mL) and brine (500 mL). The organic layer was dried over anhyd. Na_2SO_4 , filtered and concentrated under vacuum. The residue was purified by flash chromatography (Cyclohexane/EtOAc 8:2 to 6:4 + 1% Et_3N) to give hemiacetal **19** (24.0 g, 80%, α/β 4:1) as an

off-white to yellow oil. The later had ^1H NMR (CDCl_3) δ 7.89-7.78 (m, 4H, H_{ArNap}), 7.53-7.45 (m, 3H, H_{ArNap}), 5.30 (dd, 0.25H, $J_{2,1} = 1.1$ Hz, $J_{2,3} = 3.4$ Hz, H-2 β), 5.13-5.03 (m, 2.5H, H-1 α , H-2 α , $\text{CH}_{2\text{Nap}}$), 4.85-4.77 (d, 1.25H, H-1 β , $\text{CH}_{2\text{Nap}}$), 4.28 (dd, 0.75H, $J_{3,2} = 2.7$ Hz, $J_{3,4} = 9.0$ Hz, H-3 α), 4.00 (dq, 0.75H, H-5 α), 3.92 (dd, 0.25H, $J_{3,4} = 8.8$ Hz, H-3 β), 3.46 (m, 0.25H, H-5 β), 3.43 (t, 0.75H, $J_{4,5} = 9.3$ Hz, H-4 α), 3.37 (t, 0.25H, $J_{4,5} = 9.0$ Hz, H-4 β), 2.93-2.61 (m, 4H, $\text{CH}_{2\text{Lev}}$), 2.21 (s, 3H, $\text{CH}_{3\text{Lev}}$), 1.36 (d, 0.75H, $J_{6,5} = 6.2$ Hz, H-6 β), 1.30 (d, 2.25H, $J_{6,5} = 6.3$ Hz, H-6 α), 1.02-0.94 (m, 9H, $\text{CH}_{3\text{TES}}$), 0.68 (q, 6H, $\text{CH}_{2\text{TES}}$). ^{13}C NMR (CDCl_3) δ 206.2 (CO_{Lev}), 172.2 ($\text{CO}_{2\text{Lev}}$), 136.0 (C_{IVNap}), 133.3 (C_{IVNap}), 133.0 (C_{IVNap}), 128.0-125.8 (C_{ArNap}), 93.1 (C-1 β , $^1J_{\text{C,H}} = 158.5$ Hz), 92.2 (C-1 α , $^1J_{\text{C,H}} = 171.3$ Hz), 81.3 (C-4 α), 80.9 (C-4 β), 75.3 ($\text{CH}_{2\text{Nap}}$), 74.2 (C-2 β), 73.8 (C-2 α), 73.6 (C-3 β), 71.7 (C-5 β), 70.6 (C-3 α), 68.0 (C-5 α), 38.0 ($\text{CH}_{2\text{Lev}}$), 29.8 ($\text{CH}_{3\text{Lev}}$), 28.2 ($\text{CH}_{2\text{Lev}}$), 18.2 (C-6 β), 18.1 (C-6 α), 6.8 ($\text{CH}_{3\text{TES}}$), 4.9 ($\text{CH}_{2\text{TES}}$). HRMS (ESI $^+$): m/z 539.2441 (calcd for $\text{C}_{28}\text{H}_{40}\text{O}_7\text{SiNa}$ [$\text{M}+\text{Na}$] $^+$ m/z 539.2451).

2-O-Levulinoyl-4-O-(2-naphthylmethyl)-3-O-triethylsilyl- α/β -L-rhamnopyranosyl trichloroacetimidate (15). To a solution of hemiacetal **17** (24.0 g, 46.5 mmol) in anhyd. DCE (210 mL) stirred at rt under an Ar atmosphere were successively added CCl_3CN (13.97 mL, 139.3 mmol, 3.0 equiv.) and DBU (3.47 mL, 23.2 mmol, 0.5 equiv.). After stirring the RM for 3 h under Ar at rt, a TLC control (Cyclohexane/EtOAc 7:3) showed complete consumption of the starting material ($R_f = 0.26$) into a less polar product ($R_f = 0.53$). Volatiles were evaporated under reduced pressure. The residue was purified by flash chromatography (Cyclohexane/EtOAc 9:1 to 5:5 + 1% Et_3N) to give donor **15** (28.0 g, 91%, α/β 95:5) as a whitish crystalline solid. Trichloroacetimidate **15** had ^1H NMR (CDCl_3) δ 8.67 (s, 1H, NH), 7.89-7.79 (m, 4H, H_{ArNap}), 7.54-7.47 (m, 3H, H_{ArNap}), 6.15 (bd, 1H, $J_{1,2} = 2.0$ Hz, H-1), 5.26 (q, 1H, H-2), 5.10 (d, 1H, $\text{CH}_{2\text{Nap}}$), 4.82 (d, 1H, $\text{CH}_{2\text{Nap}}$), 4.32 (dd, 1H, $J_{3,2} = 3.4$ Hz, H-3), 3.95 (dq, 1H, H-5), 3.52 (t, 1H, $J_{4,3} = J_{4,5} = 9.3$ Hz, H-4), 2.84-2.74 (m, 4H, $\text{CH}_{2\text{Lev}}$), 2.23 (s, 3H, $\text{CH}_{3\text{Lev}}$), 1.32 (d, 3H, $J_{6,5} = 6.2$ Hz, H-6), 1.03-0.96 (m, 9H, $\text{CH}_{3\text{TES}}$), 0.69 (q, 6H, $\text{CH}_{2\text{TES}}$). ^{13}C (CDCl_3) δ 205.9 (CO_{Lev}), 171.3 ($\text{CO}_{2\text{Lev}}$), 160.1 (CO_{TCA}) 135.7 (C_{IVNap}), 133.3 (C_{IVNap}), 133.0 (C_{IVNap}), 128.1-125.9 (C_{ArNap}), 95.0 (C-1, $^1J_{\text{C,H}} = 179.7$ Hz), 80.5 (C-4), 75.7 ($\text{CH}_{2\text{Nap}}$), 71.9 (C-2), 71.0-70.9 (C-3 + C-5), 38.0 ($\text{CH}_{2\text{Lev}}$), 29.8 ($\text{CH}_{3\text{Lev}}$), 28.1 ($\text{CH}_{2\text{Lev}}$), 18.0 (C-6), 6.8 ($\text{CH}_{3\text{TES}}$), 4.8 ($\text{CH}_{2\text{TES}}$). HRMS (ESI $^+$): m/z 682.1537 (calcd for $\text{C}_{30}\text{H}_{40}\text{Cl}_3\text{NO}_7\text{SiNa}$ [$\text{M}+\text{Na}$] $^+$ m/z 682.1526).

Allyl 2-O-chloroacetyl-4-O-(2-naphthylmethyl)- α -L-rhamnopyranoside (14). To a solution of diol **16** (15.3 g, 44.5 mmol) in anhyd. MeCN (306 mL) stirred at rt was added 2-chloro-1,1,1-trimethoxyethane (18.9 mL, 133.4 mmol, 3.0 equiv.) and PTSA $\cdot\text{H}_2\text{O}$ (0.85 g, 4.45 mmol, 0.1 equiv.). After stirring for 1 h at rt, TLC (Cyclohexane/EtOAc 7:3) showed complete consumption of the starting material (R_f 0.13) into a mix of two less polar products ($R_f = 0.74, 0.79$). Then, 90% aq. TFA (214 mL) was added to the reaction medium at 0 $^\circ\text{C}$. After stirring for 30 min at rt, a TLC control (Cyclohexane/EtOAc 7:3) showed complete consumption of the intermediate

material into a more polar product (R_f 0.49). H_2O was added until the mixture became completely cloudy (100 mL). The product was extracted with EtOAc (2 x 200 mL). The organic phase was washed with satd aq. $NaHCO_3$ (2 x 200 mL) and brine (200 mL). The organic layer was dried over anhyd. Na_2SO_4 , filtered and concentrated under vacuum to obtain a 95:5 mixture of alcohol **14** and its regioisomer **14a**. Acceptor **14** had 1H NMR ($CDCl_3$) δ 7.92-7.81 (m, 4H, H_{ArNap}), 7.55-7.48 (m, 3H, H_{ArNap}), 5.97-5.85 (m, 1H, $CH=CH_2$), 5.33 (dq, 1H, $CH=CH_2$), 5.27-5.22 (m, 2H, H-2, $CH=CH_2$), 5.00 (d, 1H, CH_{2Nap}), 4.92 (d, 1H, CH_{2Nap}), 4.84 (bd, 1H, $J_{1,2} = 1.6$ Hz, H-1), 4.25-4.15 (m, 2H, H-3, CH_{2All}), 4.18 (d, 2H, CH_{2CA}), 4.01 (ddt_{app}, 1H, CH_{2All}), 3.85 (dq, 1H, H-5), 3.45 (t, 1H, $J_{4,3} = J_{4,5} = 9.4$ Hz, H-4), 2.21 (bs, 1H, OH), 1.41 (d, 3H, $J_{6,5} = 6.9$ Hz, H-6). ^{13}C NMR ($CDCl_3$) δ 167.1 (CO_{CA}), 135.6 (C_{IVNap}), 133.4 ($CH=CH_2$), 133.2 (C_{IVNap}), 133.1 (C_{IVNap}), 129.0-125.9 (C_{ArNap}), 117.7 ($CH=CH_2$), 96.2 (C-1, $^1J_{C,H} = 170.4$ Hz), 81.5 (C-4), 75.3 (CH_{2Nap}), 74.6 (C-2), 70.2 (C-3), 68.2 (CH_{2All}), 67.7 (C-5), 40.9 (CH_{2CA}), 18.0 (C-6). HRMS (ESI⁺): m/z 443.1284 (calcd for $C_{22}H_{25}ClO_6Na$ [$M+Na$]⁺ m/z 443.1232).

Allyl 2-O-levulinoyl-4-O-(2-naphthylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-chloroacetyl-4-O-(2-naphthylmethyl)- α -L-rhamnopyranoside (18). To a solution of crude acceptor **14** (44.5 mmol, 1.1 equiv.) in anhyd. Toluene (424 mL), stirred at rt under an Ar atmosphere, were successively added donor **15** (28.0 g, 42.4 mmol) and activated 4 Å MS (5.6 g). The suspension was stirred for 15 min at rt, then cooled to -60 °C. After 15 min of stirring at this temperature, TBSOTf (781 μ L, 3.4 mmol, 0.08 equiv.) was slowly added. After stirring for 1.5 h under Ar while the temperature wise rising from from -60 to -15 °C, a TLC control (Cyclohexane/EtOAc 7:3) showed complete consumption of donor **15** ($R_f = 0.60$) into a less polar product ($R_f = 0.72$). The mixture was cooled down to -60 °C and Et_3N (700 μ L, 5.0 mmol, 0.12 equiv.) was added for neutralization. The reaction mixture was stirred for 15 min at -60 °C and then filtered on a pad of Celite®. The filtrate was concentrated under reduced pressure. The residue was purified by flash chromatography (Toluene/EtOAc 95:5) to give disaccharide **18** (32.1 g, 82 %) as a white to off-white crystalline solid. The glycosylation product had 1H NMR ($CDCl_3$) δ 7.90-7.80 (m, 8H, H_{ArNap}), 7.53-7.45 (m, 6H, H_{ArNap}), 5.95-5.83 (m, 1H, $CH=CH_2$), 5.29 (dq, 1H, $CH=CH_2$), 5.24 (m, 1H, H-2_C), 5.22 (dq, 1H, $CH=CH_2$), 5.20 (m_{app}, 1H, H-2_B), 5.08 (bs, 1H, H-1_B), 5.05 (2d, 2H, CH_{2Nap}), 4.82 (bs, 1H, H-1_C), 4.80 (2d, 2H, CH_{2Nap}), 4.26 (dd, 1H, $J_{3,2} = 3.3$ Hz, $J_{3,4} = 9.5$ Hz, H-3_C), 4.19 (dd, 1H, $J_{3,2} = 3.5$ Hz, $J_{3,4} = 9.1$ Hz, H-3_B), 4.17 (bd, 2H, CH_{2CA}), 4.16 (dt, 1H, CH_{2All}), 3.99 (dt, 1H, CH_{2All}), 3.81 (dq, 1H, H-5_C), 3.77 (dq, 1H, H-5_B), 3.56 (t, 1H, $J_{4,5} = 9.5$ Hz, H-4_C), 3.40 (t, 1H, $J_{4,5} = 9.3$ Hz, H-4_B), 2.69-2.57 (m, 4H, CH_{2Lev}), 2.13 (s, 3H, CH_{3Lev}), 1.32 (d, 3H, $J_{6,5} = 6.2$ Hz, H-6_C), 1.27 (d, 3H, $J_{6,5} = 6.2$ Hz, H-6_B), 0.92 (t, 9H, CH_{3TES}), 0.62 (q, 6H, CH_{2TES}). ^{13}C NMR ($CDCl_3$) δ 206.0 (CO_{Lev}), 171.8 (CO_{2Lev}), 166.7 (CO_{CA}), 136.1 (C_{IVNap}), 135.4 (C_{IVNap}), 133.4 ($CH=CH_2$, C_{IVNap}), 133.3 (C_{IVNap}), 133.1 (C_{IVNap}), 132.9 (C_{IVNap}), 128.1-125.7 (C_{ArNap}), 117.7 ($CH=CH_2$), 99.6 (C-1_B, $^1J_{C,H} = 173.2$ Hz), 95.8 (C-1_C, $^1J_{C,H} = 172.0$ Hz), 80.5 (C-4_B), 80.3 (C-4_C), 76.7 (C-3_C), 75.5 (CH_{2Nap}), 75.1 (CH_{2Nap}), 74.2 (C-2_C), 73.5 (C-2_B), 70.9 (C-3_B), 68.9 (C-5_B), 68.3 (CH_{2All}), 67.9 (C-5_C), 40.8

(CH_{2CA}), 37.9 (CH_{2Lev}), 29.7 (CH_{3Lev}), 28.1 (CH_{2Lev}), 18.0 (C-6_C), 17.9 (C-6_B), 6.8 (CH_{3TES}), 4.7 (CH_{2TES}). HRMS (ESI⁺): *m/z* 941.3675 (calcd for C₅₀H₆₃ClO₁₂SiNa [M+Na]⁺ *m/z* 941.3787)

Allyl 4-O-(2-naphthylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-chloroacetyl-4-O-(2-naphthylmethyl)- α -L-rhamnopyranoside (S3). To a solution of disaccharide **18** (15.0 g, 16.3 mmol) in a mix of 3:2 Pyridine/AcOH (204 mL), stirred at rt, was added NH₂NH₂·H₂O (1.58 mL, 32.6 mmol, 2.0 equiv.). After stirring the RM for 1.5 h at rt, TLC (cHex/EtOAc 7:3) showed complete consumption of the starting material (R_f = 0.74) into a less polar product (R_f = 0.82). H₂O (50 mL) and EtOAc (200 mL) were added. The aq. layer was extracted with EtOAc (2 x 100 mL) and the combined organic phases were washed times with 10% aq. CuSO₄ (3 x 50 mL), H₂O (2 x 50 mL), satd aq. NaHCO₃ (50 mL) and brine (2 x 50 mL). The organic layer was dried over anhyd. Na₂SO₄, filtered and concentrated under vacuum. Disaccharide **21** had ¹H NMR (CDCl₃) δ 7.90-7.80 (m, 8H, H_{ArNap}), 7.53-7.45 (m, 6H, H_{ArNap}), 5.95-5.83 (m, 1H, CH=CH₂), 5.31 (m, 2H, CH=CH₂, H-2_C), 5.23 (dq, 1H, CH=CH₂), 5.19 (d, 1H, H-1_B), 5.01 (2d, 2H, CH_{2Nap}), 4.82 (m, 3H, CH_{2Nap}, H-1_C), 4.27 (dd, 1H, *J*_{3,2} = 3.3 Hz, *J*_{3,4} = 9.5 Hz, H-3_C), 4.19 (m, 3H, CH_{2All}, CH_{2CA}), 4.07 (dd, 1H, *J*_{3,2} = 3.3 Hz, *J*_{3,4} = 9.5 Hz, H-3_B), 4.02 (dt, 1H, CH_{2All}), 3.86 (m, 2H, H-2_B, H-5_C), 3.78 (dq, 1H, *J*_{5,4} = 9.5 Hz, H-5_B), 3.57 (t, 1H, *J*_{3,4} = *J*_{4,5} = 9.5 Hz, H-4_C), 3.43 (t, 1H, *J*_{3,4} = *J*_{4,5} = 9.2 Hz, H-4_B), 2.56 (m, 1H, OH-2_B), 1.35 (d, 3H, *J*_{6,5} = 6.2 Hz, H-6_C), 1.28 (d, 3H, *J*_{6,5} = 6.3 Hz, H-6_B), 0.95 (t, 9H, *J* = 7.9 Hz, CH_{3TES}), 0.63 (q, 6H, CH_{2TES}). ¹³C NMR (CDCl₃) δ 166.7 (CO_{CA}), 136.0 (2C, C_{IVNap}), 135.3 (2C, C_{IVNap}), 133.3 (CH=CH₂), 133.1-128.2 (4C, C_{ArNap}), 127.9-125.7 (14C, C_{ArNap}), 117.7 (CH=CH₂), 101.4 (C-1_B, ¹*J*_{C,H} = 172.9 Hz), 95.9 (C-1_C, ¹*J*_{C,H} = 171.4 Hz), 80.8 (C-4_B), 80.3 (C-4_C), 77.5 (C-3_C), 75.5 (CH_{2Nap}), 75.2 (CH_{2Nap}), 74.4 (C-2_C), 73.0 (C-3_B), 72.5 (C-2_B), 68.4 (C-5_B), 68.3 (CH_{2All}), 67.9 (C-5_C), 40.8 (CH_{2CA}), 18.0 (C-6_C), 17.9 (C-6_B), 6.7 (CH_{3TES}), 5.0 (CH_{2TES}). HRMS (ESI⁺): *m/z* 838.3753 (calcd for C₄₅H₅₇ClO₁₀SiNa [M+Na]⁺ *m/z* 838.3743).

Allyl 2-O-levulinoyl-4-O-(2-naphthylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 2)-4-O-(2-naphthylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-chloroacetyl-4-O-(2-naphthylmethyl)- α -L-rhamnopyranoside (19). To a solution of disaccharide **18** (15.0 g, 16.3 mmol) in 3:2 Pyridine/AcOH (204 mL) stirred at rt was added NH₂NH₂·H₂O (1.58 mL, 32.6 mmol, 2.0 equiv.). After stirring for 1.5 h at rt, TLC (Cyclohexane/EtOAc 7:3) showed complete consumption of the starting material (R_f = 0.74) into a less polar product (R_f = 0.82). H₂O (50 mL) and EtOAc (200 mL) were added. The aq. layer was extracted with EtOAc (2 x 100 mL) and the combined organic phases were washed times with 10% aq. CuSO₄ (3 x 50 mL), H₂O (2 x 50 mL), satd aq. NaHCO₃ (50 mL) and brine (2 x 50 mL). The organic layer was dried over anhyd. Na₂SO₄, filtered and concentrated under vacuum. The crude residue was dissolved in anhyd. Tol (326 mL), stirred at rt under an Ar atmosphere, and donor **15** (13.7 g, 21.2 mmol, 1.3 equiv.) and activated 4 Å MS (3.4 g) were successively added. The suspension was stirred for 15 min at rt, then cooled to -40 °C. After

15 min of stirring at this temperature, TBSOTf (215 μ L, 0.815 mmol, 0.05 equiv.) was slowly added. After stirring for 2 h under Ar from -40 to -10 $^{\circ}$ C, a TLC control (Cyclohexane/EtOAc 7:3) showed complete consumption of the disaccharide **18** (R_f = 0.54) into a less polar product (R_f = 0.75). The mixture was cooled down to -60 $^{\circ}$ C and Et₃N (204 μ L, 1.47 mmol, 0.09 equiv.) was then added for neutralization. The reaction mixture was stirred for 15 minutes at -60 $^{\circ}$ C and then filtered on a pad of Celite®. The filtrate was concentrated under reduced pressure. The residue was purified by flash chromatography (Cyclohexane/EtOAc 9:1 to 8:2) to give trisaccharide **19** (17.9 g, 83 % over two steps) as a colorless oil. Trisaccharide **19** had ¹H NMR (CDCl₃) δ 7.90-7.78 (m, 12H, H_{ArNap}), 7.53-7.44 (m, 9H, H_{ArNap}), 5.96-5.84 (m, 1H, CH=CH₂), 5.30 (dq, 1H, CH=CH₂), 5.28 (dd, 1H, H-2_A), 5.25 (dd, 1H, H-2_C), 5.21 (dq, 1H, CH=CH₂), 5.10 (bs, 1H, H-1_B), 5.06 (3d, 3H, CH_{2Nap}), 5.04 (bs, 1H, H-1_A), 4.82 (bd, 1H, H-1_C), 4.80-4.73 (3d, 3H, CH_{2Nap}), 4.28 (dd, 1H, H-3_A), 4.23 (dd, 1H, H-3_C), 4.20-4.13 (m, 2H, H-3_B, CH_{2All}), 4.16 (bs, 2H, CH_{2CA}), 3.99 (dt, 1H, CH_{2All}), 3.94 (t, 1H, H-2_B), 3.85 (dq, 1H, H-5_A), 3.81 (dq, 1H, H-5_C), 3.68 (dq, 1H, H-5_B), 3.54 (t, 1H, H-4_C), 3.45 (t, 1H, H-4_B), 3.40 (t, 1H, H-4_A), 2.80-2.70 (m, 4H, CH_{2Lev}), 2.21 (s, 3H, CH_{3Lev}), 1.26 (d, 3H, H-6_C), 1.24 (d, 3H, H-6_B), 1.15 (d, 3H, H-6_A), 0.99 (t, 9H, CH_{3TES}), 0.93 (t, 9H, CH_{3TES}), 0.69 (q, 6H, CH_{2TES}), 0.62 (q, 6H, CH_{2TES}). ¹³C NMR (CDCl₃) δ 206.0 (CO_{Lev}), 171.6 (CO_{2Lev}), 166.7 (CO_{CA}), 136.3 (C_{IVNap}), 136.1 (C_{IVNap}), 135.3 (C_{IVNap}), 133.4 (CH=CH₂), 133.4 (C_{IVNap}), 133.3 (C_{IVNap}), 133.3 (C_{IVNap}), 133.1 (C_{IVNap}), 132.9 (C_{IVNap}), 132.9 (C_{IVNap}), 128.3-125.7 (C_{ArNap}), 117.7 (CH=CH₂), 101.5 (C-1_B, ¹J_{C,H} = 171.6 Hz), 99.1 (C-1_A, ¹J_{C,H} = 173.0 Hz), 95.7 (C-1_C, ¹J_{C,H} = 171.9 Hz), 81.2 (C-4_A, C-4_B), 79.7 (C-4_C), 78.7 (C-2_B), 78.1 (C-3_C), 75.5 (CH_{2Nap}), 75.3 (CH_{2Nap}), 75.2 (CH_{2Nap}), 74.5 (C-2_C), 73.3 (C-2_A), 72.7 (C-3_B), 71.1 (C-3_A), 69.2 (C-5_B), 68.6 (C-5_A), 68.3 (CH_{2All}), 67.8 (C-5_C), 40.7 (CH_{2CA}), 38.1 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.2 (CH_{2Lev}), 18.0 (C-6_A, C-6_B, C-6_C), 17.9 (C-6_B), 6.9 (CH_{3TES}), 6.8 (CH_{3TES}), 5.0 (CH_{2TES}), 4.9 (CH_{2TES}). HRMS (ESI⁺): m/z 1336.6191 (calcd for C₇₃H₉₅ClO₁₆Si₂NH₄ [M+NH₄]⁺ m/z 1336.6224)

2-O-Levulinoyl-4-O-(2-naphtylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 2)4-O-(2-naphtylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-chloroacetyl-4-O-(2-naphtylmethyl)- α/β -L-rhamnopyranose (20). [Ir] (103 mg, 0.12 mmol, 0.02 equiv.) was dissolved in anhyd. THF (10 mL) and the solution was degassed repeatedly then stirred under a satd. H₂ atmosphere for 30 min. The resulting yellow solution was degassed repeatedly before being poured into a solution of trisaccharide **19** in anhyd. THF (67 mL). The mixture was stirred under an Ar atmosphere at rt for 4 h. A TLC control (Cyclohexane/EtOAc 7:3) showed complete conversion of the starting material (R_f = 0.62) into a less polar product (R_f = 0.64). A solution of NIS (1.64 g, 7.27 mmol, 1.2 equiv.) in 1:5 H₂O/THF (34 mL) and then additional H₂O (84 mL) were then poured into the mixture. After stirring for 4 h at rt, a TLC control (Cyclohexane/EtOAc 7:3) showed complete conversion of the intermediate into a more polar product (R_f = 0.35). Satd aq. Na₂S₂O₅ (200 mL) and then EtOAc (500 mL) were added. The aq. layer was extracted with EtOAc (2 x 250 mL) and the combined organic phases were

washed with satd aq. NaHCO₃ (200 mL) and brine (200 mL). The organic layer was dried over anhyd. Na₂SO₄, filtered and concentrated under vacuum. The residue was purified by flash chromatography (cHex/EtOAc 8:2 to 6:4) to give hemiacetal **20** (6.2 g, 80%, α/β 8:2) as a yellow oil. The α anomer had ¹H NMR (CDCl₃) δ 7.90-7.77 (m, 12H, H_{ArNap}), 7.52-7.43 (m, 9H, H_{ArNap}), 5.27 (bs, 2H, H-2_A, H-2_C), 5.20 (bs, 1H, H-1_A), 5.12 (bs, 1H, H-1_B), 5.05 (d, 3H, CH_{2Nap}), 5.04 (bs, 1H, H-1_C), 4.80-4.73 (2d, 3H, CH_{2Nap}), 4.28 (dd, 2H, H-3_A, H-3_C), 4.17 (m, 1H, H-3_B), 4.15 (bs, 2H, CH_{2CA}), 4.01 (dq, 1H, H-5_C), 3.93 (dt, 1H, H-2_B), 3.85 (dq, 1H, H-5_A), 3.69 (dq, 1H, H-5_B), 3.54 (t, 1H, H-4_C), 3.45 (t, 1H, H-4_B), 3.40 (t, 1H, H-4_A), 2.80-2.70 (m, 4H, CH_{2Lev}), 2.21 (s, 3H, CH_{3Lev}), 1.24 (d, 3H, H-6_B), 1.23 (d, 3H, H-6_C), 1.15 (d, 3H, H-6_A), 0.99 (t, 9H, CH_{3TES}), 0.93 (t, 9H, CH_{3TES}), 0.69 (q, 6H, CH_{2TES}), 0.62 (q, 6H, CH_{2TES}). ¹³C NMR (CDCl₃) δ 206.0 (CO_{Lev}), 171.6 (CO_{2Lev}), 166.7 (CO_{CA}), 136.3 (C_{IVNap}), 136.1 (C_{IVNap}), 135.3 (C_{IVNap}), 133.4 (C_{IVNap}), 133.3 (C_{IVNap}), 133.3 (C_{IVNap}), 133.1 (C_{IVNap}), 132.9 (C_{IVNap}), 132.9 (C_{IVNap}), 128.3-125.7 (C_{ArNap}), 101.4 (C-1_B, ¹J_{C,H} = 171.6 Hz), 99.1 (C-1_C, ¹J_{C,H} = 172.5 Hz), 91.4 (C-1_A, ¹J_{C,H} = 172.8 Hz), 81.2 (C-4_A, C-4_C), 79.6 (C-4_B), 78.8 (C-2_B), 77.2 (C-3_C), 75.4 (CH_{2Nap}), 75.3 (CH_{2Nap}), 75.2 (CH_{2Nap}), 74.7 (C-2_C), 73.3 (C-2_A), 72.7 (C-3_B), 71.1 (C-3_A), 69.3 (C-5_B), 68.7 (C-5_A), 67.9 (CH_{2All}), 67.8 (C-5_C), 40.7 (CH_{2CA}), 38.1 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.2 (CH_{2Lev}), 18.0 (C-6_A, C-6_B, C-6_C), 17.9 (C-6_B), 6.9 (CH_{3TES}), 6.8 (CH_{3TES}), 5.0 (CH_{2TES}), 4.9 (CH_{2TES}). HRMS (ESI⁺): *m/z* 1296.5878 (calcd for C₇₀H₉₁ClO₁₆Si₂NH₄ [M+NH₄]⁺ *m/z* 1296.5891).

2-O-Levulinoyl-4-O-(2-naphtylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 2)4-O-(2-naphtylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-chloroacetyl-4-O-(2-naphtylmethyl)- α/β -L-rhamnopyranosyl *N*-(phenyl)trifluoroacetimidate (21**).** To a solution of hemiacetal **20** (5.74 g, 4.48 mmol) in anhyd. acetone (90 mL), stirred at rt, were successively added PTFA-Cl (1.07 mL, 6.73 mmol, 1.5 equiv.) and K₂CO₃ (1.24 g, 8.97 mmol, 2.0 equiv.). After stirring for 3 h at rt, TLC (Cyclohexane/EtOAc 7:3) showed complete consumption of hemiacetal **20** (R_f = 0.35) into a less polar product (R_f = 0.67). The mixture was filtered on a pad of Celite® and the filtrate was concentrated under reduced pressure. The residue was purified by flash chromatography (Toluene/EtOAc 8:2 to 5:5) to give donor **21** (6.0 g, 91%, α/β 8:2) as a colorless oil. Imidate **21** had ¹H NMR (CDCl₃) δ 7.91-7.78 (m, 12H, H_{ArNap}), 7.53-7.42 (m, 9H, H_{ArNap}), 7.30 (t, 2H, H_{ArPh}), 7.12 (t, 1H, H_{ArPh}), 6.90-6.82 (m, 2H, H_{ArPh}), 6.18 (bs, 1H, H-1_C), 5.42 (bs, 1H, H-2_C), 5.29 (dt, 1H, H-2_A), 5.17 (bs, 1H, H-1_B), 5.12 (bs, 1H, H-1_A), 5.11-5.02 (m, 3H, CH_{2Nap}), 4.82-4.73 (m, 3H, CH_{2Nap}), 4.29 (dd, 1H, H-3_A), 4.24 (dd, 1H, H-3_C), 4.16 (bs, 2H, CH_{2CA}), 4.13 (dd, 1H, H-3_B), 3.95 (bs, 1H, H-2_B), 3.91 (dq, 1H, H-5_C), 3.84 (dq, 1H, H-5_A), 3.71 (dq, 1H, H-5_B), 3.64 (t, 1H, H-4_C), 3.48 (t, 1H, H-4_B), 3.42 (t, 1H, H-4_A), 2.83-2.71 (m, 4H, CH_{2Lev}), 2.22 (s, 3H, CH_{3Lev}), 1.30 (d, 3H, H-6_C), 1.26 (d, 3H, H-6_B), 1.16 (d, 3H, H-6_A), 1.05-0.91 (m, 18H, CH_{3TES}), 0.75-0.60 (m, 12H, CH_{2TES}). ¹³C NMR (CDCl₃) δ 206.3 (CO_{Lev}), 171.6 (CO_{2Lev}), 166.4 (CO_{CA}), 136.3 (C_{IVNap}), 136.1 (C_{IVNap}), 135.3 (C_{IVNap}), 133.4 (C_{IVNap}), 133.3 (C_{IVNap}), 133.3 (C_{IVNap}), 133.1 (C_{IVNap}), 132.9 (C_{IVNap}), 132.9 (C_{IVNap}), 128.3-125.7

(C_{ArNap}), 101.4 (C-1_B, ¹J_{C,H} = 171.6 Hz), 99.1 (C-1_C, ¹J_{C,H} = 172.5 Hz), 91.4 (C-1_A, ¹J_{C,H} = 172.8 Hz), 81.2 (C-4_A, C-4_C), 79.6 (C-4_B), 78.8 (C-2_B), 77.2 (C-3_C), 75.4 (CH_{2Nap}), 75.3 (CH_{2Nap}), 75.2 (CH_{2Nap}), 74.7 (C-2_C), 73.3 (C-2_A), 72.7 (C-3_B), 71.1 (C-3_A), 69.3 (C-5_B), 68.7 (C-5_A), 67.9 (CH_{2All}), 67.8 (C-5_C), 40.7 (CH_{2CA}), 38.1 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.2 (CH_{2Lev}), 18.0 (C-6_A, C-6_B, C-6_C), 17.9 (C-6_B), 6.9 (CH_{3TES}), 6.8 (CH_{3TES}), 5.0 (CH_{2TES}), 4.9 (CH_{2TES}). HRMS (ESI⁺): *m/z* 1467.6174 (calcd for C₇₈H₉₅ClF₃NO₁₆Si₂NH₄ [M+NH₄]⁺ *m/z* 1467.6160).

Allyl 2-O-levulinoyl-4-O-(2-naphtylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 2)-4-O-(2-naphtylmethyl)-3-O-triethylsilyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-chloroacetyl-4-O-(2-naphtylmethyl)- α -L-rhamnopyranosyl-(1 \rightarrow 3)-4,6-O-benzylidene-2-deoxy-2-

trichloroacetamido- β -D-glucopyranoside (13). To a solution of donor **21** (5.7 g, 3.93 mmol) in anhyd. 3:1 Toluene/DCM (223 mL), stirred at rt under an Ar atmosphere, were successively added acceptor **7** (3.56 g, 7.86 mmol, 2.0 equiv.) and activated 4 Å MS (11 g). The suspension was stirred for 15 min at rt, then cooled to -20 °C. After 15 min of stirring at this temperature, TBSOTf (0.135 mL, 0.590 mmol, 0.15 equiv.) was slowly added. After stirring for 1 h under Ar from -20 to 0 °C, TLC (Cyclohexane/EtOAc 7:3) showed complete consumption of donor **21** (R_f = 0.67) into a more polar product (R_f = 0.48). The mixture was cooled down to -40 °C and Et₃N (103 μ L, 0.74 mmol, 0.19 equiv.) was added for neutralization. The suspension was stirred for 15 minutes at -40 °C and then filtered on a pad of Celite®. The filtrate was concentrated under reduced pressure. The residue was purified by flash chromatography (Toluene/EtOAc 95:5 to 5:5) to give tetrasaccharide **13** (5.62 g, 84 %) as a colorless to yellow oil. The fully protected **13** had ¹H NMR (CDCl₃) δ 7.90-7.78 (m, 12H, H_{ArNap}), 7.70 (s, 1H, H_{ArBzl}), 7.54-7.34 (m, 11H, 9H_{ArNap}, 2H_{ArBzl}), 7.22-7.16 (m, 2H, H_{ArBzl}), 7.04 (d, 1H, NH), 5.93-5.81 (dq, 1H, CH=CH₂), 5.31 (d, 1H, CH=CH₂), 5.27 (bs, 2H, H-2_A, H-2_C), 5.22 (d, 1H, CH=CH₂), 5.20 (bs, 1H, H-1_A), 5.12 (bs, 1H, H-1_B), 5.05 (d, 3H, CH_{2Nap}), 5.04 (bs, 1H, H-1_C), 4.80-4.73 (2d, 3H, CH_{2Nap}), 4.28 (dd, 2H, H-3_A, H-3_C), 4.17 (m, 1H, H-3_B), 4.15 (bs, 2H, CH_{2CA}), 4.01 (dq, 1H, H-5_C), 3.93 (dt, 1H, H-2_B), 3.85 (dq, 1H, H-5_A), 3.69 (dq, 1H, H-5_B), 3.54 (t, 1H, H-4_C), 3.45 (t, 1H, H-4_B), 3.40 (t, 1H, H-4_A), 2.80-2.70 (m, 4H, CH_{2Lev}), 2.21 (s, 3H, CH_{3Lev}), 1.24 (d, 3H, H-6_B), 1.23 (d, 3H, H-6_C), 1.15 (d, 3H, H-6_A), 0.99 (t, 9H, CH_{3TES}), 0.93 (t, 9H, CH_{3TES}), 0.69 (q, 6H, CH_{2TES}), 0.62 (q, 6H, CH_{2TES}). ¹³C NMR (CDCl₃) δ 206.0 (CO_{Lev}), 171.6 (CO_{2Lev}), 166.7 (CO_{CA}), 136.3 (C_{IVNap}), 136.1 (C_{IVNap}), 135.3 (C_{IVNap}), 133.4 (C_{IVNap}), 133.3 (C_{IVNap}), 133.3 (C_{IVNap}), 133.1 (C_{IVNap}), 132.9 (C_{IVNap}), 132.9 (C_{IVNap}), 128.3-125.7 (C_{ArNap}), 101.4 (C-1_B, ¹J_{C,H} = 171.6 Hz), 99.1 (C-1_C, ¹J_{C,H} = 172.5 Hz), 91.4 (C-1_A, ¹J_{C,H} = 172.8 Hz), 81.2 (C-4_A, C-4_C), 79.6 (C-4_B), 78.8 (C-2_B), 77.2 (C-3_C), 75.4 (CH_{2Nap}), 75.3 (CH_{2Nap}), 75.2 (CH_{2Nap}), 74.7 (C-2_C), 73.3 (C-2_A), 72.7 (C-3_B), 71.1 (C-3_A), 69.3 (C-5_B), 68.7 (C-5_A), 67.9 (CH_{2All}), 67.8 (C-5_C), 40.7 (CH_{2CA}), 38.1 (CH_{2Lev}), 29.8 (CH_{3Lev}), 28.2 (CH_{2Lev}), 18.0 (C-6_A, C-6_B, C-6_C), 17.9 (C-6_B), 6.9 (CH_{3TES}), 6.8 (CH_{3TES}), 5.0 (CH_{2TES}), 4.9 (CH_{2TES}). HRMS (ESI⁺): *m/z* 1729.6128 (calcd for C₈₉H₁₀₉Cl₄NO₂₁Si₂NH₄ [M+NH₄]⁺ *m/z* 1729.6145).

Allyl α -L-rhamnopyranosyl-(1 \rightarrow 2)- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-chloroacetyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-deoxy-2-trichloroacetamido- β -D-glucopyranoside (12**).** To a solution of tetrasaccharide **13** (3.02 g, 1.76 mmol) in 3:2 Pyridine/AcOH (35 mL) stirred at rt, was added $\text{NH}_2\text{NH}_2\cdot\text{H}_2\text{O}$ (0.34 mL, 7.05 mmol, 4.0 equiv.). After stirring for 1 h at rt, a TLC control (Cyclohexane/EtOAc 8:2) showed complete consumption of the starting material ($R_f = 0.22$) into a less polar product ($R_f = 0.46$). Satd aq. NaHCO_3 (200 mL) and DCM (250 mL) were added. The aq. layer was extracted with DCM (150 mL) and the combined organic phases were washed with brine (150 mL). The organic layer was dried over anhyd. Na_2SO_4 , filtered, co-evaporated five times with Toluene and concentrated under vacuum. The residue was then diluted in Toluene (21 mL) and the mixture was cooled down to 0 °C. TFA (189 mL) was then slowly added and the RM was stirred for 20 h at rt under an Ar atmosphere. A TLC control (DCM/MeOH 8:2) showed complete consumption of the starting material ($R_f = 0.9$) into a more polar product ($R_f = 0.15$). Solvents were evaporated under reduced pressure and co-evaporations with Toluene (3 x 100 mL) then MeOH (3 x 100 mL) were performed. The residue was purified by semi-preparative RP-HPLC (215 nm, 20.5% isocratic gradient of MeCN/Water over 20 min) to give, after freeze-drying, the lightly protected **ABC'D'** **12** (559 mg, 36%) as a white powder. The lightly protected tetrasaccharide **12** had ^1H NMR (800 MHz, D_2O), δ 5.84 (m, 1H, $\text{CH}=\text{CH}_{2\text{All}}$), 5.25 (m, 1H, $J_{\text{trans}} = 17.3$ Hz, $\text{CH}=\text{CH}_{2\text{All}}$), 5.19 (m, 1H, $J_{\text{cis}} = 10.5$ Hz, $\text{CH}=\text{CH}_{2\text{All}}$), 5.13 (dd, 1H, $J_{2,3} = 3.0$ Hz, $J_{2,1} = 1.9$ Hz, H-2C), 5.11 (d, 1H, $J_{1,2} = 1.3$ Hz, H-1B), 4.86 (d, 1H, $J_{1,2} = 1.6$ Hz, H-1C), 4.86 (d, 1H, $J_{1,2} = 1.5$ Hz, H-1A), 4.66 (d, 1H, $J_{1,2} = 8.0$ Hz, H-1D), 4.28 (m, 1H, $\text{CH}_2\text{-CH}=\text{All}$), 4.27 (d, 1H, $J_{\text{gem}} = 15.4$ Hz, H_{CA}), 4.23 (d, 1H, $J_{\text{gem}} = 15.4$ Hz, H_{CA}), 4.12 (m, 1H, $\text{CH}_2\text{-CH}=\text{All}$), 4.05 (m, 1H, H-5C), 3.99 (dd, 1H, $J_{2,3} = 3.3$ Hz, $J_{2,1} = 1.7$ Hz, H-2A), 3.93 (dd, 1H, $J_{3,4} = 9.7$ Hz, $J_{3,2} = 3.1$ Hz, H-3C), 3.89 (dd, 1H, $J_{2,3} = 4.7$ Hz, $J_{2,1} = 1.7$ Hz, H-2B), 3.88 (d, 1H, $J_{6,5} = 2.1$ Hz, H-6bD), 3.86 (m, 1H, H-2D), 3.72 (m, 2H, H-3D, H-3A), 3.71 (d, 1H, $J_{6,5} = 3.3$ Hz, H-6aD), 3.69 (m, 1H, H-5B), 3.63 (dd, 1H, $J_{3,4} = 9.9$ Hz, $J_{3,2} = 3.5$ Hz, H-3B), 3.62 (m, 1H, H-5A), 3.54 (dd, 1H, $J_{4,5} = 9.9$ Hz, $J_{4,3} = 9.9$ Hz, H-4D), 3.52 (dd, 1H, $J_{4,5} = 9.8$ Hz, $J_{4,3} = 9.8$ Hz, H-4C), 3.42 (m, 1H, H-5D), 3.38 (dd, 1H, $J_{4,5} = 9.8$ Hz, $J_{4,3} = 9.8$ Hz, H-4B), 3.37 (dd, 1H, $J_{4,5} = 9.8$ Hz, $J_{4,3} = 9.8$ Hz, H-4A), 1.25 (d, 3H, $J_{6,5} = 6.2$ Hz, H-6B), 1.20 (d, 3H, $J_{6,5} = 6.4$ Hz, H-6A), 1.19 (d, 3H, $J_{6,5} = 6.4$ Hz, H-6C). ^{13}C NMR (800 MHz, D_2O), δ 168.6 (CO_{CA}), 164.7 ($\text{CO}_{\text{NHC(O)CCl}_3}$), 133.0 ($\text{CH}=\text{CH}_{2\text{All}}$), 118.7 ($\text{CH}=\text{CH}_{2\text{All}}$), 102.3 (C-1A), 100.5 (C-1B), 99.1 (C-1D), 98.5 (C-1C), 91.6 ($\text{C}_{\text{IVCCl}_3}$), 81.7 (C-3D), 78.0 (C-2B), 76.1 (C-5D), 74.5 (C-3C), 73.4 (C-2C), 72.2 (C-4C), 72.0 (2C, C-4B, C-4A), 70.8 ($\text{CH}_2\text{-CH}=\text{All}$), 70.0 (2C, C-3A, C-2A), 69.9 (C-3B), 69.5 (C-5B), 69.2 (C-5C), 69.1 (C-5A), 68.5 (C-4D), 60.7 (C-6D), 57.1 (C-2D), 40.7 ($\text{CH}_{2\text{CA}}$), 16.7 (C-6B), 16.6 (C-6A), 16.2 (C-6C). HRMS (ESI⁺): m/z 895.1860 (calc. for $\text{C}_{31}\text{H}_{47}\text{Cl}_4\text{NO}_{19}\text{NH}_4$ [$\text{M}+\text{NH}_4$]⁺: 895.1840). RP-HPLC t_R (210 nm): 18.04 min, t_R (ELSD): 18.08 min.

Allyl α -L-rhamnopyranosyl-(1 \rightarrow 2)- α -L-rhamnopyranosyl-(1 \rightarrow 3)-[α -D-glucopyranosyl-(1 \rightarrow 3)]-2-O-chloroacetyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-deoxy-2-trichloroacetamido- β -D-glucopyranoside (P1**).** A 1 mL scale reaction was performed using 1 $\text{U}\cdot\text{mL}^{-1}$ of purified

BRS-B Δ 1 in presence of acceptor **12** (44 mg, 50 μ mol, 50 mM) and sucrose (343 mg, 20 equiv., 1 M), in sodium acetate buffer pH 5.1 for 16 h. Product **P1** was purified by automated fractionation on an Agilent 1260 Infinity HPLC using a C18.RP Fusion column by means of a 20 min run eluting at a flow of 1 mL.min⁻¹ with H₂O/CH₃CN 70:30 for 4 min and a linear gradient of H₂O/CH₃CN 70:30 to 40:60 from 4 min to 11 min, followed by washing and re-equilibration steps. Fractions containing the products as detected by UV were collected, and re-analyzed by HPLC, using the same analytical conditions. Fractions containing a single peak product were pooled, concentrated to dryness using a SpeedVac to give pentasaccharide **P1** (1 mg, 2.3%) and analyzed (Figure S2).

Methods for enzymology

Production of the native GH70 branching sucrases and purification of BRS-B Δ 1

The six branching sucrases were recombinantly expressed in *E. coli* as previously reported (Vuillemin et al. 2014, 2016). Briefly, *E. coli* BL21 StarTM DE3 and *E. coli* BL21 AI competent cells (Life Technologies) were transformed by pET53-55 plasmids containing the genes *brsB* Δ 1, *brsC*, *brsD*, *brsE*, Δ N₁₂₃-GBD-CD2; and pBAD49 plasmid containing the gene *brsA*, respectively. Twenty milliliters of LB-Miller medium, supplemented with ampicillin (100 μ g.mL⁻¹), were inoculated with 100 μ L of transformation mix and incubated overnight at 37 °C under agitation (200 rpm).

Enzyme productions were performed in Erlenmeyer flasks with modified ZYM-5052 medium (Studier 2005) that contains i) 0% lactose, 0% glucose, 0.5% glycerol and 0.01% L-arabinose for BRS-A production, ii) 0.1% lactose, 0% glucose and 1% glycerol for BRS-B Δ 1, BRS-C, BRS-D, BRS-E production or iii) 0.75% lactose, 0.05% glucose and 1.5% glycerol for Δ N₁₂₃-GBD-CD2 production (Vuillemin et al. 2014). All culture media were supplemented with ampicillin (100 μ g mL⁻¹) and inoculated with the corresponding starter culture at an OD_{600nm} of 0.05. Cultures were incubated at 21 °C for 26 hours (BRS-A, BRS-B Δ 1, BRS-C, BRS-D, BRS-E) or 23 °C for 32 hours (Δ N₁₂₃-GBD-CD2) under agitation (150 rpm). Cells were then harvested by centrifugation, dispersed in 50 mM sodium acetate buffer (pH 5.75) at a final OD_{600nm} of 80 for BRS-A, BRS-B Δ 1, BRS-C, BRS-D, BRS-E, and an OD_{600nm} of 30 for Δ N₁₂₃-GBD-CD2. Cells were disrupted by sonication and recombinant enzymes were recovered in the soluble fraction after a cycle of centrifugation (15,500 g, 30 min, 8 °C) of the crude cell extract.

The BRS-B Δ 1 enzyme was purified by affinity chromatography onto nickel resins (Brison et al. 2012). Briefly, cells were centrifuged and re-suspended in binding buffer (20 mM phosphate sodium buffer, pH 7.4, 500 mM NaCl, 20 mM imidazole) at a final OD_{600nm} of 200. After disruption by sonication, centrifugation (18,000 g, 30 min, 4 °C) and filtration through a 0.22

μ m Ministart® syringe filter (Sartorius), lysates were applied onto a 1 mL HisTrap FF® column (GE Healthcare Life Sciences) that had been equilibrated with the binding buffer, using an AKTAXpress system (GE Healthcare Life Sciences) kept at 10 °C. Proteins were eluted by a 400 mM imidazole step and immediately stored in a 5 mL loop for online automated re-injection into a Superose12 resin (GE Healthcare Life Sciences). Elution of this gel filtration step was performed during 3 h 30 with 50 mM sodium acetate buffer at pH 5.75 supplemented with 100 mM NaCl at a flow of 1 mL.min⁻¹. Protein fractions were automatically collected and purity was controlled by sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) using Mini-PROTEAN® TGX Stain-Free™ precast gels (BioRad).

Measurement of enzyme activity on sucrose (DNS assay)

One unit of branching sucrose is defined as the amount of enzyme which catalyzes the production of one micromole of fructose per min from 292 mM sucrose, at 30 °C, in 50 mM sodium acetate buffer at pH 5.1 or pH 5.75. The enzyme activities were determined by measuring the amount of reducing sugars using the dinitrosalicylic acid (DNS) method (Miller 1959), corresponding to the initial velocity of released fructose.

Enzyme screening for tetrasaccharide **12** glucosylation

After optimization of the reaction conditions (see Complementary Results of this chapter) transglucosylation assays were performed at 30 °C in 50 mM sodium acetate buffer containing 0.34 mM CaCl₂, pH 5.1, supplemented with 0.4 U.mL⁻¹ of enzyme, 1 M sucrose, and 50 mM tetrasaccharide **12** corresponding to an acceptor:sucrose molar ratio of 1:20. Reactions were incubated in glass tubes for 16 h.

Analytical methods for the detection of glucosylated products

Glucose, fructose, leucrose and sucrose present in reaction medium were analyzed by HPAEC-PAD (high performance anion exchange chromatography with pulsed amperometric detection), using a CarboPac™ PA100 analytical column (4 mm x 250 mm) coupled with a CarboPac™ PA100 guard (4 mm x 50 mm). Product separation was performed using a sodium acetate gradient (6–500 mM) in 150 mM NaOH over 45 min (1 mL/min).

Tetrasaccharide and related glucosylated products were analyzed. The analyses were carried out on a LC-MS platform composed of a Dionex/Thermo Scientific™ HPLC UltiMate 3000 system with VW Detector, coupled to Thermo Scientific™ LTQ Orbitrap Velos™ hybrid FT mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA).

Reaction media were diluted 5 times in H₂O/acetonitrile (70:30, v/v) + 0.08% trifluoroacetic acid (TFA). The presence of residual tetrasaccharide **12** and glucosylated products (pentasaccharides **P1** and **P2** and poly-glucosylated products) was determined using a C₁₈RP

Fusion (4 μm , 80 \AA , 250 x 4.6 mm) analytical column placed in an oven at 30 $^{\circ}\text{C}$. The flow rate was set to 1 $\text{mL}\cdot\text{min}^{-1}$ and the solvent system consisted of H_2O , and acetonitrile (ACN). Elution started with an initial 20 % aq. ACN for 15 min, then was increased to 80 % aq. ACN within 15 min, which was held for 5 min, followed by a decrease to 20% aq. ACN within 0.1 min, held for 11 min to re-equilibrate the column. The injected sample amount was 10 μL and samples temperature was kept at 4 $^{\circ}\text{C}$. UV absorption spectra were monitored at 220 nm wavelength.

Mass detection was carried out in a positive heated electrospray ionization (HESI) mode. Mass spectrometer settings were as follows: spray voltage 4.0 kV, capillary and desolvation temperature were 400 and 450 $^{\circ}\text{C}$ respectively, maximum injection time 50 ms in Full Scan mode (500 - 1900 m/z). Nitrogen was used as sheath gas (pressure 75 units) and auxiliary gas (pressure 20 units). The automatic gain control (AGC) was set at $1\text{e}6$ for full scan mode with a mass resolution of 60,000 (at 400 m/z). Data acquisition was performed using Thermo Scientific Xcalibur software.

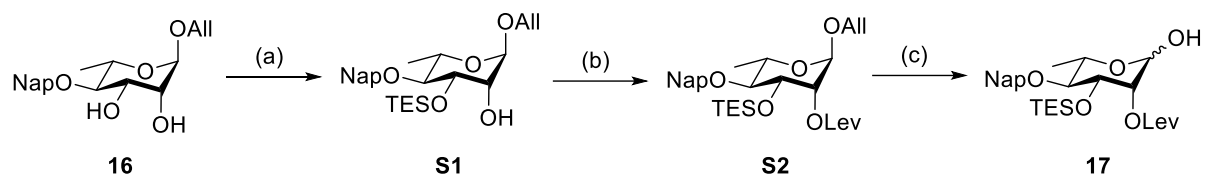
Determination of the structure of product P1 by NMR spectroscopy

The samples were dissolved in DCI-containing D_2O at pH 5.1. For NMR studies, the samples were lyophilized three times and dissolved in 180 μL of 99.9% DCI-containing D_2O .

All NMR spectra were recorded on a Bruker Avance spectrometer operating at a proton frequency of 950 MHz (TGIR- RMN-THC Fr3050 CNRS, Gif-sur-Yvette) and at a carbon frequency of 238 MHz with a 5-mm gradient indirect cryoprobe. All spectra were processed and analyzed with Topspin software (Bruker) and Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). ^1H and ^{13}C 1D NMR spectra were accumulated at 25 $^{\circ}\text{C}$, 65536 data points were acquired with 32 and 2048 scans respectively for proton and carbon experiments. ^1H - ^{13}C HSQC HMBC and COSY experiments were performed at 298 K. Homo and heteronuclear spectra were recorded under the following experimental conditions: 512 increments of 2048 complex points are acquired with an accumulation of 16 scans. Spectral widths were 16025 Hz for protons dimension and 44267 Hz for carbon dimension. The structural characterization of **P1** essentially derived from NMR analysis and was based firstly on the previous resonance assignments of the tetrasaccharide **12** NMR spectra. The comparison of the HSQC spectra (Figure 3A and B); showed shifted resonances of the glucosylated positions and adjacent atoms observed; the glucosylated 6D' carbon was shifted toward higher frequency from 61.78 ppm to 65.87 ppm, while the adjacent 5D' was shifted to lower frequency, from 76.23 ppm and 74.51 ppm, while the neighbor 4D' found slightly shifted around 0.1 ppm in its two ^1H and ^{13}C resonances. The cross-correlation peaks between shifted resonances were assigned using Heteronuclear Multiple-Bond Correlation spectroscopy (HMBC) and COSY experiments (Figure S4).

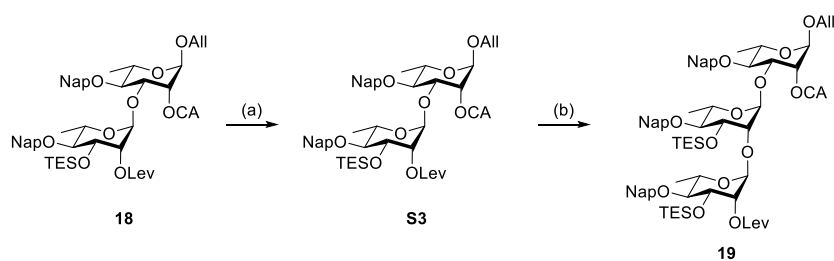
Supporting Schemes and Figures

Scheme S1: Synthesis of hemiacetal 17 from diol 16 with intermediates shown.¹



¹Reaction conditions: (a) TESCl, DIPEA, MeCN, 74%; (b) DCC, DMAP, LevOH, DCM, 74%; (c) (i) [Ir], THF, (ii) I₂/NaHCO₃, THF/H₂O 5:1, 89% (α/β 4:1).

Scheme S2. Synthesis of Trisaccharide 19 from the Fully Protected Disaccharide 18.¹



¹Reaction conditions. (a) NH₂NH₂·H₂O, Pyridine/AcOH 3:2; (b) **15**, TBSOTf, 4 Å MS, PhMe, -20 °C.

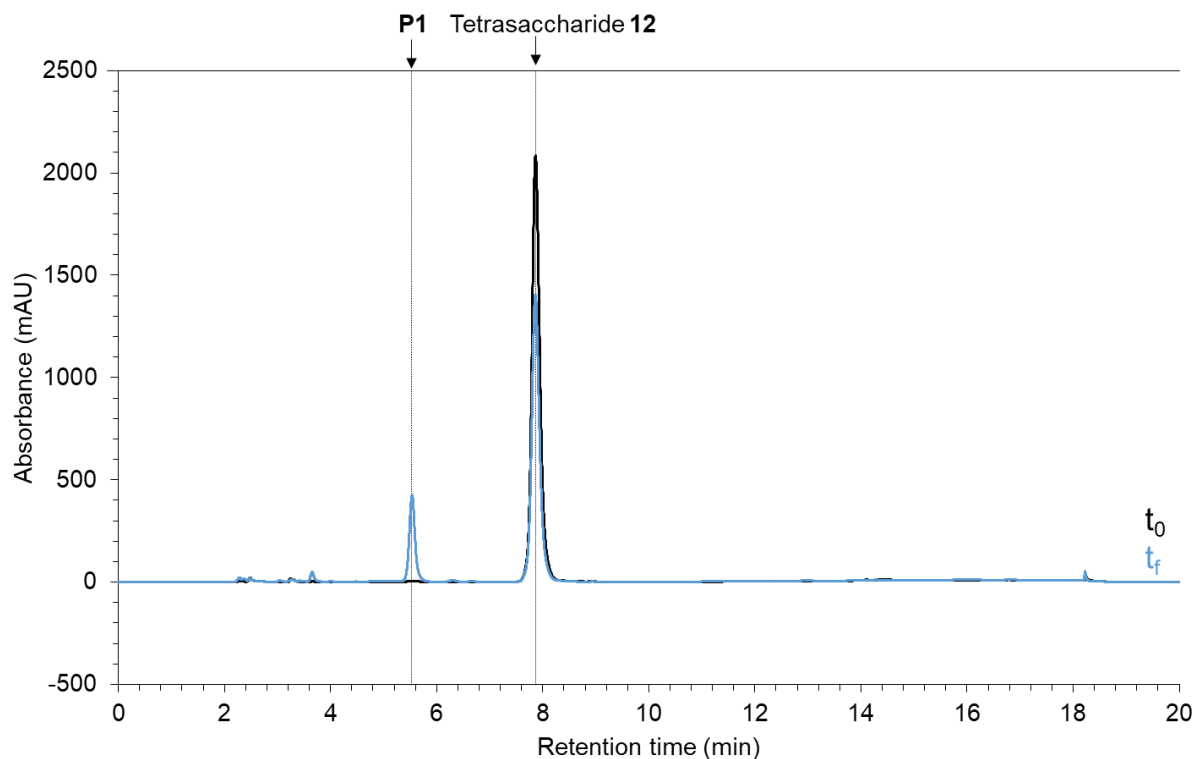


Figure S1: HPLC-UV chromatograms: comparison of initial (t_0) and final time (t_f) after 16 h of reaction in presence of BRS-B Δ 1 1 U/mL, sucrose (1 M) and tetrasaccharide **12** (**T12**) (50 mM) at pH 5.1 in 500 μ L reaction. The acceptor conversion yield was calculated as follows: $\frac{T12_{initial} - T12_{final}}{T12_{initial}} \times 100$: BRS-B Δ 1 yielded 31.3% conversion of the acceptor X. The main product of conversion was **P1**, which was further purified for structure elucidation by NMR.

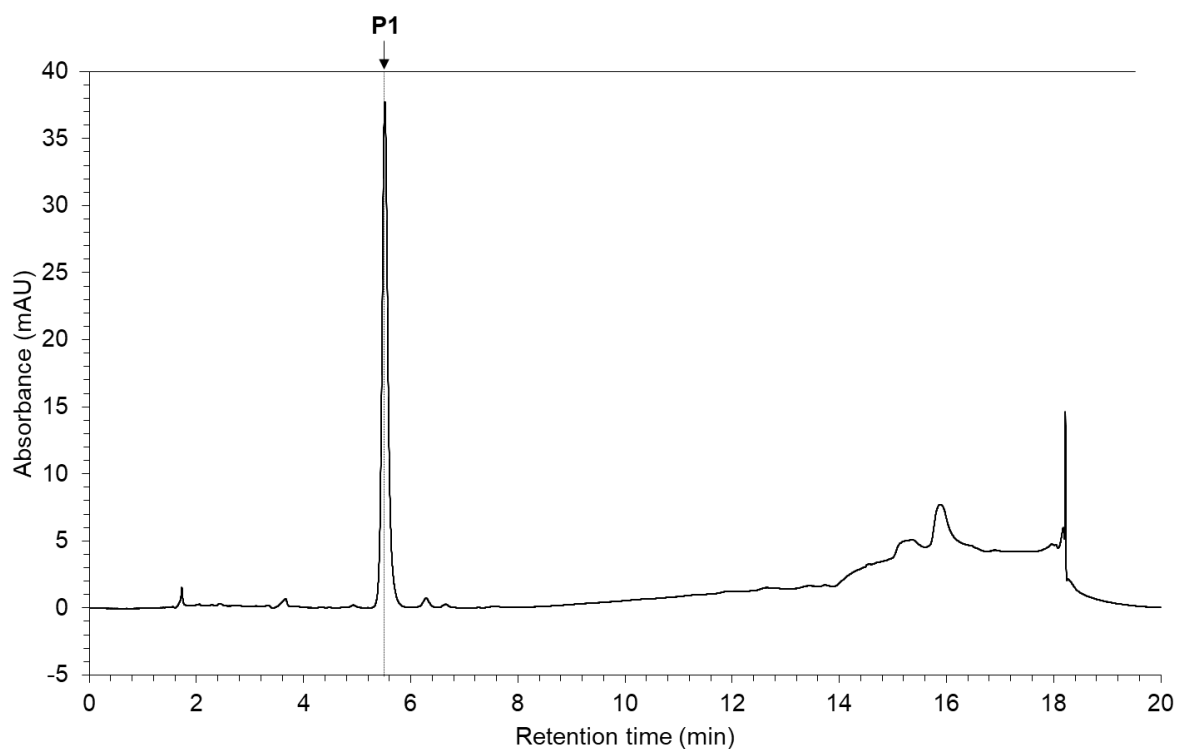


Figure S2: HPLC-UV chromatogram of the isolated **P1**.

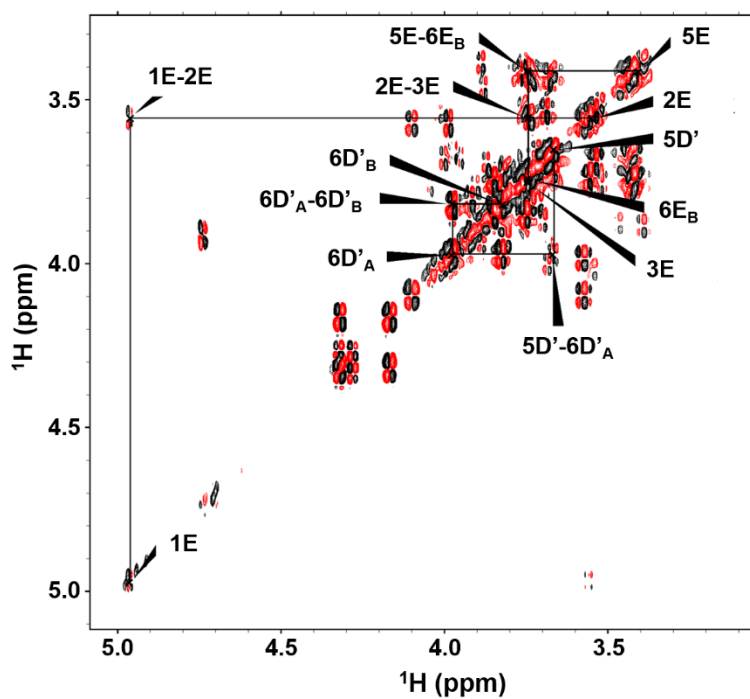


Figure S3: ^1H - ^1H QDF-COSY spectrum at 950 MHz of pentasaccharide **P1**. Positive and negative peaks are shown in black and red, respectively. The resonances of the E unit and the shifted resonances of the D' unit ($5\text{D}'$, $6\text{D}'_A$ and $6\text{D}'_B$) compared to those of tetrasaccharide **12**, are labeled.

Supporting Tables

Table S1. Transglucosylation of tetrasaccharide **12** using six recombinant branching sucrases of the GH70 family in optimized conditions (acceptor:sucrose ratio 1:20 (50 mM/1 M)). The ratio acceptor:sucrose used was 1:20 (50 mM/1 M) after 16 h of reaction with 0.4 U/mL of enzymes.

Enzyme	P1 (relative area %) ^a	P2 (relative area %) ^a	Poly-glucosylated products (relative area %) ^a	Conversion based on acceptor glucosylation (%) ^b	Conversion based on acceptor consumption (%) ^c	Sucrose consumption (%) ^c
BRS-A	0.42	1.37	0.66	2.45	6	97.2
BRS-D	0.81	1.23	3.28	5.32	11.2	100
ΔN_{123} -GBD- CD2	0.7	0.66	1.93	3.29	14.3	97.2
BRS-B $\Delta 1$	21.41	1.1	0.25	21.76	28.9	96.8
BRS-C	2.16	1.06	0.28	3.5	8.2	97.1
BRS-E	13.35	0.68	0.68	14.71	12.8	96.8

^a The relative area was calculated by dividing the area of the UV peak (or sum of the UV peaks for poly-glucosylated products) by the total area of the acceptor and product UV peaks.

^b Acceptor conversion calculated as follow: $Conversion = \frac{A_{P1}^{final} + A_{P2}^{final} + A_{Polyglucosylated}^{final}}{A_{12}^{initial}} \times 100$

^c The acceptor conversion was calculated using the formula: $Conversion = \frac{A_{12}^{initial} - A_{12}^{final}}{A_{12}^{initial}} \times 100$. The sucrose consumption was calculated after HPAEC-PAD

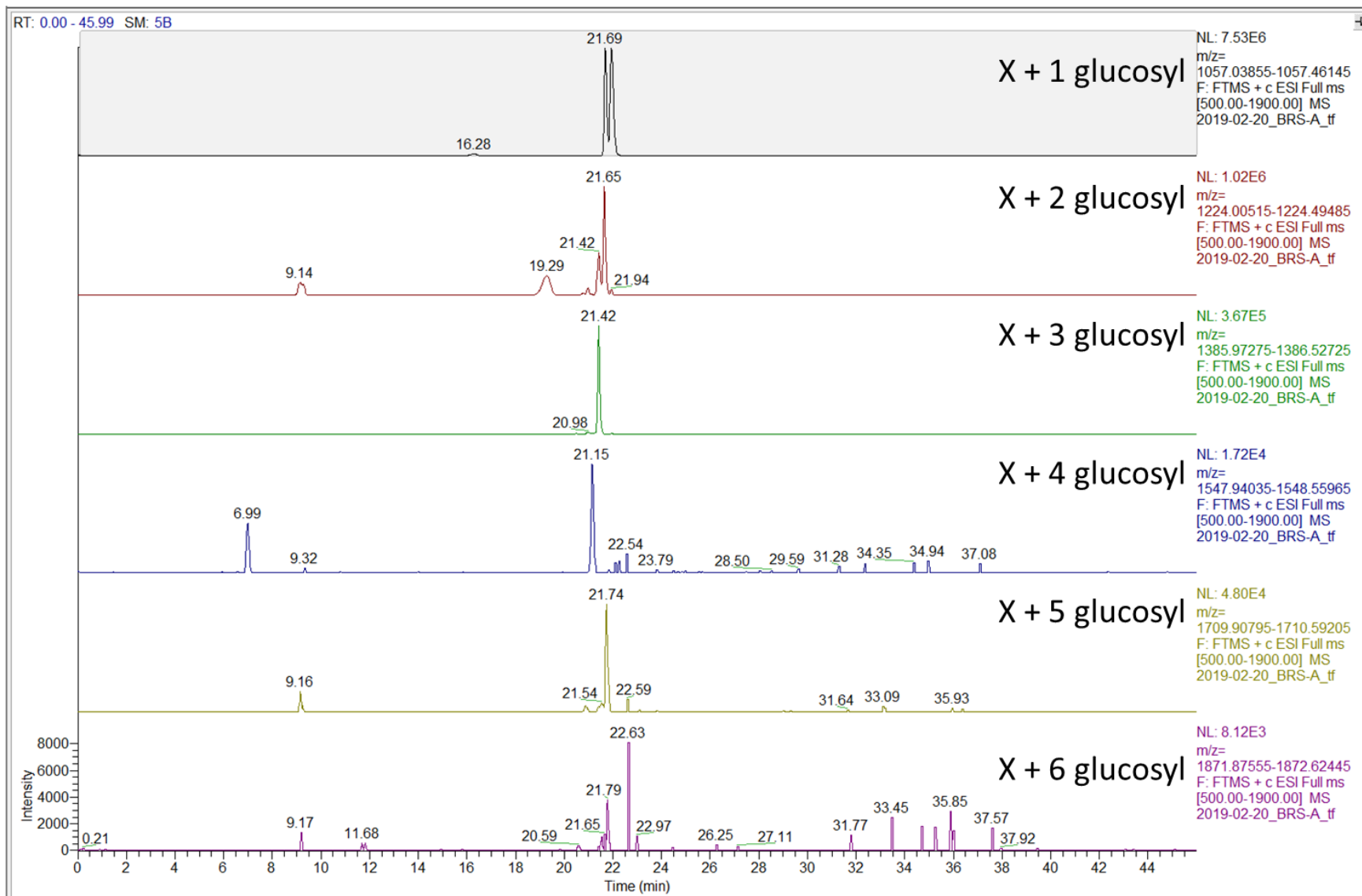
analyses using the formula: $Sucrose\ consumption = \frac{A_{sucrose}^{initial} - A_{sucrose}^{final}}{A_{sucrose}^{initial}} \times 100$.

^d The tetrasaccharide acceptor consumption of 24.8% was calculated for a 50 μ L scale reaction using 0.4 U/mL

MS data

The following figures show the high-resolution mass spectrometry analysis for each reaction medium. Mass ranges of specific ions corresponding to NH_4^+ or Na^+ adducts of the different glycosylation products were extracted. All data were in accordance with theoretical predictions of m/z for the different products. Trace amounts corresponding to an intensity below 10^5 were not considered for quantification.

m/z detected for BRS-A



NL: 7.53E6
 m/z=
 1057.03855-1057.46145
 F: FTMS + c ESI Full ms
 [500.00-1900.00] MS
 2019-02-20_BRS-A_tf

NL: 1.02E6
 m/z=
 1224.00515-1224.49485
 F: FTMS + c ESI Full ms
 [500.00-1900.00] MS
 2019-02-20_BRS-A_tf

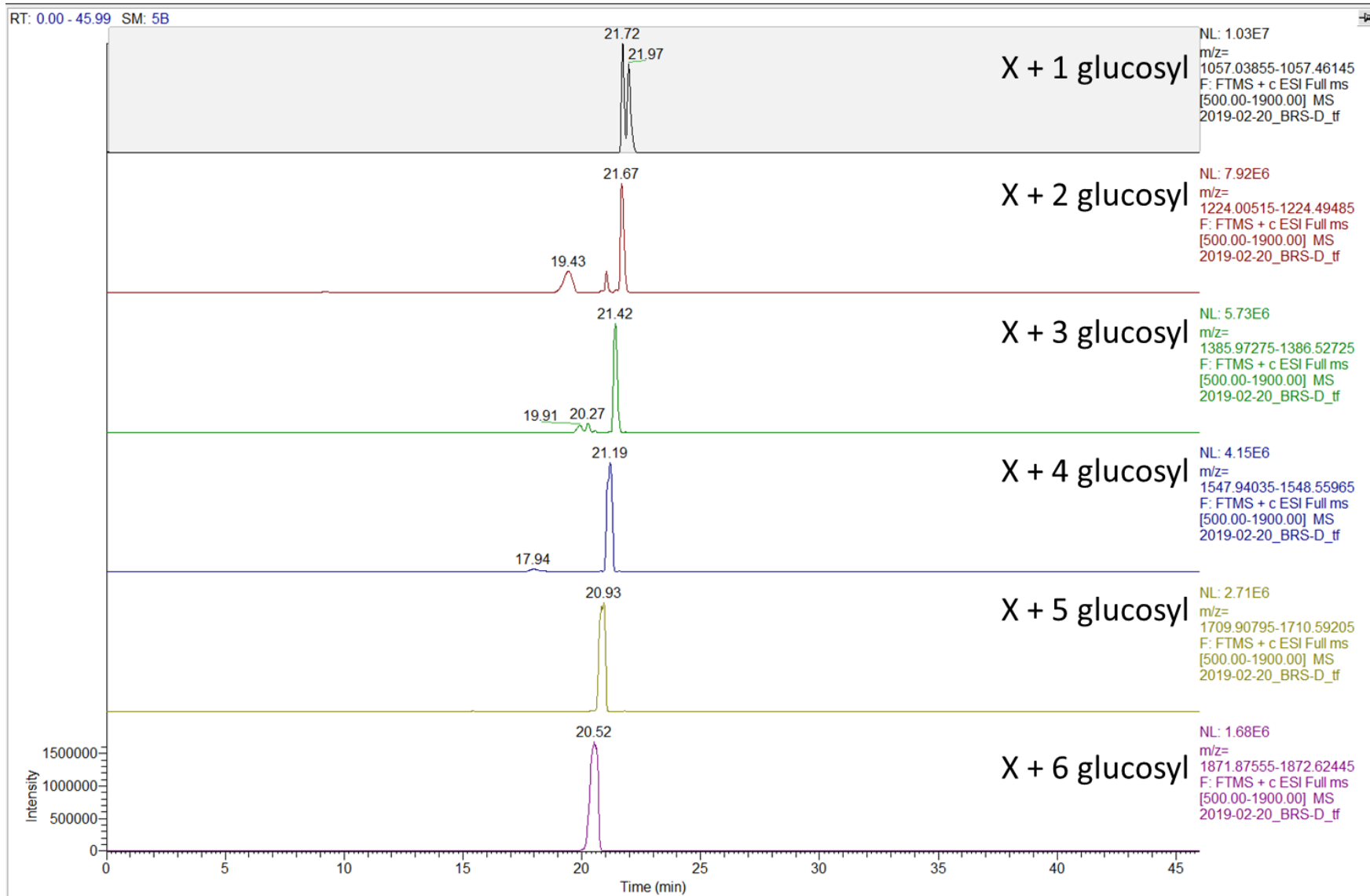
NL: 3.67E5
 m/z=
 1385.97275-1386.52725
 F: FTMS + c ESI Full ms
 [500.00-1900.00] MS
 2019-02-20_BRS-A_tf

NL: 1.72E4
 m/z=
 1547.94035-1548.55965
 F: FTMS + c ESI Full ms
 [500.00-1900.00] MS
 2019-02-20_BRS-A_tf

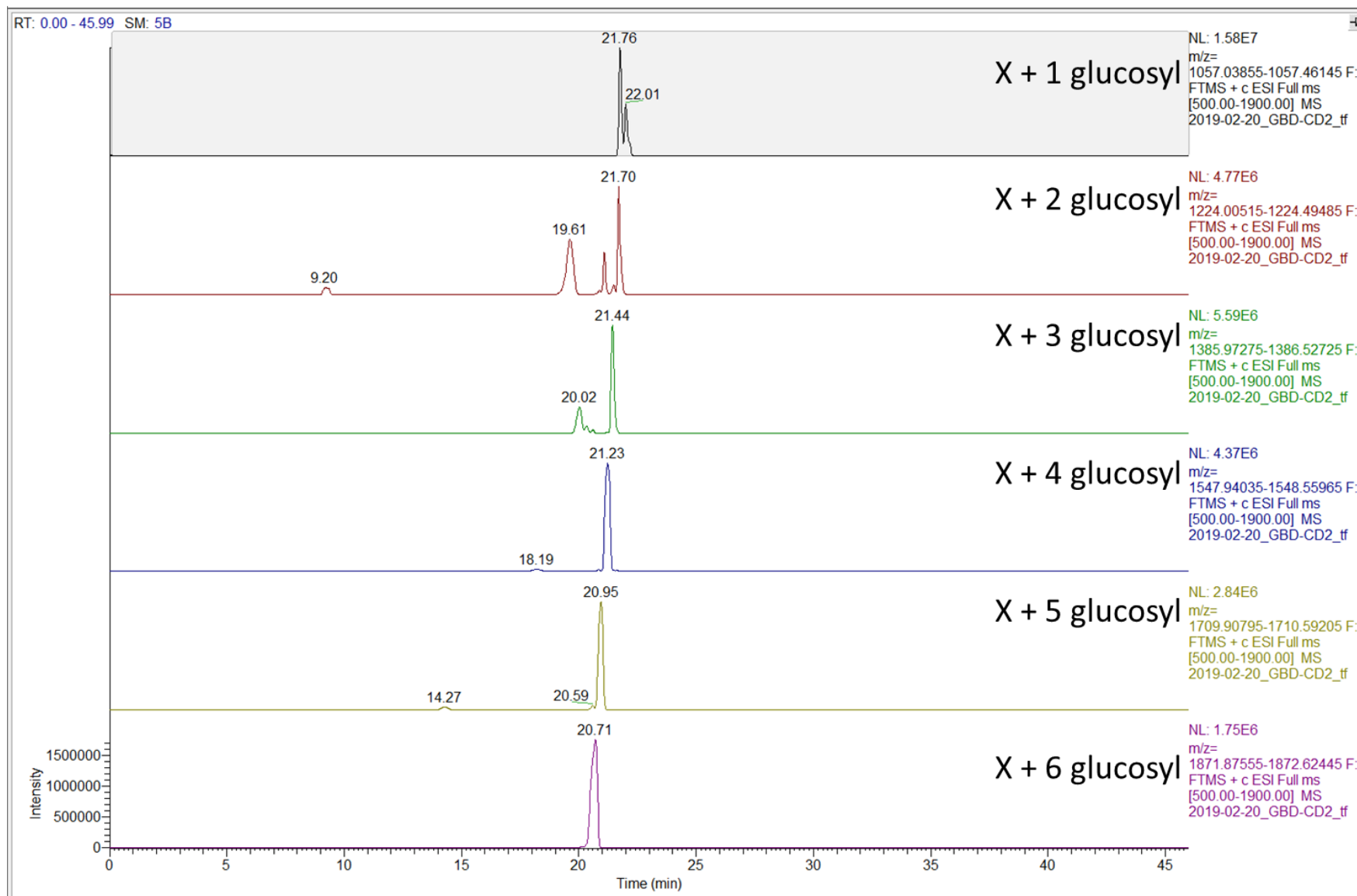
NL: 4.80E4
 m/z=
 1709.90795-1710.59205
 F: FTMS + c ESI Full ms
 [500.00-1900.00] MS
 2019-02-20_BRS-A_tf

NL: 8.12E3
 m/z=
 1871.87555-1872.62445
 F: FTMS + c ESI Full ms
 [500.00-1900.00] MS
 2019-02-20_BRS-A_tf

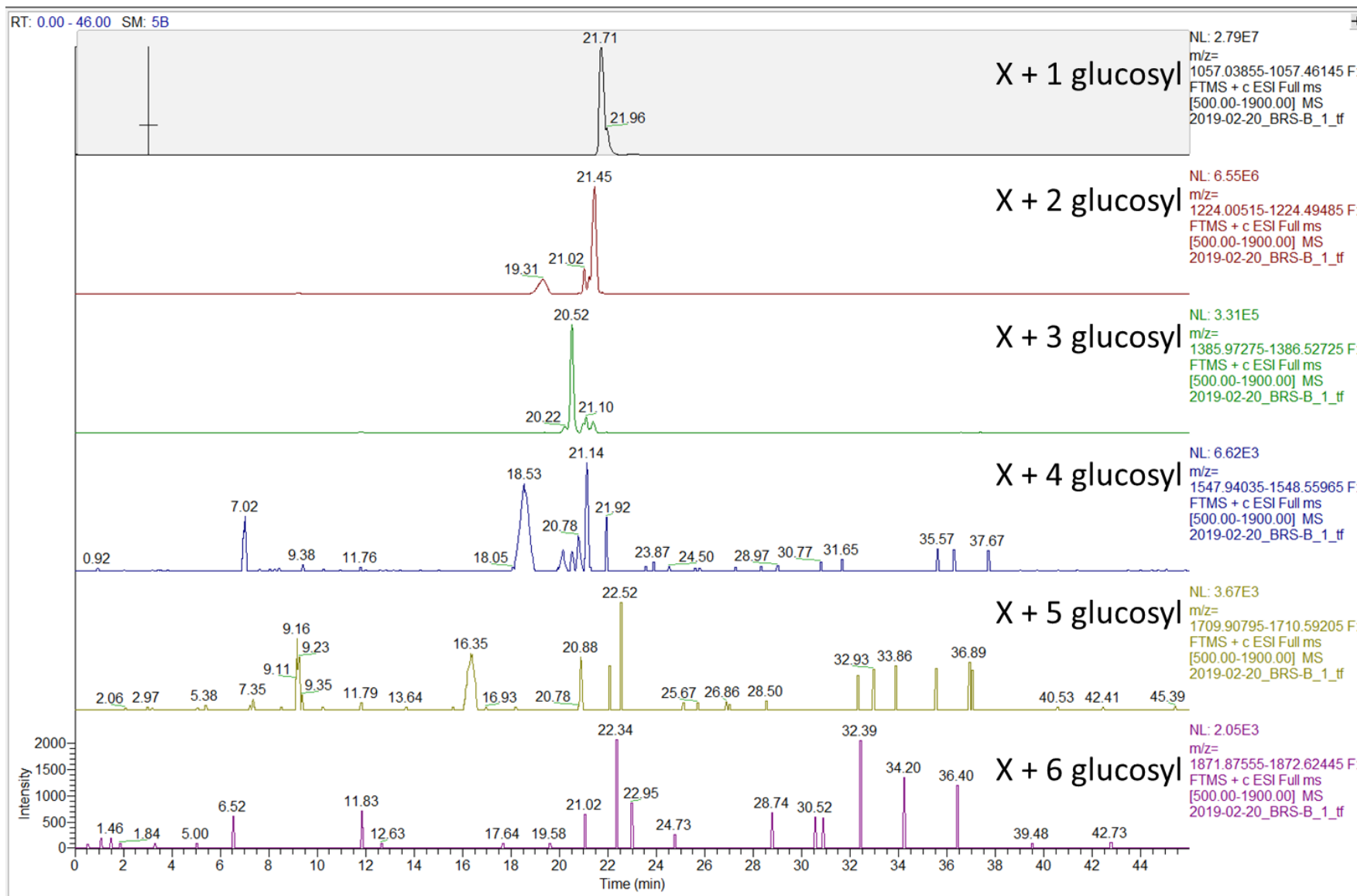
m/z detected for BRS-D



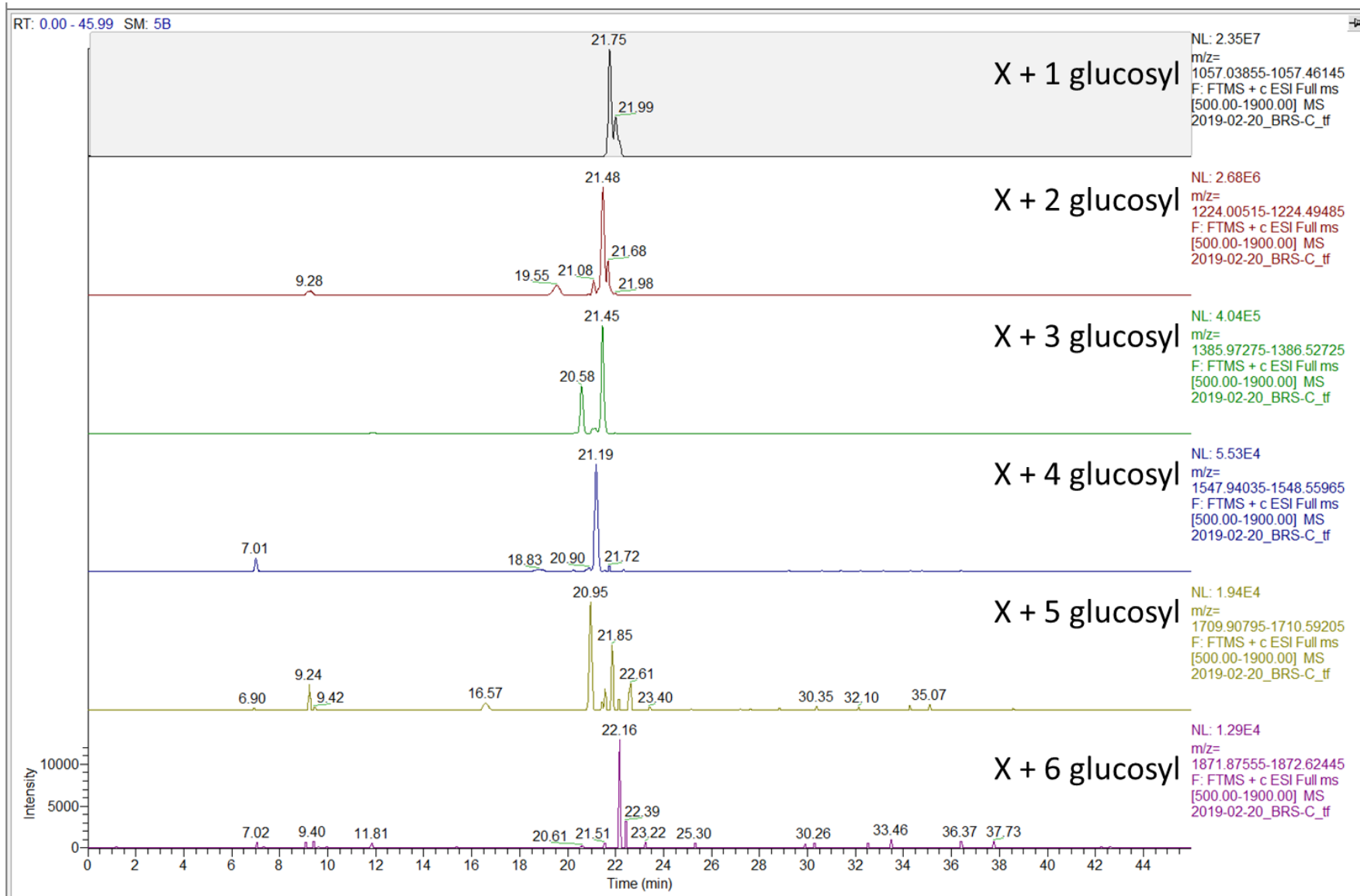
m/z detected for GBD-CD2



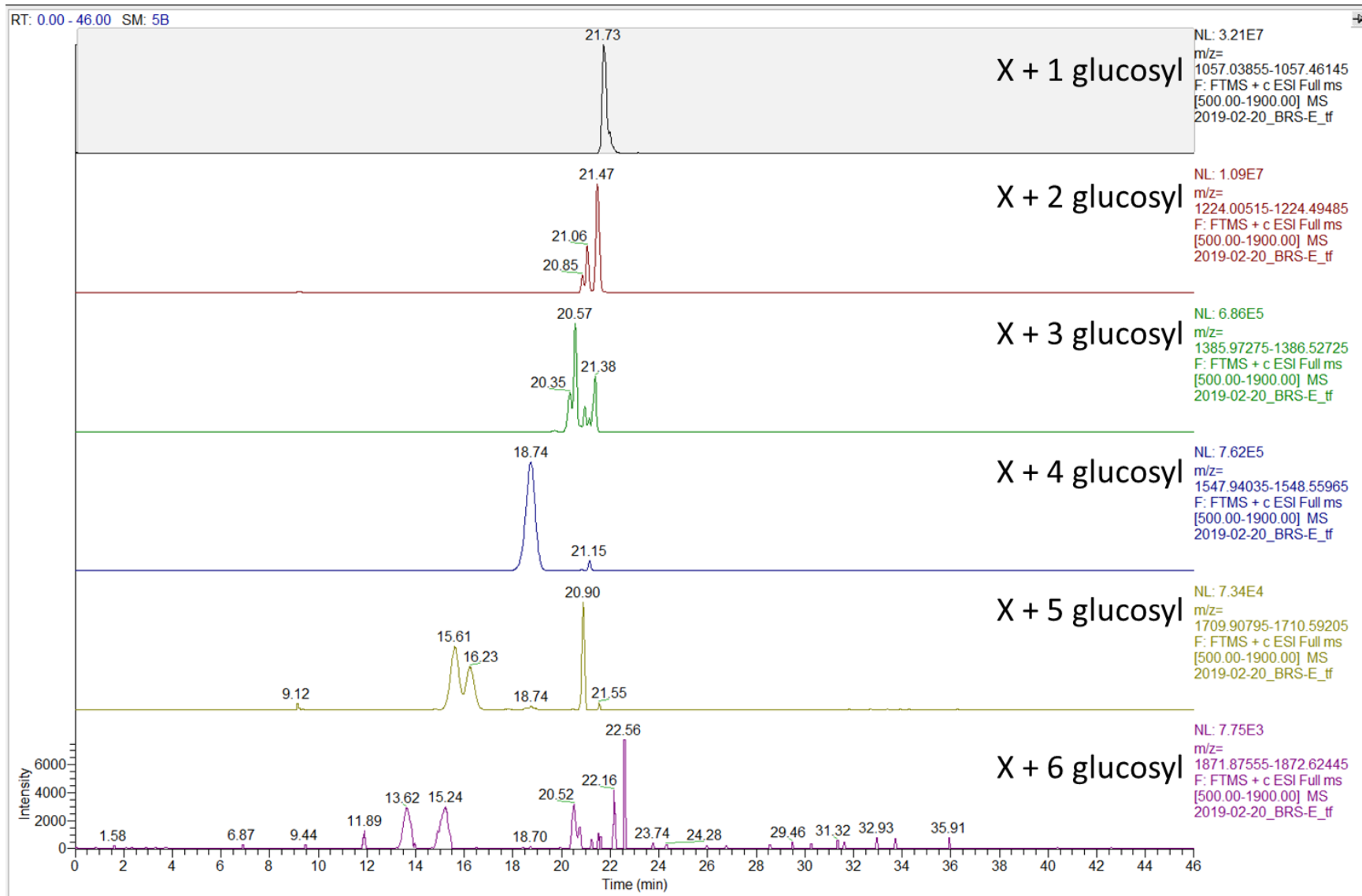
m/z detected for BRS-B



m/z detected for BRS-C



m/z detected for BRS-E



Complementary Results

This complementary part briefly presents additional results concerning the preliminary screening performed using the six native branching sucrases with sucrose and tetrasaccharide **12**, followed by an optimization of the reaction conditions resulting in those presented in the chapter II (article formatted for publication as a communication).

Preliminary screening

Transglucosylation assays were first performed at 30 °C in 50 mM sodium acetate buffer containing 0.34 mM CaCl₂, in a miniaturized 20 µL reaction volume at pH 5.75, which is the optimal pH for these enzymes, and supplemented with 0.4 U.mL⁻¹ of enzyme, 150 mM sucrose, and 50 mM tetrasaccharide **12** corresponding to an acceptor:sucrose molar ratio of 1:3. Reactions were incubated for 16 h. Analysis of the reaction products by HPLC-UV-MS revealed that all enzymes could readily transfer a glucosyl unit onto acceptor **12** to synthesize two mono-glucosylated products, namely **P1** and **P2**, confirmed by mass spectrometry (Figure C1). BRS-B Δ1 produced the highest amount of **P1** product. Another peak with a mass corresponding to tetrasaccharide **12** without its chloracetyl protecting group ([M+NH₄]⁺ = 819.27), and named tetrasaccharide **22**, was detected in all reaction media in significant amounts.

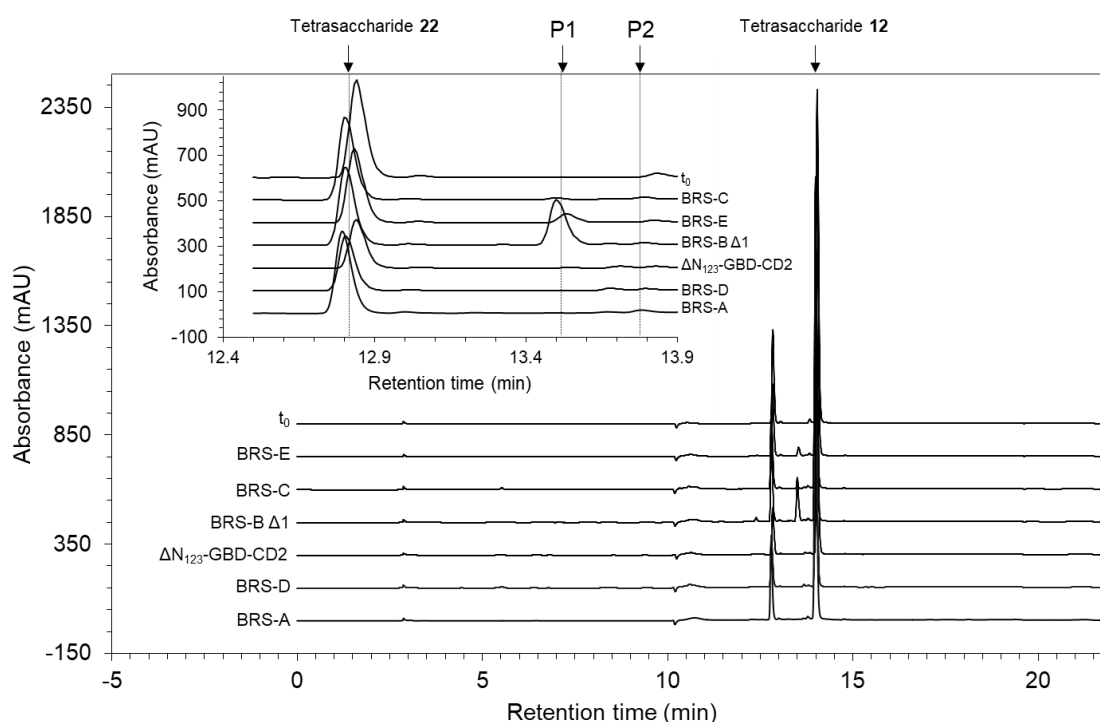


Figure C1: Chromatograms of the HPLC-UV_{220nm} analysis for the enzymatic glucosylation of tetrasaccharide **12** by six selected GH70 branching sucrases after 16 h of reaction in presence of sucrose (150 M) and tetrasaccharide **12** (50 mM) at pH 5.75. The region between 12.4 and 13.9 minutes is represented in the top left square. The initial time (t_0) is indicated as reference. **P1** (RT ≈ 13.5 min), **P2** (RT ≈ 13.8 min): mono-glucosylated products of the enzymatic glucosylation. A shift in molecular mass by 162 Da, characteristic of glucosylation, was observed for **P1** and **P2**. A molecule corresponding to a dechloroacetyl form of tetrasaccharide **12** (named tetrasaccharide **22**) was detected, in significant amounts.

In this preliminary screening, analytical separation was performed as follows: the reaction media were diluted 10 times in H₂O. The presence of residual tetrasaccharide **12** and glycosylated products (pentasaccharides **P1** and **P2**) was determined using a C₁₈RP Fusion (4 μm, 80 Å, 250 x 4.6 mm) analytical column placed in an oven at 40 °C. The flow rate was set to 1 mL.min⁻¹ and the solvent system consisted of H₂O, and acetonitrile (ACN). The gradient program was as following: it started with an initial 100 % H₂O for 15 min, then was increased to 67 % ACN within 2 min. Re-equilibration of the column was performed for 5 min with 100 % H₂O. The injected sample amount was 10 μL and samples temperature was kept at 4°C. UV absorption spectra were monitored at 220 nm wavelength.

As observed, retentions times were not consistent through multiple injections (for example for tetrasaccharide **22**, retention time varied between 12.75 and 12.85 min). This was attributed to sample dilutions in H₂O and more generally to the viscosity of the reaction media. Later, dilution in a mixture of H₂O and ACN allowed to obtain more consistent retention times.

Optimization of the sucrose and acceptor concentrations and their molar ratio

Optimization of the synthesis of the glycosylation products was carried out at pH 5.75 with 1 U/ml of BRS-B Δ1 enzyme in the presence of various acceptor initial concentrations (25, 50 and 75 mM, respectively), and different acceptor:donor ratio (1:3 and 1:12; 1:9, 1:12 and 1:20; 1:9 and 1:12, respectively) (Figure C2). The best compromise between acceptor consumption and appearance of the product was found to be 50 mM of tetrasaccharide **12** and 1 M of sucrose (ratio 1:20).

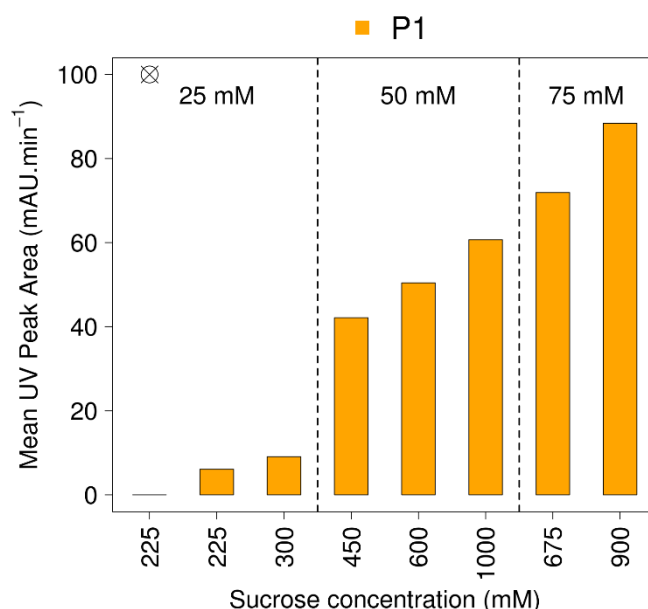


Figure C2: UV peak area of the product **P1** detected after reactions using BRS-B Δ1 at 1 U/mL, at pH 5.75. Optimization of the reaction conditions was performed by varying the initial sucrose concentration (225, 300, 450, 675, 900 or 1 000 mM) and the initial tetrasaccharide **12** concentration (25, 50, or 75 mM), resulting in different molar ratio [tetrasaccharide **12**]/[sucrose] (1:9, 1:12 and 1:20) Circle cross: negative control with no enzyme

Optimization of the pH of reaction

Instability of the tetrasaccharide **12** at pH higher than 5 could be an issue in larger-scale enzymatic synthesis of mono-glucosylated products. Previous work on the enzyme BRS-B $\Delta 1$ (Vuillemin et al. 2016) showed that it could cope with pH change and remain active between pH 3.5 and 7, with respectively 40 % and 30 % of relative activity compared to the optimal activity at pH 5.75. We therefore decided to lower the pH to a value of 4.5. As shown in Figure C3, tetrasaccharide **12** conversion is equivalent for reactions performed at pH 4.5 and 5.75 with the same amount of product detected, while only traces of tetrasaccharide **22** were detected in more acidic conditions (Figure C3). However having lost 60 % of relative activity compared to the optimal pH and looking for the best compromise between enzyme activity and tetrasaccharide **12** stability, we decided to use a pH of 5.1 (see Main results of this chapter).

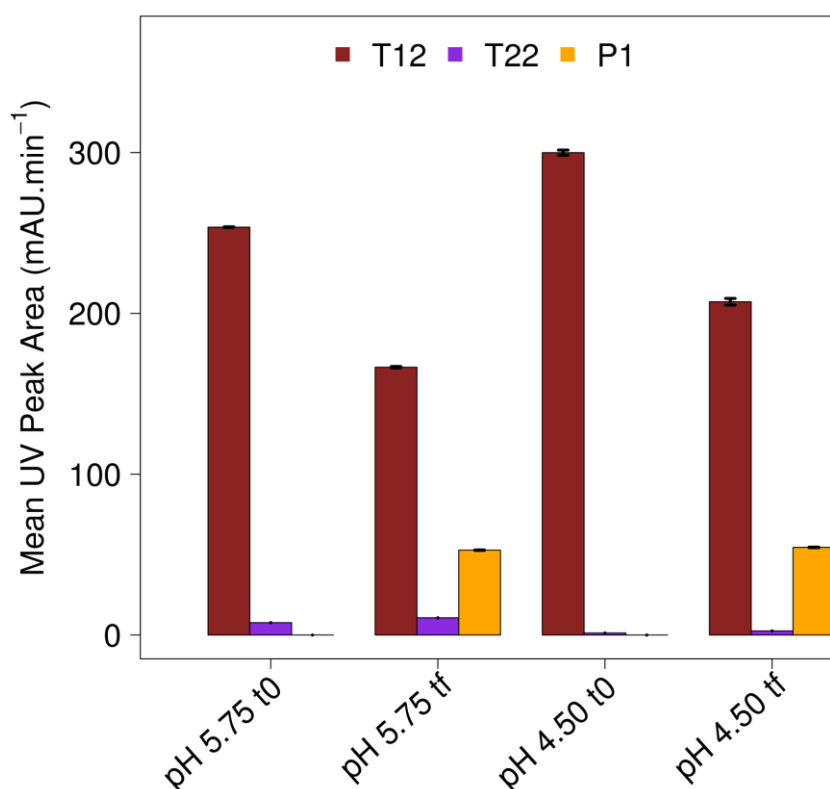


Figure C3: UV peak areas of the tetrasaccharide **12** (T12), tetrasaccharide **22** (T22) and the product **P1** detected in initial and final times of reactions at pH 5.75 and 4.5. Optimization of the pH of the reaction using 1 U/mL BRS-B $\Delta 1$ with 50 mM of tetrasaccharide **12** and 1 M of sucrose, during 16 h of reaction. Lowering the pH allowed a better stability of tetrasaccharide **12** with only traces of tetrasaccharide **22** detected. Conversion of tetrasaccharide **12** was not impacted and the same amount of product **P1** could be detected.

Similar analyses were performed by collaborators at Pasteur Institute (presented below).

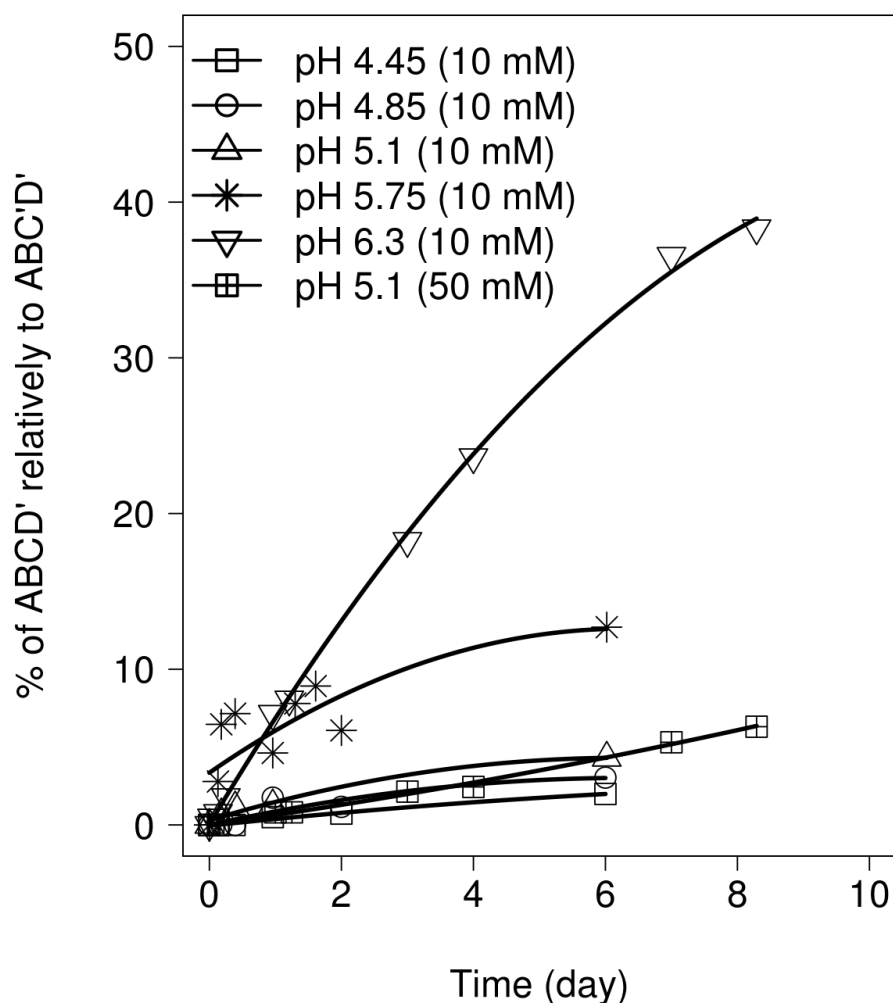


Figure C4: Stability analysis of tetrasaccharide **12** over time at 10 mM or 50 mM in an aqueous solution at different pH. Data reflect the proportion of dechloroacetylated tetrasaccharide **22** relative to the total tetrasaccharide content ($[\mathbf{22}]/([\mathbf{22}]+[\mathbf{12}])$) based on monitoring by RP-HPLC-UV_{210nm}.

Since the CA ester had to comply with conditions most suitable for a transglycosylase-mediated glycosylation, these observations prompted the evaluation of tetrasaccharide **12** integrity over time for a range of pH (4.45 – 6.3). RP-HPLC analysis demonstrated ABC'D' stability at pH below 5 (Figure C4). At pH 6.3, the hydrolyzed product **22** was clearly visible at six hours (1.8%), to reach 8% after a day while steadily increasing over time (over 37% after seven days). As expected, reducing the pH was found favorable. At pH 5.75, as used by default for enzymatic glycosylation mediated by GH70 α -transglycosylase, hydrolysis remained visible to provide tetrasaccharide **22** in close to 5% amount after 24 hours. At pH 4.45, dechloroacetylation was hardly visible reaching 0.5% and 2% after one and six days, respectively, while no other degradation product could be detected. Tetrasaccharide **12** was considered to be stable at this pH. At pH 4.85, the presence of tetrasaccharide **22** was detected to a 1.8% and 3% extent at one and six days, respectively, suggesting that the upper pH limit for long term stability was reached. This tendency was confirmed when running the analysis at pH 5.1. Under these conditions, partial dechloroacetylation was detected at 4 h (0.7%) to

then rise slowly over time (1.25% after 24 h). Moreover, increasing the concentration of ABC'D' by a factor five to reach 50 mM, had a positive impact and the extent of dechloroacetylation (0.8% of **22**) observed during a 24 h timeframe, which exceeds that of the enzymatic glucosylation step (16 h, see below), was considered to be acceptable (Figure C5).

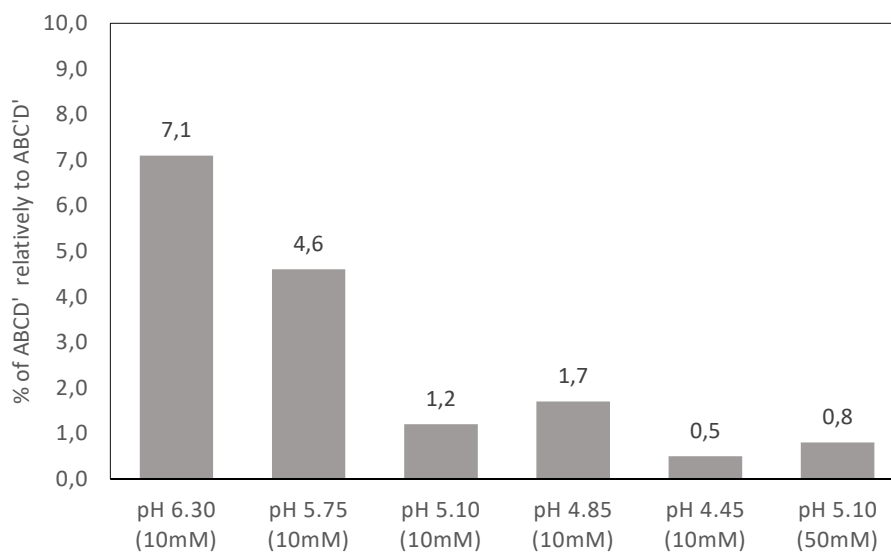


Figure C5: Stability of tetrasaccharide **12** in an aqueous solution (10 mM or 50 mM) at different pH as estimated from the proportion of tetrasaccharide **22** relative to the total amount of tetrasaccharide ($[\mathbf{22}]/([\mathbf{22}]+[\mathbf{12}])$), present at 24 h in the solution, based on RP-HPLC-UV_{210nm} analysis.

CHAPTER III

Summary: *This chapter presents the use of an available collection of mutants of the branching sucrose ΔN_{123} -GBD-CD2 for the glucosylation of the target tetrasaccharide molecule and the characterization of a product representative of *S. flexneri* prevalent serotype 3a, as well as the modelling studies performed to understand the impact of the mutations on tetrasaccharide recognition and glucosylation.*

Main achievements: *Modulation of the production of two mono-glucosylated products was achieved with mutants of ΔN_{123} -GBD-CD2, depending on the introduced mutation. Some mutants displayed higher product yields than parental enzyme. One mutant was selected for the production and characterization of the product **P2**, further shown to be representative of *Shigella flexneri* serotype 3a. Some mutants were found able to form novel products compared to the parental enzyme, namely **P2'**, **P3** and hexasaccharide **H1**. Effects of the mutations on the tetrasaccharide recognition and glucosylation were investigated using molecular modelling.*

Contributions to the work: *Tetrasaccharide was chemically synthesized by L-A. Barel at the Pasteur Institute.*

I produced the mutant enzymes in recombinant form, screened their activity, and purified the products as well as the mutants of interest for further characterization. All the modelling work was performed by A. Ben Imeddourene.

The structure of the pentasaccharide products was elucidated by 2D NMR by A. Ben Imeddourene. High resolution mass spectrometry was performed by H. Barbier.

Redirecting substrate regioselectivity using engineered ΔN_{123} -GBD-CD2 branching sucrases for the production of *S. flexneri* 3a, 4a and 4b pentasaccharide repeating units

Introduction

Carbohydrate-active enzymes have emerged as a practical synthetic alternative to chemical catalysts, avoiding multiple steps of protection and deprotection required in chemistry to control the reactivity of the sugar hydroxyl groups and the reaction regio- and stereo-selectivity. Although Nature has evolved enzymes to achieve specific needs, native carbohydrate-active enzymes are versatile biocatalysts, catalyzing a wide range of chemical reactions and often displaying naturally a relaxed substrate specificity. This promiscuity can be further exacerbated by enzyme engineering to either broaden or narrow down the range of recognized substrates and/or control the reaction selectivity (reviewed in (Benkoulouche et al. 2019)). In particular, mutagenesis targeting the enzyme active site has led to tremendous achievements to alter substrate specificity and diversify the structures of accessible glyco-products. In recent years, our group has been particularly active in this field, notably by engineering sucrose-active α -transglucosylases from GH13 and GH70 families, to produce a variety of carbohydrate derivatives and glyco-conjugates (Champion et al. 2012; Vergès et al. 2015; Champion et al. 2009a; Malbert et al. 2018). In particular, the amylosucrase from *Neisseria polysaccharea*, a GH13 sucrose-active enzyme, was purposely tailored to act on chemically modified substrates and produce glucosylated derivatives programmed to enter various chemo-enzymatic pathways combining in the most convenient manner chemical and enzymatic steps involved in the synthesis of antigenic oligosaccharides. In the continuity of these studies, we recently tested the ability of branching sucrases - native GH70 sucrose-active transglucosylases specialized in dextran branching - to glucosylate regio-selectively a lightly protected tetrasaccharide (tetrasaccharide **12** in Chapter II named herein **ABC'D'**) and obtain pentasaccharide precursors of *Shigella flexneri* serotype-specific haptens that could enter in the composition of broad coverage vaccines. Altogether six branching sucrases were assessed and found to produce majorly a mono-glucosylated product called **P1**, harboring the specific α -1,6 glucosylation pattern of *S. flexneri* serotype 4a/4b, trace amounts of a second mono-glucosylated tetrasaccharide named **P2** that remained uncharacterized due to limited availability, as well as a range of poly-glucosylated products. Among the tested enzymes, one of the most intriguing ones was the so-called ΔN_{123} -GBD-CD2 branching sucrose (Fabre et al. 2005; Brison et al. 2012), as it was the most efficient to glucosylate the tetrasaccharide of interest, although in a multitude of poly-glucosylated products in addition to **P1** and **P2**. This enzyme is however interesting as it is the only GH70 branching sucrose of known three-dimensional structure to date. Its molecular dynamics was studied in details through multi-level

modelling methods that revealed the high flexibility of several loops surrounding the catalytic pocket, which could play a beneficial role in the recognition of a broader range of acceptors. Furthermore, (semi-)rational engineering of this enzyme enabled the diversification of bulky flavonoid glucosides (Malbert et al. 2018). A small and focused mutant library targeting four amino acid residues from acceptor subsites +1, +2 and +3 of the catalytic pocket, was constructed by saturation mutagenesis. Screening assays enabled the identification of a small set of 22 highly promiscuous variants displaying remarkably improved activities toward several flavonoid acceptors (luteolin, morin, naringenin, quercetin, apigenin and chrysin) and producing a novel diversity of flavonoid glucosides.

In light of these encouraging results, we undertook herein the screening of this focused library for the glucosylation of tetrasaccharide **ABC'D'** in the hope that we can identify mutants with distinguished selectivity. Indeed, we hypothesized that the mutations introduced in the acceptor subsites could help to re-orient catalytically productive binding of the bulky tetrasaccharide, diversify the glucosylation pattern or improve the yield of the two pentasaccharides originally synthesized by the parental enzyme. Upon screening, we identified mutants that produced new compounds, pentasaccharide **P3** and hexasaccharide **H1**. Six mutants, containing single, double or triple substitutions were selected for further characterization as they produced higher amounts of **P1** and **P2** pentasaccharides than the parental ΔN_{123} -GBD-CD2 branching sucrose and/or the new products **P3** and **H1**. The structure of pentasaccharide **P2** was characterized and shown to be representative of the *S. flexneri* 3a prevalent serotype (Livio et al. 2014). The synthesis of both **P1** and **P2** showed that the introduced mutations drive the glucosylation toward one extremity or the other of the tetrasaccharide **ABC'D'** (**A** vs **D'**). To identify determinants possibly involved in the control of the regioselectivity, we used Molecular Dynamics (MD) simulations combined with in-depth analyses of amino acid residue networks.

This study highlights the potential of GH70 branching sucrose engineering for glucodiversification of a *S. flexneri* tetrasaccharide backbone as well as control of the reaction regioselectivity. These promising results pave the way for a more exhaustive reshaping of branching sucrose active site to further extend the panel of accessible glucosylation patterns and ultimately, provide synthetic tools for the development of broad coverage vaccines against *S. flexneri*.

Results & Discussion

Screening of the mutant library

The glucosylation of tetrasaccharide **ABC'D'** had to be conducted at pH values lower than the optimal pH (pH 5.75) of these enzymes (Passerini et al. 2015; Vuillemin et al. 2016) to avoid deterioration of the chloroacetyl protecting group (Chapter II). To that extent, we first checked

the activity of the parental and mutant enzymes on sucrose only at pH 4.7, for which the tetrasaccharide was stable. The activity decreased with pH for all the tested enzymes and some mutants were more affected than others (Figure 1). Without establishing the pH profile of each enzyme, it is difficult to provide a rational interpretation explaining the effect of changes in pH on each enzyme activity as both their K_M and V_{max} can be affected by the pH variations. Nonetheless, all the mutants retained activity and could thus be further tested for the subsequent glucosylation of the tetrasaccharide **ABC'D'**. To limit the quantity of tetrasaccharide used, the acceptor reactions were first miniaturized down to 10 μ L volume and performed at pH 4.7 in the presence of 50 mM of tetrasaccharide **ABC'D'** and 1 M of sucrose as previously described (Chapter II).

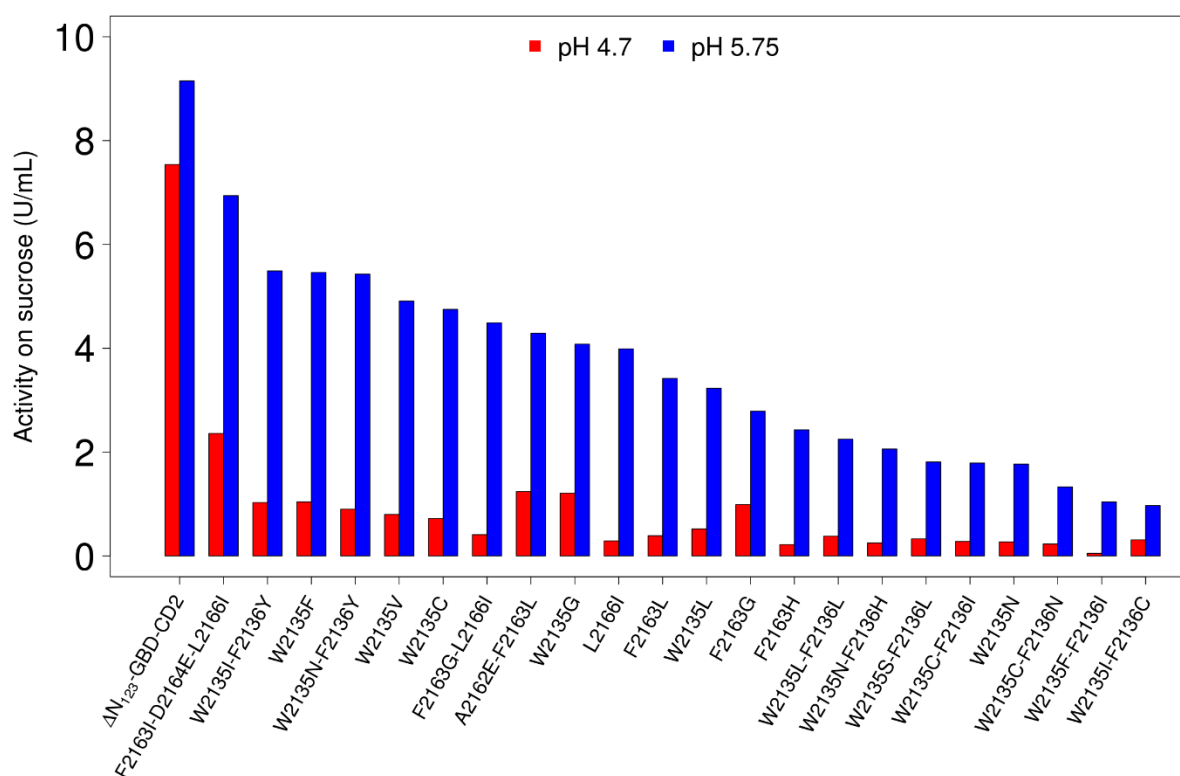


Figure 1: Determination of the activity on sucrose of parental ΔN_{123} -GBD-CD2 branching sucrose and its 22 mutants, at pH 4.7 (red) and pH 5.75 (blue). The activity was determined using the DNS assay with 100 g/L sucrose. The enzymes were more active at pH 5.75, consistent with results already determined for BRS-B $\Delta 1$ branching sucrose (Vuillemin et al. 2016).

With the exception of mutant W2135F-F2136I (the least active mutant on sucrose alone at pH 4.7), all variants were able to glucosylate the tetrasaccharide **ABC'D'** into two mono-glucosylated tetrasaccharides whose liquid chromatography retention times and mass value correspond to previously reported **P1** and **P2** pentasaccharides found to be synthesized by native branching sucraes (Chapter II). **P1** and **P2** were produced in varying amounts depending on the mutant (Figures 2 and S1).

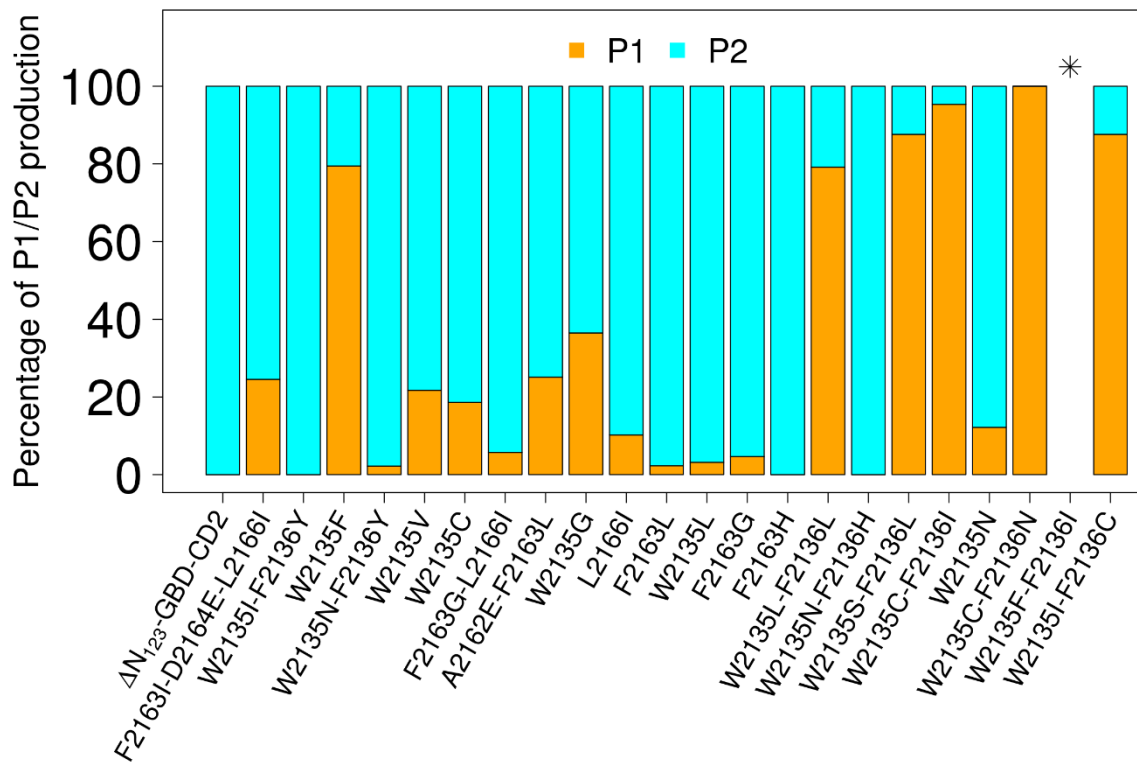


Figure 2: Stacked bar chart of the **P1** and **P2** amounts formed by ΔN_{123} -GBD-CD2 branching sucrose and its mutants from 10 μL reaction volume. Pentasaccharides **P1** (orange) and **P2** (blue). Results of mutant W2135S-F2136L were taken from the experiment performed at a 50 μL scale. * No glucosylation observed.

In these miniaturized conditions, parental ΔN_{123} -GBD-CD2 branching sucrose was found to exclusively produce **P2** while it was shown in Chapter II to produce **P1**, **P2**, and several poly-glucosylated products. Comparison of the product profiles led to the identification of three main families depending on the products mostly observed (Figure 2). The first group gathers 6 mutants (W2135F, W2135L-F2136L, W2135S-F2136L, W2135C-F2136I, W2135C-F2136N and W2135I-F2136C), out of the 21 active ones, which showed a partial or total loss of ability to produce **P2**, but produced however almost exclusively **P1**. Conversely, the second group comprises 6 mutants (F2163L-D2164E-L2166I, W2135V, A2162E-F2163L, L2166I, F2163G and F2163L) that produce **P2** in higher amounts than parental ΔN_{123} -GBD-CD2 while still forming **P1** in varying amounts. In the third group are found all 9 remaining mutants that either displayed similar product profile to parental enzyme (mutant W2135I-F2136Y) or lost most of their ability to glucosylate tetrasaccharide **ABC'D'**.

Interestingly, it should be noted that mutations favoring formation of **P1** were found at residues W2135 and F2136 (located in the first shell of the acceptor subsites +2/+3, in the characteristic motif V of GH70) while mutations generally favoring formation of **P2** were located in the first shell (L2166) and in the second shell (F2163) of the acceptor subsites +1 and +2, respectively. In the seminal work that led to the construction of the mutant library, these amino acids were targeted with the aim of gaining space in order to better accommodate bulky flavonoid molecules in the acceptor subsites (Malbert et al. 2018). We hypothesized that the same

mutations could also help to accommodate the large tetrasaccharide **ABC'D'**, which harbors bulky protecting groups, in a catalytically productive manner. It turned out that substituting the cumbersome residues W2135 and F2136 by smaller residues such as leucine or serine favored formation of **P1**. Conversely, the substitution of F2163 by aliphatic and more flexible residues (glycine, isoleucine, leucine) favored formation of **P2**. Other bulky residues like histidine (mutants W2135N-F2136H or F2163H) led to less formation of products than with the parental enzyme. Altogether, these results reveal the possibility to balance the production of **P1** and **P2** depending on the mutations introduced in the active site.

As previously mentioned, we did not detect the production of **P1** in the reactions carried out with parental ΔN_{123} -GBD-CD2 at 10 μ L scale and pH 4.7 which differs from results obtained with reactions in 50 μ L volume at pH 5.1 (Chapter II). Such discrepancies could result from the weaker activity observed on sucrose at pH 4.7 and less accurate screening in smaller volumes. To confirm the product profiles, we selected mutants distinguishable from their product specificity and reproduced the glucosylation experiment at 50 μ L scale and pH 5.1 to limit the effect of lowering pH on enzyme activity and stability.

Reaction scale-up of selected mutants

Three mutants producing **P2** in higher amounts (F2163G, L2166I and F2163L-D2164E-L2166I) and three mutants producing favorably **P1** (W2135S-F2136L, W2135I-F2136C and W2135L-F2136L) were used for scaling-up the reaction. Acceptor reactions were carried out at 50 μ L scale and pH 5.1 using 1 U/mL enzyme, 50 mM of tetrasaccharide **ABC'D'** and 1 M of sucrose. At this scale, differences were observed in the product profiles compared to those obtained for reactions performed at 10 μ L scale. First, parental ΔN_{123} -GBD-CD2 branching sucrose was here found able to produce also small amounts of **P1**, as well as poly-glucosylated products as previously reported in these conditions (Chapter II). We confirmed that the three enzymes mutated at positions 2135-2136 formed majorly **P1**. Interestingly, they also yielded a third mono-glucosylated product called herein **P3**, not observed in 10 μ L scale reactions and also not found for the parental enzyme. Detected only in traces, it was not possible to isolate enough **P3** for NMR structure elucidation. These three mutants also produced a poly-glucosylated tetrasaccharide, named **H1**, substituted by two additional glucosyl moieties. However, it remained unclear whether this double glucosylation occurred at the same or distinct positions. Of note, this hexasaccharide displayed a different retention time (t_R = 19.45 min) (Figure S2) from the poly-glucosylated products described in Chapter II (t_R = 21.45 min), suggesting different molecular structures. Interestingly, the poly-glucosylated products produced by the parental enzyme with up to 6 glucosyl units transferred were not produced by the mutants (Figure S2). Regarding the three other mutants targeting positions 2163 and 2166, we confirmed their inclination to produce preferentially **P2**, although only mutant F2163G was finally able to outperform parental ΔN_{123} -GBD-CD2 branching sucrose (Figure 3).

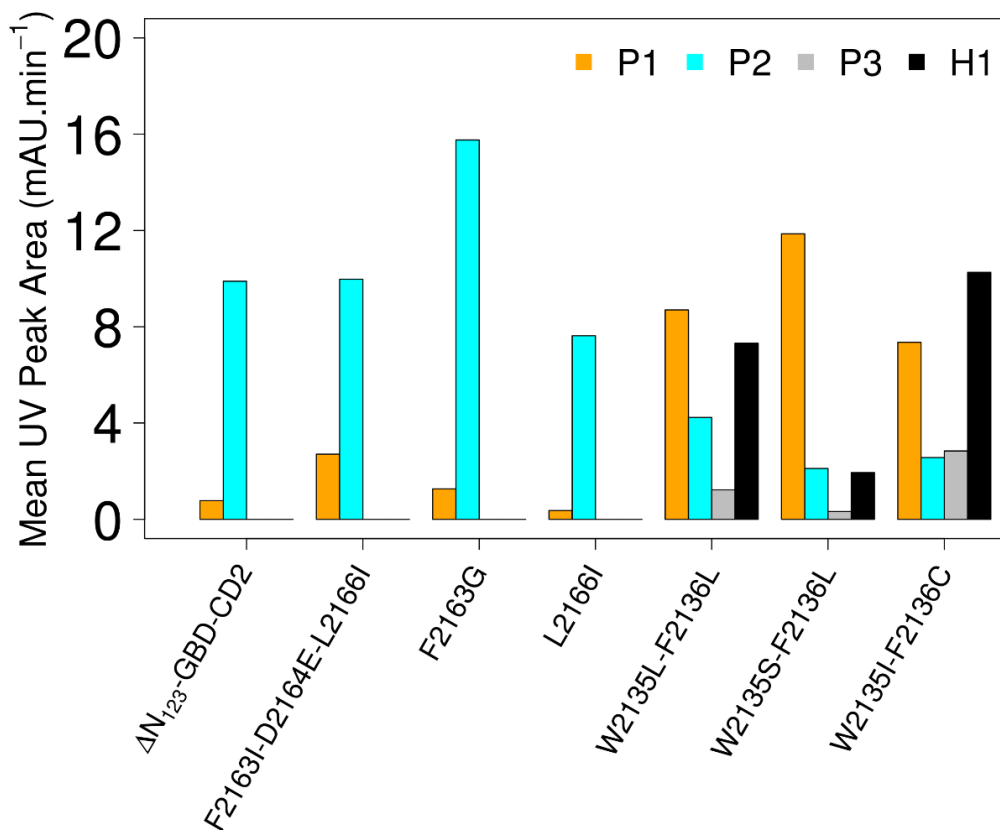


Figure 3: Detection of products formed by ΔN_{123} -GBD-CD2 branching sucrose and six selected mutants. The bar plot illustrates the amount of pentasaccharides (**P1** (orange), **P2** (cyan), **P3** (gray)) and the hexasaccharide (**H1** black) produced by ΔN_{123} -GBD-CD2 and the mutants F2136I-D2164E-L2166I, F2163G, L2166I, W2135L-F2136L, W2135S-F2136L and W2135I-F2136C, detected by mass spectrometry coupled with HPLC-UV (wavelength at 220 nm). Reactions were performed in presence of 1 M sucrose and 50 mM tetrasaccharide **ABC'D'**, during 16 hours reaction using 1 U/mL enzymes at pH 5.1 and 30°C.

Production and structural characterization of novel pentasaccharides

F2163G mutant was selected for further production of **P2** at larger scale. The mutant was purified to homogeneity and used in a 2 mL reaction in the presence of 50 mM of tetrasaccharide acceptor and 1 M sucrose donor, during 16 h of reaction. Reaction products were analyzed by HPLC-UV-HRMS which revealed that **P2** was in fact a mixture of two co-eluted products that we called respectively **P2** and **P2'**. A mixture of **P2/P2'** (~ 2 mg) was first isolated but it did not allow structural characterization of the products. Therefore, we undertook a new enzymatic reaction followed by a two-step purification of **P2** and **P2'**. About 5 mg of **P2'** could then be obtained which allowed its structure elucidation by NMR analysis in combination with previous resonance assignments of **ABC'D'** NMR spectra. Data showed that glucosylation of **ABC'D'** occurred at OH-4_A, leading to allyl [α -D-glucopyranoside-(1→4)]- α -L-rhamnopyranosyl-(1→2)- α -L-rhamnopyranosyl-(1→3)-2-O-chloroacetyl- α -L-rhamnopyranosyl-(1→3)-2-deoxy-2-trichloroacetamido- β -D-glucopyranoside, also named **E(1→4)ABC'D'**, or **P2'** in HPLC analysis. **P2** was further characterized to be glucosylated at OH-3_A, corresponding to allyl [α -D-glucopyranoside-(1→3)]- α -L-rhamnopyranosyl-(1→2)- α -L-rhamnopyranosyl-(1→3)-2-O-chloroacetyl- α -L-rhamnopyranosyl-(1→3)-2-deoxy-2-

trichloroacetamido- β -D-glucopyranoside, also named **E(1 \rightarrow 3)ABC'D'**, or **P2** in HPLC analysis (Figure 4).

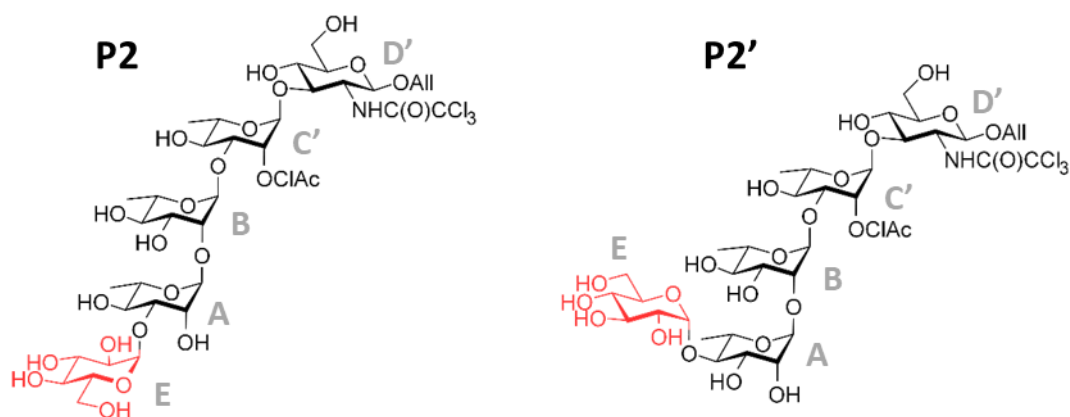


Figure 4: Structures of pentasaccharides **P2** and **P2'** (E(1 \rightarrow 3)ABC'D' and E(1 \rightarrow 4)ABC'D', respectively) characterized by NMR spectroscopy.

In details, analysis of HSQC spectra showed shifted resonances of the glycosylated positions and adjacent atoms observed (Figure S3); the glycosylated 4A carbon was shifted toward higher frequency from 72.17 ppm to 81.21 ppm, while the adjacent 3A and 5A were shifted to lower frequency, from 70.12 ppm and 69.21 ppm to 68.28 ppm and 68.38 ppm, respectively. The cross-correlation peaks between shifted resonances were assigned using Heteronuclear Multiple-Bond Correlation spectroscopy (HMBC) experiments (Figure S3C), and the Double Quantum Filtered COReLation SpectroscopY (QDF COSY) experiments (Figure S4).

The resonance assignments of **P2** was extracted from NMR spectra representative of a mixture of products (referenced below as a mixture including pentasaccharides **P2** and **P2'**, as well as at least one dechloroacetylated pentaccharide) and was performed firstly by the subtraction of signals known to belong to the spectra of the **P2'** isomer (Figure S5A). The E units belonging to **P2** resonance was found by subtraction of **P2'** and dechloroacetylated tetrasaccharide signals. All connectivities of the latter E units were found by QDF-COSY spectrum (Figure S6). The inspection of the HSQC spectrum showed that only the resonance of 3A from the non-superimposed tetrasaccharide peaks, was shifted, the 4A and 4B were superimposed in the tetrasaccharide spectrum. This observation suggested two assumptions; the glycosylation occurred on 3A or on 2A of the tetrasaccharide. It is known that a glycosylation leads to slightly lower frequency shift of the adjacent resonance of the glycosylated position. However, no peak was observed in the anomeric region (Figure S5B), which confirmed that the glycosylation position is the 3A. The HMBC spectrum showed the cross-correlation peak of the E units and the assigned 3A resonance.

Structural insight on the impact of the mutations on the regioselective glucosylation of tetrasaccharide **ABC'D'**

Results described above revealed that glucosylation can occur either on **D'** (**P1**) or **A'** (**P2**) moiety of the tetrasaccharide depending on the mutations introduced in the active site of ΔN_{123} -GBD-CD2 branching sucrases. This suggests that tetrasaccharide can bind in at least two distinct modes in the active site and still be glucosylated by the enzyme. To further investigate at atomic level the versatile binding mode of the tetrasaccharide acceptor, we used computational methods to understand the influence of the mutations on activity toward sucrose donor and on the selective glucosylation of acceptor substrate (Figure 4). In particular, MD simulations (of various lengths: 1 μ s, 100 ns and 2 ns) were carried out in free form or in complex with sucrose or formed products **P1** or **P2**, for parental ΔN_{123} -GBD-CD2 branching sucrase, and its mutants F2163G, W2135S-F2136L, and W2135I-F2136C. The free energy landscapes (FELs) from the two first eigenvectors of each 1 μ s MD simulation reflect the structure and dynamics perturbation due to the mutations on the global motion of the enzymes (Figure 5). Interestingly, the mutants W2135S-F2136L, and W2135I-F2136C showed a high resemblance in their FEL; the lowest energy basins shifted to the positive value for the first eigenvector. Furthermore the negative regions of this eigenvector were more populated than the parental enzyme, while the second component does not seem affected by the mutations. The F2163G mutant exhibited the broadest FEL indicating the high flexibility of this mutant. The principal component analyses, performed on the whole enzymes, reported therein provide the trend of the global dynamics modes. However, to understand the impact of the mutations on product profiles, further analyses on the active sites should be carried out. Thus, MD simulations were analyzed using Fruchterman-Reingold algorithm and the spin-glass algorithm to identify the amino acid network in the active site and define corresponding amino acid clusters (also called communities). Network analysis is often used in biology to understand cell functional organization via protein network or to clusterize metagenomics sequences (Pell et al. 2012; Marbouty et al. 2014). Network analysis derived from MD simulations (correlation matrix) were previously used to identify with success the signaling pathways in bacterial glutamyl-tRNA synthetase (GluRS):tRNA^{Glu} and an archaeal leucyl-tRNA synthetase (LeuRS):tRNA^{Leu} complexes (Sethi et al. 2009). In the current study, we used graph theory to map the distance matrices of the active site residues and examine the effect of the mutation on the catalytic pocket topology, and possibly relate it to the product profile. To our knowledge, this is the first time such approaches are used in this context, providing essential information on the catalytic behavior of this family of enzymes.

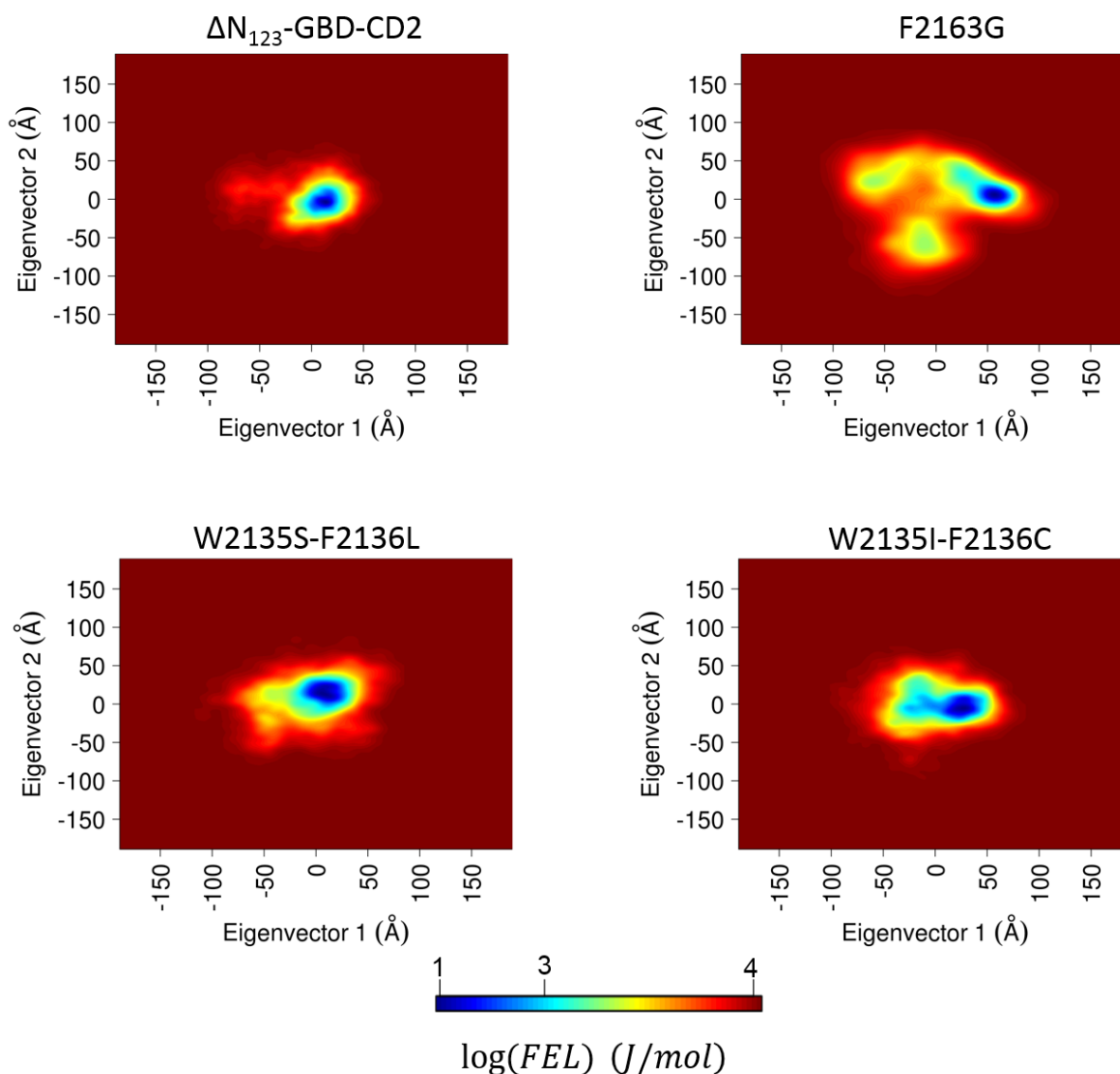


Figure 5: Free energy landscape (FEL) of ΔN_{123} -GBD-CD2, F2163G, W2135I-F2136C and W2135I-F2136C were determined using as reaction coordinates the projection of the first and second principal components from one microsecond MD simulation of the free ligand form for each system. The bottom legend shows the color scale of the logarithm of FEL in $J \cdot mol^{-1}$

Except for the parental enzyme, the clustering algorithm found 5 communities (I (orange), II (green), III (red), IV (yellow) and V (cyan)) (Figure 6); for all mutants the residues 2689 to 2694 form the cluster II, while the clusters I (orange) and II (green) were gathered in one single large cluster in the case of the parental enzyme. The F2163G mutation weakened the interaction inside the community III compared to the parental enzyme and other mutants. Furthermore, the cluster III in the case of the F2163G mutation had a different localization, nearby the cluster V. Remarkably, this proximity between these two communities situated on both sides of the active site led to the clear separation of clusters I and IV (Figure 6). Interestingly, mutants W2135S-F2136L and W2135I-F2136C, which display the same product profile, showed a high similarity in their networks and community interactions: (i) all the vertices of the cluster II are linked to the two clusters I and IV; (ii) all nodes belonging to the cluster III are connected to

each other with the exception of S2135 and L2167 residues in the case of mutant W2135S-F2136L; (iii) the two leucines 2166 and 2167 from cluster III bridged the clusters I, II and IV except the L2167 residue in the case of mutant W2135S-F2136L; (iv) the community V is connected to cluster IV and not to cluster I.

In the present study, the population shifts observed in Figure 6 could help to explain the variability of product profiles between parental enzyme and its mutants. Furthermore, our graph analysis illustrates the thermodynamical displacement toward one conformation or another among the different mutants. Preexistence of such conformation changes is also underlined by the correlation found between the network derived from MD simulations and the product profiles of mutants W2135S-F2136L and W2135I-F2136C obtained from distinct acceptors (tetrasaccharide studied herein, Figure 2, and flavonoids (Malbert et al. 2018)).

Active site networks are represented on 3D structures of the free ligand forms of the enzymes (Figure 6) and mutations are highlighted in the modelled complexes with pentasaccharides (Figure 7). Mutation of the bulky residues W2135 and F2136 clearly results in a gain of space that facilitates the accommodation of the bulky protecting groups of residue D' and allowed the acceptor to be oriented with the OH-6_{D'} in a catalytically more favorable conformation to produce **P1** (Figure 7). The higher amount of **P1** obtained for variant W2135S-F2136L in comparison to variant W2135I-F2136C (Figure 3) could result from the re-orientation of S2135 residue toward the solvent (Figure 7), further widening the +1 subsite. This conformation is reflected by the absence of edge between the S2135 and L2167 mentioned above (Figure 6). The flexibility of the loops surrounding the active site in the mutant F2163G was increased due to the introduction of the glycine residue that could destabilize the interactions mediated by this F2163 residue and thus leading to variations of the RMSD and high B-factor values observed for F2163G mutant compared to parental enzyme (Figure S7). The population shift in this case led to more proximity between the clusters III and V. In this topological organization, the displacement of the loop containing residues W2135 and F2136 toward community V disfavors the accommodation of the tetrasaccharide with **C'** and **D'** moieties in productive conformation for **P1** production. Indeed, the simulation illustrates that the closeness of the loop containing the residues W2135 and F2136 to L2166 and L2167 from cluster V, promoting a bent conformation of the pentasaccharide, with the D' well stacked to the W2135 and F2136 residues through strong C-H π interactions. These favorable interactions could compensate for the entropy loss due to this conformation bending in comparison to the parental enzyme for which the **C'** and **D'** moieties are found rather exposed to the solvent. Altogether, this could help to better glucosylation of the **A** moiety in mutant F2163G.

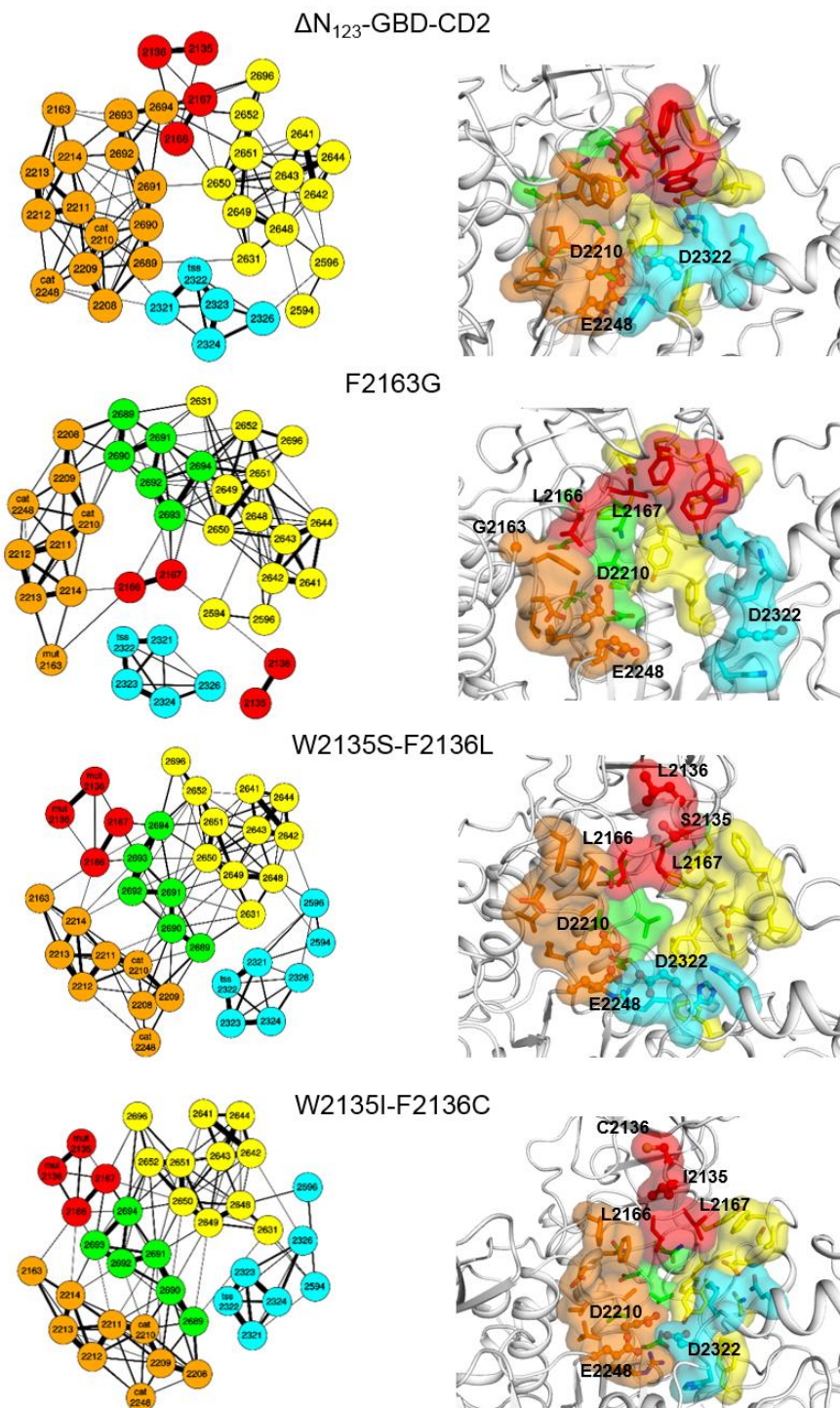


Figure 6: Comparison of network communities (left) and structural representation (right) derived from MD analysis.

Left: Active site residues network presented by spring layout using Fruchterman-Reingold algorithm from distance matrixes from 1 μ s MD simulations of the parental ΔN_{123} -GBD-CD2, the F2163G mutant, and the double mutants W2135S-F2136L and W2135I-F2136C. Amino acid communities highlighted with orange, red, green, yellow and cyan colors were found by spin-glass algorithm, the catalytic, the transition state stabilizer, and the mutated residues are labelled respectively by “cat”, “tss” and “mut” for each corresponding graph. Right: View of active site residue clusters from the spin-glass algorithm for the parental ΔN_{123} -GBD-CD2 and the mutants F2163G, W2135S-F2136L, and W2135I-F2136C. The 5 communities are presented by I (orange), II (green), III (red), IV (yellow) and V (cyan). Three dimensional structures are taken from the low energy regions of the free energy landscapes of each system (**Figure 5**).

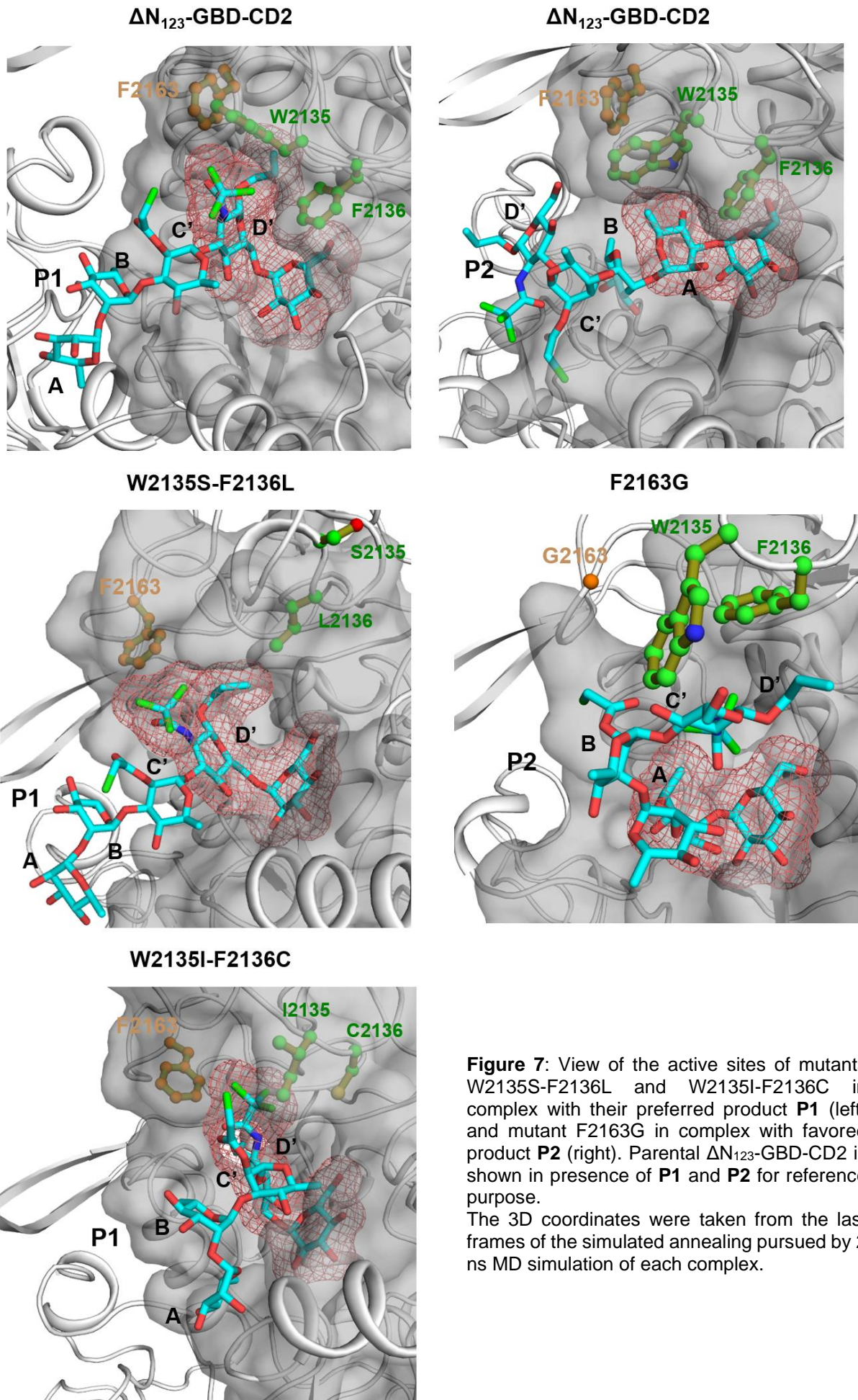


Figure 7: View of the active sites of mutants W2135S-F2136L and W2135I-F2136C in complex with their preferred product **P1** (left) and mutant F2163G in complex with favored product **P2** (right). Parental ΔN_{123} -GBD-CD2 is shown in presence of **P1** and **P2** for reference purpose. The 3D coordinates were taken from the last frames of the simulated annealing pursued by 2 ns MD simulation of each complex.

Effect of mutations on the sucrose kinetics of mutants

To investigate further the effect of mutations on sucrose utilization, we determined kinetic parameters on sucrose of the parental ΔN_{123} -GBD-CD2 branching sucrose and its mutants F2163G and W2135S-F2136L. Mutant W2135I-F2136C was discarded from this evaluation as it displayed the same profile as W2135S-F2136L. The parental ΔN_{123} -GBD-CD2 enzyme and its mutants displayed Michaelis-Menten kinetics for hydrolysis of the sucrose substrate (Figure S9). Interestingly, K_M , V_{max} and k_{cat} values were not consistent with those that we previously reported on ΔN_{123} -GBD-CD2 (Brison et al. 2012) which were performed at 5.4 instead of 5.1 in our case. Moreover, in this latter study, the purification of the enzyme was performed by renaturation of unfolded proteins extracted from inclusion bodies. The difference in enzyme preparation could also affect apparent kinetic parameters. A K_M value of 7.5 mM with a V_{max} of $36.3 \mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ was earlier reported at pH 5.4 for the parental ΔN_{123} -GBD-CD2. In comparison, these parameters determined at pH 5.1 revealed no variation of K_M but a 2-fold gain in V_{max} .

When comparing the kinetic parameters of ΔN_{123} -GBD-CD2 with the mutants F2163G and W2135S-F2136L, we note a ~ 3.5 -fold decrease of K_M and a 4.5 fold increase of V_{max} indicating that the mutations have affected both parameters (Table 1).

Table 1: Determination of kinetic parameters (K_M , V_{max} and k_{cat}) for sucrose hydrolysis of ΔN_{123} -GBD-CD2 parental enzyme and mutants F2163G and W2135S-F2136L. Values were obtained by the least square fit of the Michaelis-Menten equation (Table S1) and using increasing concentrations of sucrose (ranging from 10 to 300 mM), at pH 5.1 and 30°C with 0.25 U/mL of purified enzymes.

	ΔN_{123} -GBD-CD2	W2135S-F2136L	F2163G
$K_{M,suc}$ (mM)	30.47 ± 7.22	8.75 ± 1.03	7.02 ± 1.15
$V_{max,suc}$ ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	77.50 ± 6.09	17.67 ± 0.42	16.82 ± 0.45
$k_{cat,suc}$ (s^{-1})	161.7 ± 12.71	36.9 ± 0.88	35.1 ± 0.94

The impact of the mutations on the topology and dynamics of the active site for the mutants studied by MD simulations is clear (Figure 6), and this behavior could thus affect the enthalpy component of the free binding energy. The perturbation of the enzyme structure by the introduction of mutations could also affect the pKa of the catalytic residues (Ludwiczek et al. 2013). The free energy landscapes taken from the two eigenvectors from the MD simulations of the free ligand enzymes (Figure 5) and in complex with sucrose (Figures S9) revealed a different behavior of the enzymes. Except for the F2163G mutant, which showed a tightened

free energy landscape for its complex with sucrose, the effect of sucrose binding on parental enzyme and the mutants W2135S-F2136L and W2135S-F2136L is not clear. This analysis suggests that the effect of the mutation on sucrose consumption is more dominated by enthalpy variation than the loss of entropy.

Conclusion

Overall, our study shows how amino acid substitutions introduced in the acceptor subsites +1, +2 and +3 of ΔN_{123} -GBD-CD2 branching sucrose could drastically shift the preferred binding mode of a non-natural tetrasaccharide acceptor, a lightly protected precursor in the synthesis of *S. flexneri* haptens, and thereby control its regioselective glucosylation. From the screening of a focused library of 22 mutants earlier reported (Malbert et al. 2018), we identified 6 mutants exhibiting singular product profiles with respect to the parental enzyme which was earlier reported to produce mainly an uncharacterized mono-glucosylated product **P2**. By introducing a limited number of mutations, up to three, it was possible to drive the transglucosylation toward one extremity or the other of the tetrasaccharide **ABC'D'** (**A** vs **D'**), producing in different amounts four distinct pentasaccharides (**P1**, **P2**, **P2'**, and **P3**) of which two are characteristic of *S. flexneri* serotype-specific pentasaccharide repeating units (*S. flexneri* serotype 4a/4b for **P1** and prevalent serotype 3a for **P2**). Interestingly, two of these pentasaccharides (**P2'** and **P3**) and an additional di-glucosylated tetrasaccharide (**H1**) were not produced by parental ΔN_{123} -GBD-CD2.

Molecular dynamics simulations were further carried out and analyzed in details using a graph approach that enabled to map distance matrices of active site residues along the simulation. The resulting amino acid residue networks computed for parental enzyme and its mutants revealed very interesting features and specific patterns of residue interactions that helped to understand product profile specificity. Taken together, this information enabled to shed some light on the molecular and dynamical determinants responsible for the regioselectivity of the tetrasaccharide transglucosylation as well as activity toward sucrose.

Given the diversity of products obtained upon introduction of just a few mutations in the very exposed active site of ΔN_{123} -GBD-CD2, having more than 50 residues within 10 Å radius of the sucrose in its productive conformation, one can hope that purposely re-designing branching sucrose active site using computer-aided engineering methods could enable further glucodiversification and improve enzyme catalytic efficiency. Such strategies will undoubtedly provide major opportunities in the near future to develop biocatalysts needed for the development of chemo-enzymatic routes toward *Shigella flexneri* serotype-specific haptens that could enter in the composition of broad coverage vaccines.

Material & Methods

Bacterial strains, plasmids and chemicals

The same biological material and chemicals previously described (Chapter II) were used herein.

Production of the library of ΔN_{123} -GBD-CD2 in flasks

The library of 22 mutants previously constructed (Malbert et al. 2018) was here produced in 50 mL flasks according to the procedure described in Chapter II.

Sucrose activity assay of the library of ΔN_{123} -GBD-CD2 branching sucrases

After recovery of the enzymes in the soluble fraction (Chapter II), the activity on sucrose was assayed on sucrose in 1 mL format using the DNS method as previously described (Chapter II). One unit of activity is described as the amount of enzyme necessary to release one micromole of fructose in one minute at pH 5.75 (or pH 5.1 or pH 4.7 when stated) and 30°C.

Acceptor glucosylation reaction of the library of ΔN_{123} -GBD-CD2 branching sucrases

The enzymatic glucosylation assays were performed in miniaturized format at 10 μ L or 50 μ L scale in glass inserts in the optimized conditions described previously (Chapter II). The final concentrations of the acceptor **ABC'D'** and sucrose donor were 50 mM and 1 M, respectively. When the scale was 10 μ L, only 0.07 U/mL enzyme was added in order to normalize the amount on the less active enzyme, whereas when the scale was 50 μ L, 1 U/mL enzyme was added. Reactions were stopped by diluting them 5 times in a solution of H₂O/acetonitrile (70:30) containing 0.08% of TFA, allowing at the same time to stabilize the products. The reaction mixture was used for HPLC-MS analysis.

Analytical methods used for carbohydrate separation and detection

The carbohydrate content of the reaction mixtures were separated using a C₁₈RP Fusion analytical column (4 μ m, 80 Å, 250 x 4.6 mm) and detected using UV detection at 220 nm coupled with low resolution mass spectrometry detection.

Reactions performed in 10 μ L scale as well as samples collected after purification of the products were analyzed using the following method: 20 min of separation with an isocratic step of H₂O/ACN 70:30 from 0 to 4 min and a linear gradient H₂O/ACN 70:30 to 40:60 from 4 min to 11 min, followed by washing and re-equilibrations steps, at a flow of 1 mL.min⁻¹, at 40°C. Reactions performed at 50 μ L scale with ΔN_{123} -GBD-CD2 enzyme and mutants W2135I-F2136C, W2135S-F2136L and F2163G were analyzed using HPLC-UV coupled with high resolution mass spectrometry with a separation method previously described (Chapter II).

Production and purification of enzymes

ΔN_{123} -GBD-CD2, and mutants W2135S-F2136L and F2163G were produced in 1 L scale using the same method described as above but for 24 h instead of 32 h in order to avoid accumulation of aggregated or degraded forms. Purification was then performed using the same method as in Chapter II.

Synthesis pentasaccharide products P2 and P2'

Synthesis of the mixture of **P2/P2'** mono-glucosylated products was achieved in a 2 mL scale reaction using 1 U.mL⁻¹ of purified F2163G mutant with 50 mM **ABC'D'** acceptor (88 mg) and 1 M sucrose donor (685.96 mg), during 16 h of reaction. After stopping of the reaction (dilution 5 times in H₂O/ACN + 0.08% TFA) the sample was analyzed in HPLC-UV, then purified using an automated fraction collector. After re-analysis of the fractions by HPLC-UV, the ones containing pure product **P2/P2'** were pooled, acetonitrile was evaporated using a rotavap, and the samples were frozen at -80°C and lyophilized prior to re-suspension in D₂O and NMR analyses.

In a second synthesis performed in a 1.5 mL scale (66 mg acceptor and 514.47 mg sucrose), two rounds of purification were performed in order to isolate **P2'** from the mixture **P2/P2'**. The reaction was stopped by diluting the sample 2 times in H₂O/ACN (40:60) + TFA 0.16%. Modifications were applied to the elution method of the second round of purification, first an isocratic step was applied at 20% of acetonitrile during 30 min then a 10 min gradient was applied to reach 60% acetonitrile, and the column was re-equilibrated for 11 min at 20% acetonitrile. After automatic collection and re-analysis by HPLC-UV, the fractions containing pure product were evaporated to dryness using a SpeedVac, and re-suspended in D₂O prior to NMR analyses.

NMR experiments

The samples were dissolved in DCI-containing D₂O at pH 5.1. For NMR studies, the samples were lyophilized three times and dissolved in 180 μ L of 99.9% DCI-containing D₂O.

All NMR spectra were recorded on a Bruker Avance spectrometer operating at a proton frequency of 950 MHz (TGIR- RMN-THC Fr3050 CNRS, Gif-sur-Yvette) and at a carbon frequency of 238 MHz with a 5-mm gradient indirect cryoprobe. All spectra were processed and analyzed with Topspin software (Bruker).

¹H and ¹³C 1D NMR spectra were accumulated at 25 °C, 65536 data points were acquired with 32 and 2048 scans respectively for proton and carbon experiments. ¹H-¹³C HSQC (Heteronuclear Single Quantum Coherence spectroscopy), HMBC (Heteronuclear single quantum coherence spectroscopy) with J_{CH} log range of 3.5 Hz, and Double Quantum Filtered COrrrelation Spectroscopy (QDF COSY) (Rance et al. 1983) experiments were performed at 25 °C. Homo and heteronuclear spectra were recorded under

the following experimental conditions: 512 increments of 2048 complex points are acquired with an accumulation of 16 scans. Spectral widths were 16025 Hz for protons dimension and 44267 Hz for carbon dimension.

Kinetic parameter determination

Kinetic parameters on sucrose were determined using 0.25 U/mL of purified enzymes at pH 5.1 and 30°C. Sucrose was used in increasing concentrations ranging from 0 to 300 mM in order to determine catalytic efficiency. Reactions were stopped by heating the samples at 95°C for 2 minutes and analyzed by HPLC-CAD on HPX-87C columns. When sucrose is used as sole substrate, the enzymes adopted a Michaelis-Menten behavior (Johnson and Goody 2011). The k_{cat} , V_m and K_m values were determined and catalytic efficiency was defined as the ratio k_{cat}/K_m .

Computational procedures

MD simulations were performed with the AMBER ff14SB (Cornell et al. 1995) force-field for enzymes and GLYCAM_06j-1 (Kirschner et al. 2008) for sugar ligands. Carbohydrate oligomers were constructed using tleap program from AMBER software package (Pearlman et al. 1995). The chemical protecting groups; chloroacetyl and allyl carried by chloroacetyl α -L-rhamnopyranoside (C') and allyl *N*-trichloroacetyl- β -D-glucosaminopyranoside (D') units respectively were constructed by Avogadro software (Hanwell et al. 2012), the trichloroacetyl of D' was obtained through substitution of acetyl hydrogens by chlorine atoms. Quantum mechanical calculations were performed using the Gaussian software package. Geometries were optimized using the Gaussian default optimization criteria. The HF/6-31G* level of theory was employed. Resulting partial charges were fitted using restrained electrostatic potential (RESP) (Bayly et al. 1993) with 0.01 charge restraint weight (Woods and Chappelle 2000). The valence bond, angle, and torsional parameters of protecting groups were derived from gaff force field. Nevertheless, original GLYCAM geometrical parameters of sugar rings were conserved.

The 3D model for the parental enzyme was based on ΔN_{123} -GBD-CD2 PDB ID: 3TTQ (Brison et al. 2012). Mutants F2163G, W2135S-F2136L and W2135I-F2136C were constructed using Rosetta3 software (Richter et al. 2011). The X-ray conformations were conserved for all amino acids except for the residues targeted by mutation for which the side chain rotations were permitted. The H++ webserver was used to determine the protonation state of ionisable residues at pH 5.75 at which the activity of ΔN_{123} -GBD-CD2 and its mutants was experimentally tested. The enzyme:sucrose complexes were built by superposing ΔN_{123} -GBD-CD2 and its variants to the inactive mutant of GTF180, PDB ID: 3HZ3 containing sucrose in the active site (Vujcic-Zagar et al. 2010). Furthermore, this sucrose binding structural knowledge was used for docking pentasaccharides; ABC[E(1→6)]D' and [E(1→3)]ABC'D' (**P1** and **P2**, respectively)

in the active site. The strategy consisted on taking the coordinates of the glucosyl moiety from sucrose for the glucosyl of the pentasaccharides. Then, the simulated annealing (from 0 to 350 K in 10 ps and vice versa) of the systems in vacuum with harmonic positional restraints of 50.0 kcal/mol/Å² on the enzyme and glycosyl unit, and the sugar pucker rings of **ABC'D'** was carried out for each system.

MD simulation were performed using the NAMD program (Phillips et al. 2005) for the Apo forms. The complex systems MD simulations were done by pmemd.CUDA (Salomon-Ferrer et al. 2013) supporting the mixed scaling of 1-4 non-bond electrostatic and van der Waals interactions (Kirschner and Woods 2001) of amino acids and sugars.

MD simulations were performed at constant temperature (303 K) and pressure (1 bar) using the Berendsen algorithm (Berendsen et al. 1984). The integration time-step was 2 fs and covalent bonds involving hydrogen were constrained using SHAKE (van Gunsteren and Berendsen 1977). The non-bonded pair-list was updated heuristically. Long-range electrostatic interactions were treated using the particle mesh Ewald (PME) approach (Darden et al. 1993). Non-bonded interactions were treated with a 9 Å direct space cut-off. All enzyme systems were neutralized with Na⁺ ions (Åqvist 1990) (minimal salt condition), in explicit TIP3P water molecules (Jorgensen et al. 1983); the primary boxes were rectangular with solvent extending 10 Å around the enzymes. The water molecules and counterions were energy-minimized and equilibrated at 100 K around the constrained solute for 100 ps in the NVT ensemble; the entire system was then heated incrementally over 100 ps from 100 to 300 K in 5 K steps with harmonic positional restraints of 25.0 kcal/mol/Å² on the solute atoms. The MD simulations were continued in NPT, without notable change in volume. The positional restraints were gradually removed over 250 ps and followed by the production phase. MD snapshots were saved every 1 ps. The lengths of MD simulations were 1 μs for the free ligand forms, 100 ns for enzyme:sucrose complexes and 2 ns for enzyme:pentasaccharides complexes.

Principal Component Analyses (PCA), Root Mean Square Deviation (RMSD), B-factors, Free-energy landscape (FEL) analysis were carried out as described in our previous study of ΔN₁₂₃-GBD-CD2 branching sucrose (Ben Imeddourene et al. 2018). The active site residues average distance matrices from MD simulation were built using cpptraj (Roe and Cheatham 2013). The latter distance matrices were employed for the graph analyses based on the Fruchterman-Reingold algorithm (Fruchterman and Reingold 1991) and the spin-glass algorithm (Epskamp et al. 2012) implemented in R software (R Development Core Team 2018) by Borsboom group

Acknowledgements

This work was funded by the French National Research Agency (ANR Project CARBUNIVAX, *ANR-15-CE07-0019-01*). Financial support from the TGIR- RMN-THC Fr3050 CNRS (Gif-sur-Yvette) for the NMR experiments is gratefully acknowledged as well as N. Morellet for helpful advice on NMR experiments. The authors are grateful to P. Bondy for providing her help in the production of mutants. Authors thank N. Monties and G. Cioci from the PICT-ICEO screening facility from Toulouse Biotechnology Institute (TBI) for providing access to analytical and protein purification facilities. The authors also thank F. Bellvert and H. Barbier for granting access to HRMS facility of the MetaToul platform of TBI (Toulouse, France), and performing the product analysis. This work was granted access to the HPC resources on the TGCC-Curie, CINES-Occigen supercomputers and the Computing mesocenter of Région Midi-Pyrénées (CALMIP, Toulouse, France).

Supplementary Information

Supporting Figures

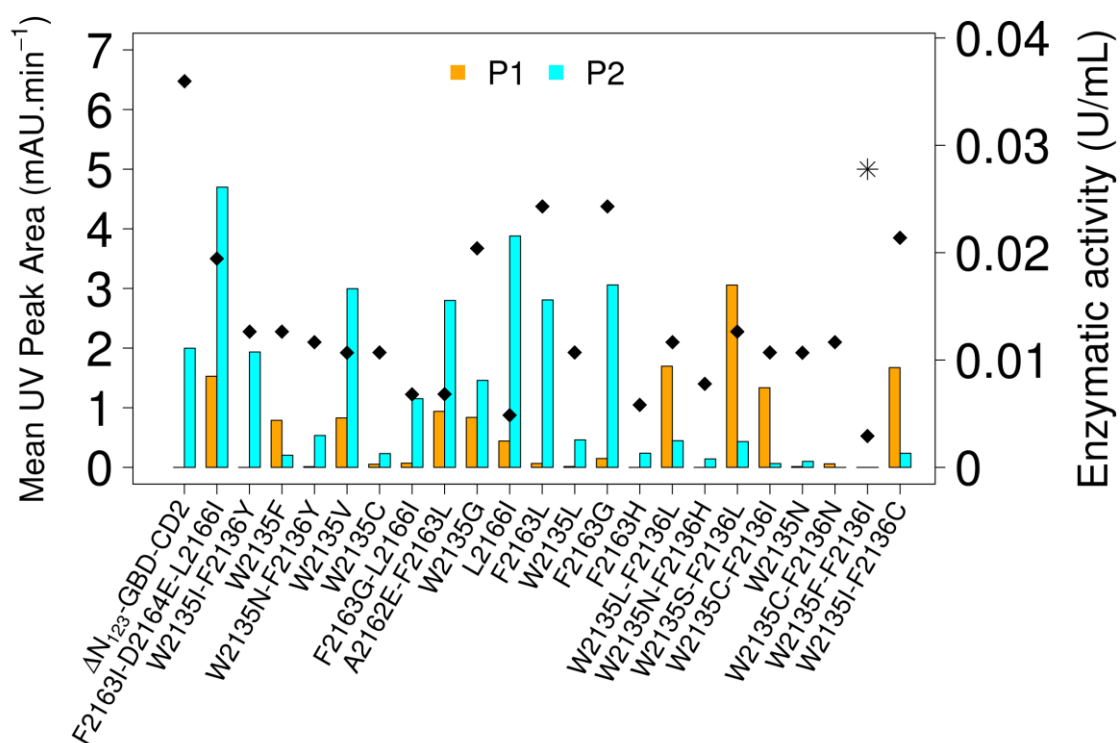


Figure S1: Products formed by ΔN_{123} -GBD-CD2 branching sucrose and its mutants in presence of 1 M sucrose and 50 mM **ABC'D'**, after 16 h reaction at pH 4.7 and 30°C, using between 0.003 and 0.037 U/mL enzymatic activity (black diamonds) and detected by mass spectrometry coupled with HPLC-UV (wavelength at 220 nm). The two main pentasaccharide products are **P1** (orange) and **P2** (cyan). * represents the mutant W2135F-F2136I that was not able to glucosylate tetrasaccharide **ABC'D'** in these conditions.

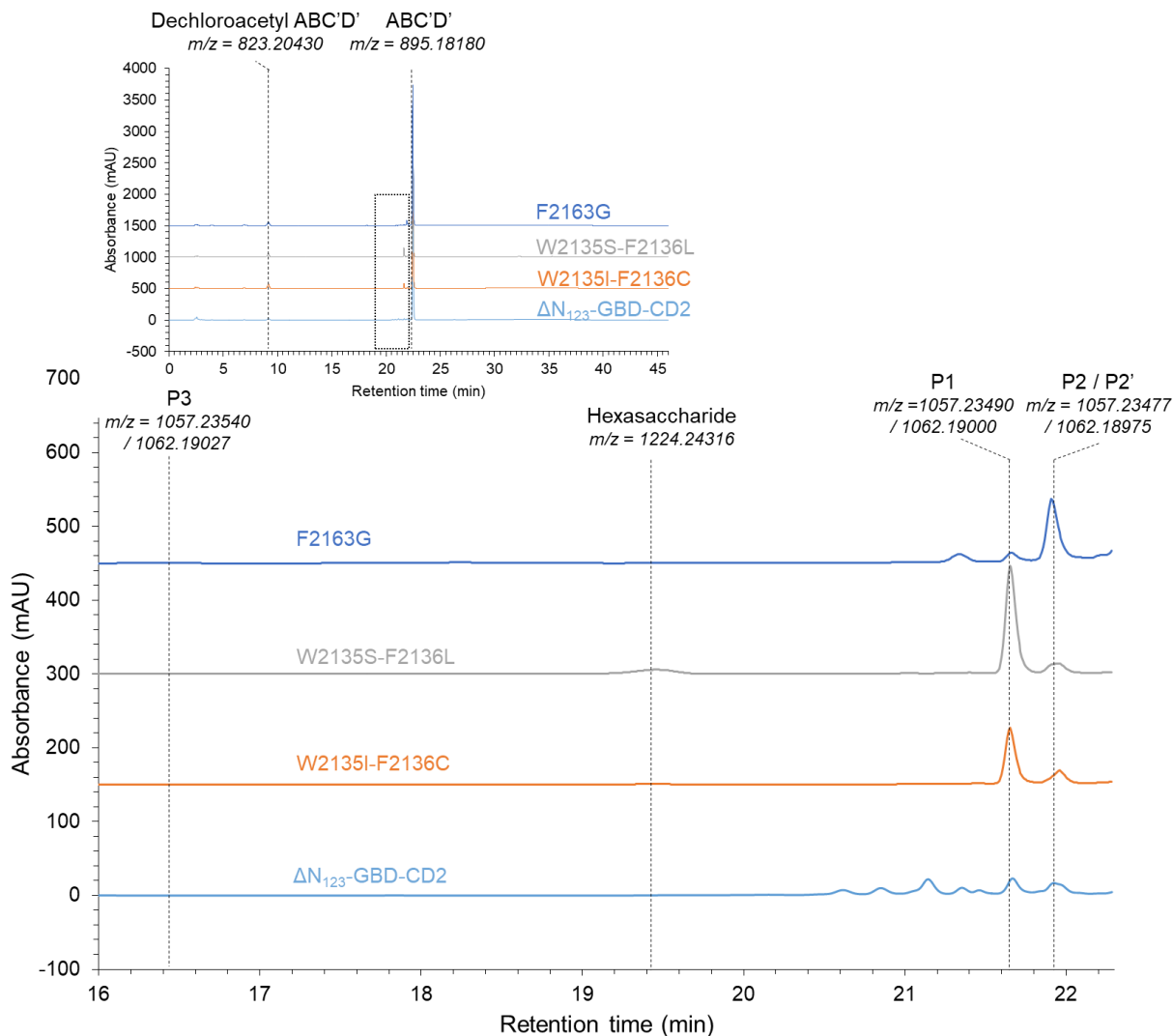


Figure S2: Comparison of product profiles obtained with ΔN_{123} -GBD-CD2 branching sucrose and its mutants W2135I-F2136C, W2135S-F2136L and F2163G after 16 h of reaction in presence of sucrose (1 M) and tetrasaccharide **ABC'D'** (50 mM) at pH 5.1 using 1 U/mL enzyme and 30°C. HPLC-UV_{220nm} analysis was performed. A zoom of the product region located between 16 and 22.3 minutes is shown. **P1** ($t_R = 21.65$ min), **P2** ($t_R = 21.9$ min), **P3** ($t_R = 16.4$ min) and **H1** ($t_R = 19.45$ min): products of the enzymatic glucosylation. A shift in molecular mass by 162 Da, characteristic of the glucosylation, was observed for **P1**, **P2** and **P3** and a shift of 362 Da was observed for the **H1** (m/z detected by HRMS indicated below each product and corresponding to Na^+ or NH_4^+ adducts in positive mode). Other **poly-glycosylated** products detected for parental ΔN_{123} -GBD-CD2 branching sucrose are listed in Chapter II.

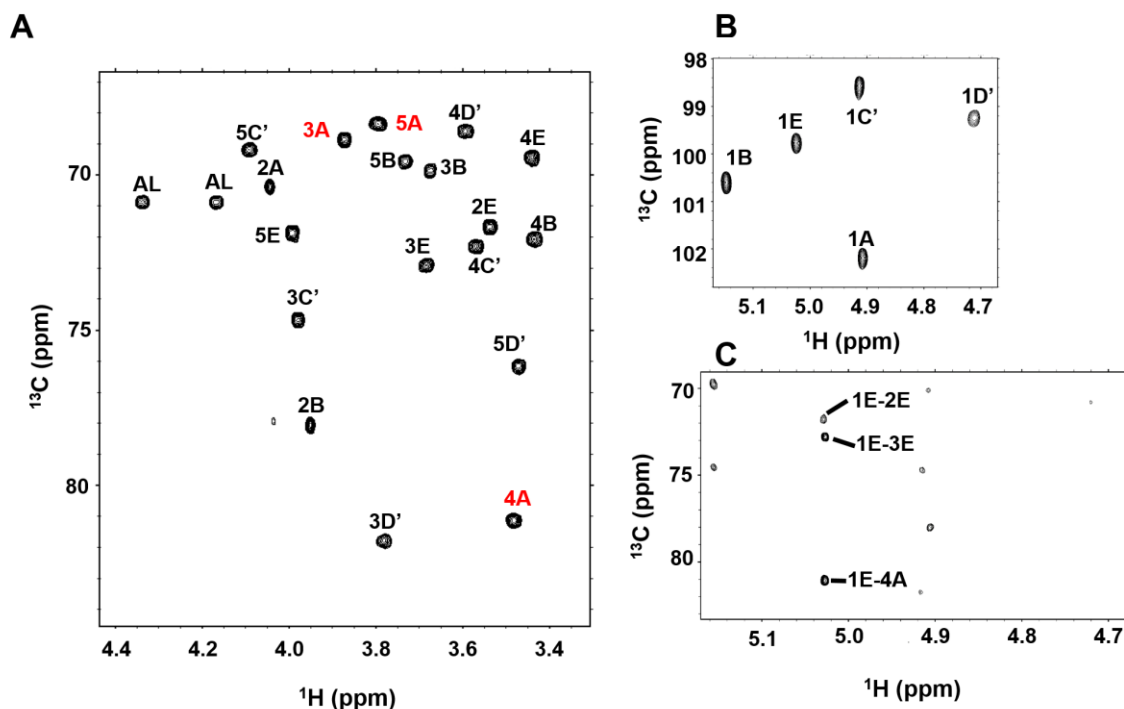


Figure S3: ^1H - ^{13}C HSQC spectra of the C2, C3, C4 and C5 resonances (A) and the anomeric region (B) of **P2'** pentasaccharide. The ^1H - ^{13}C HMBC illustrates the correlation between the E and A units (C). The shifted resonances, compared to the tetrasaccharide, were labeled in red color. All spectra were acquired at 950MHz.

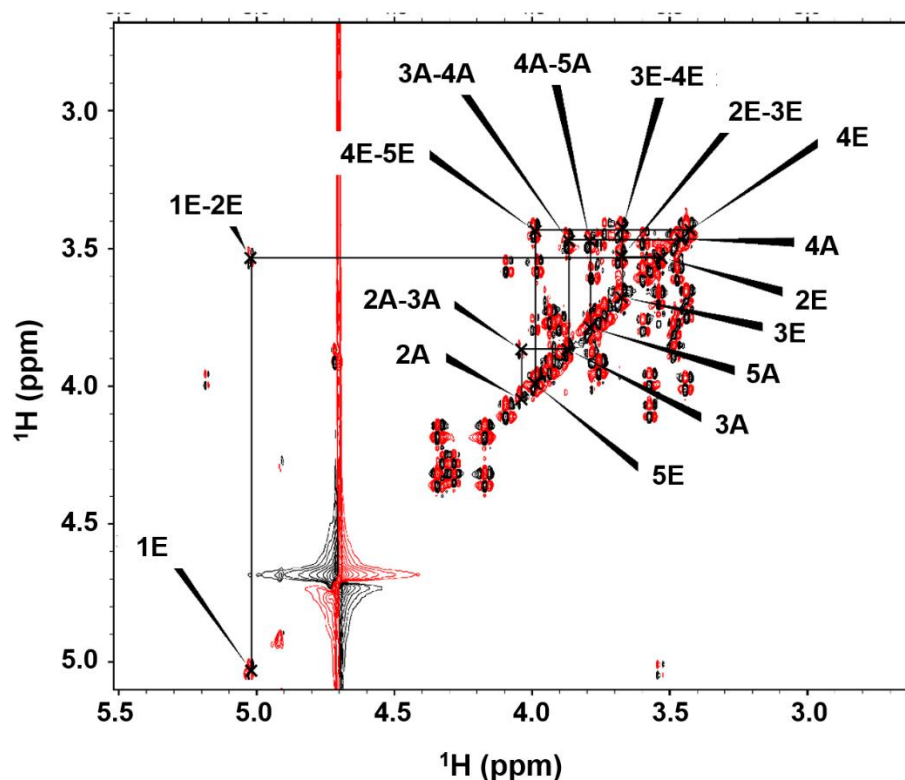


Figure S4: ^1H - ^1H QDF-COSY spectrum at 950MHz of the **P2'** pentasaccharide. Positive and negative peaks are showed in black and red respectively. The E and A units connectivities are plotted and the corresponding peaks are labeled.

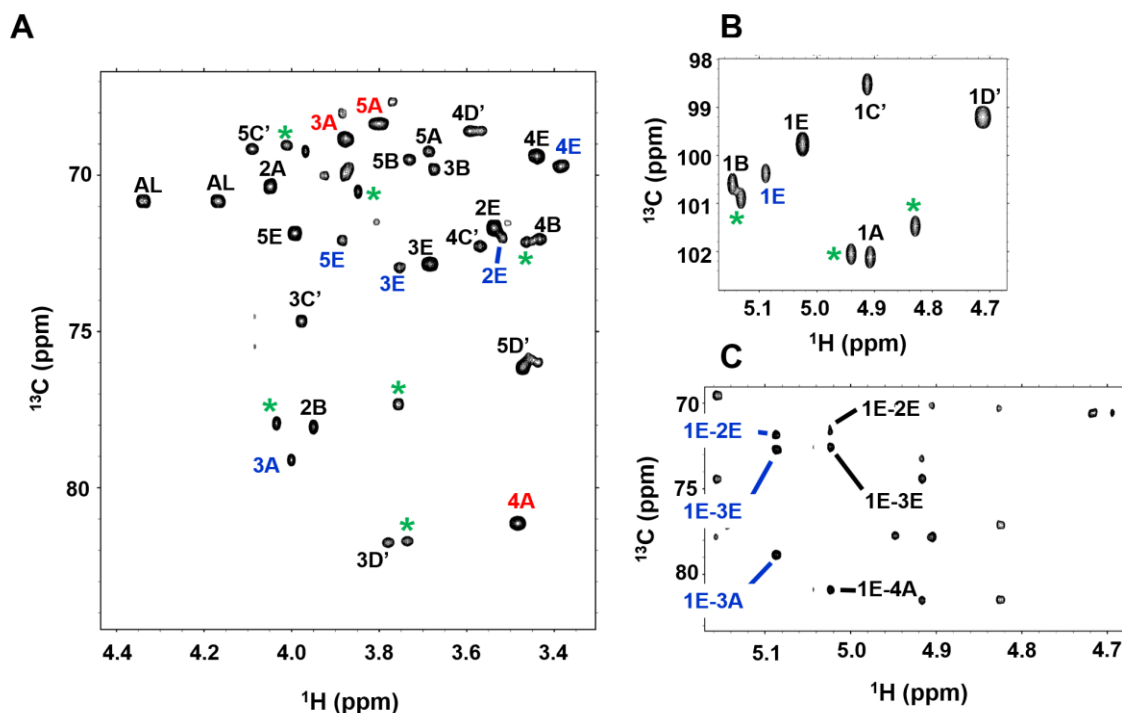


Figure S5: ^1H - ^{13}C HSQC spectra of the C2, C3, C4 and C5 resonances (A) and the anomeric region (B) of the mixture. The ^1H - ^{13}C HMBC illustrates the correlation between the E and A units of P2 and P2' (C). The shifted resonances of the P2', compared to the tetrasaccharide, were labeled in red color and the A and E units belonging to the P2 are highlighted by blue color. The peaks of the dechloroacetylated form ABCD' are labelled by green stars. All spectra were acquired at 950MHz.

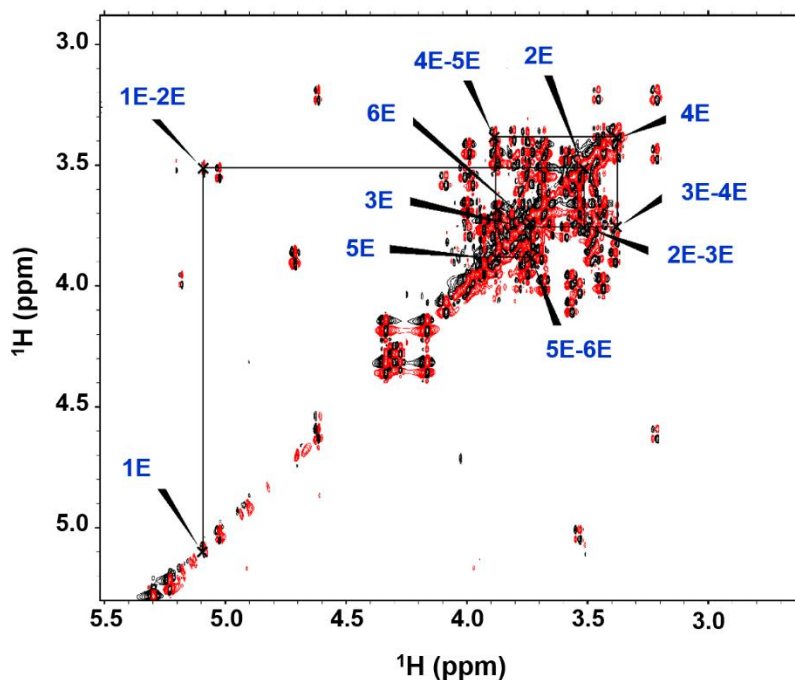


Figure S6: ^1H - ^1H QDF-COSY spectrum at 950MHz of the mixture. Positive and negative peaks are shown in black and red respectively. The E units belonging to the P2 connectivities are plotted and the corresponding peaks are labeled in blue color.

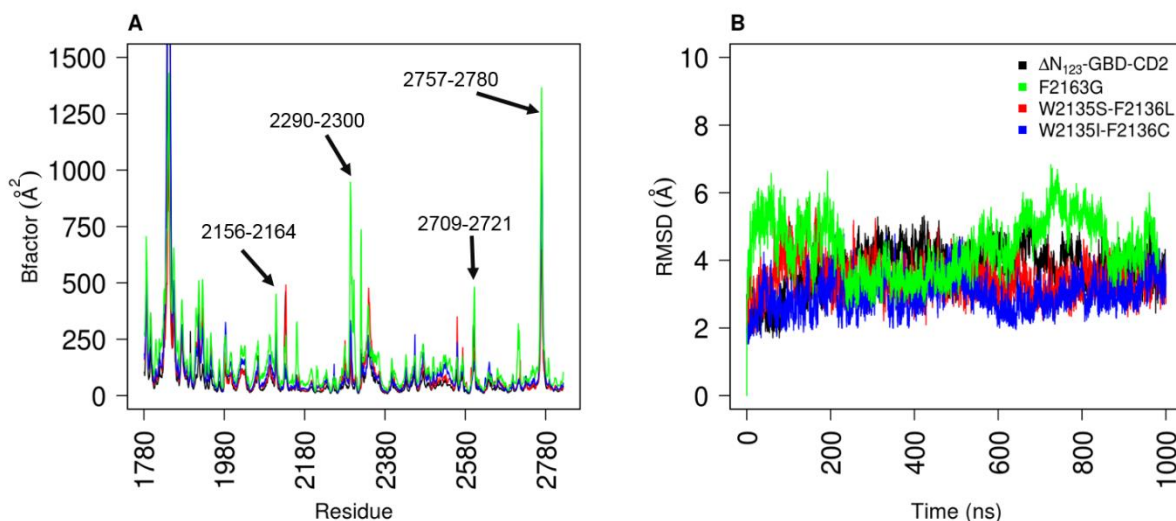


Figure S7: Analysis of MD simulations. **(A):** B-factors were calculated as function of enzyme amino acid residues along 1 μ s of MD simulation of the parental ΔN_{123} -GBD-CD2 (black), the mono mutant F2163G (green), and double mutants W2135I-F2136C (blue) and W2135S-F2136L (red). The regions pointed by the arrows are those discussed in the text. **(B)** Carbon α atoms Root Mean Square Deviation of ΔN_{123} -GBD-CD2 (black line), W2135I-F2136C (red line), W2135S-F2136L (blue line) and F2163G (green line), with respect to parental ΔN_{123} -GBD-CD2 X-ray structure (PDB ID: 3TTQ) were calculated as a function of simulation time

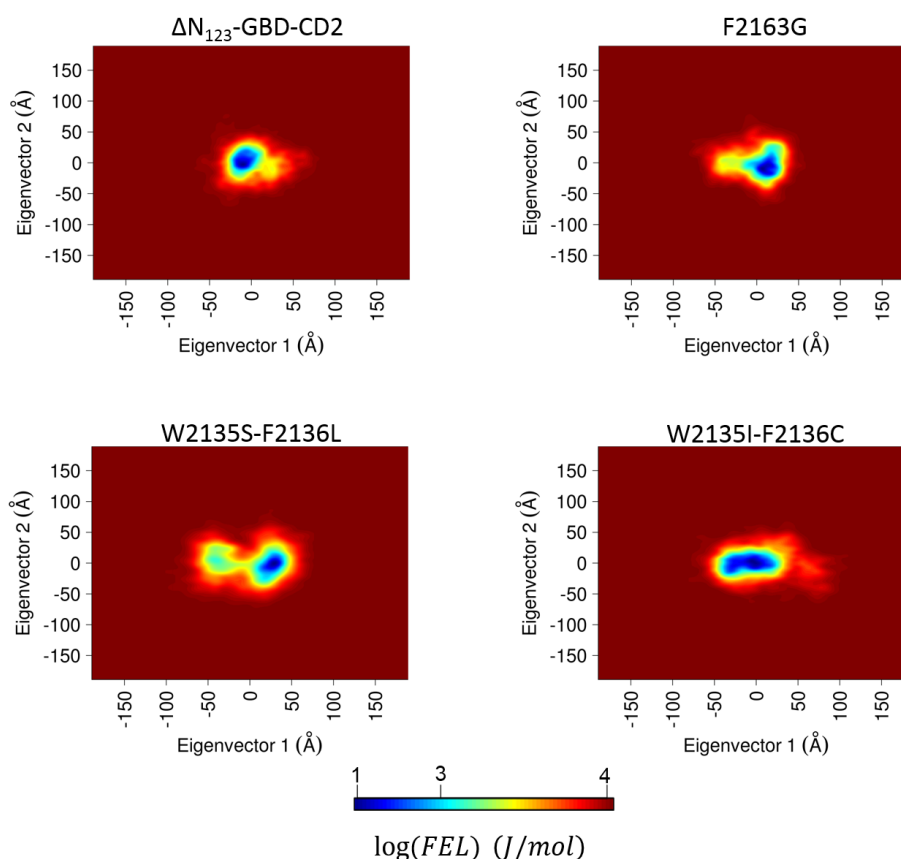


Figure S8: Free energy landscape (FEL) of enzyme:sucrose complexes of parental ΔN_{123} -GBD-CD2, F2163G, W2135S-F2136L and W2135I-F2136C were determined using as reaction coordinates the projection of the first and second principal components from 100ns MD simulation of the enzyme:sucrose complexes for each system. The bottom legend shows the color scale of the logarithm of FEL in J. mol⁻¹.

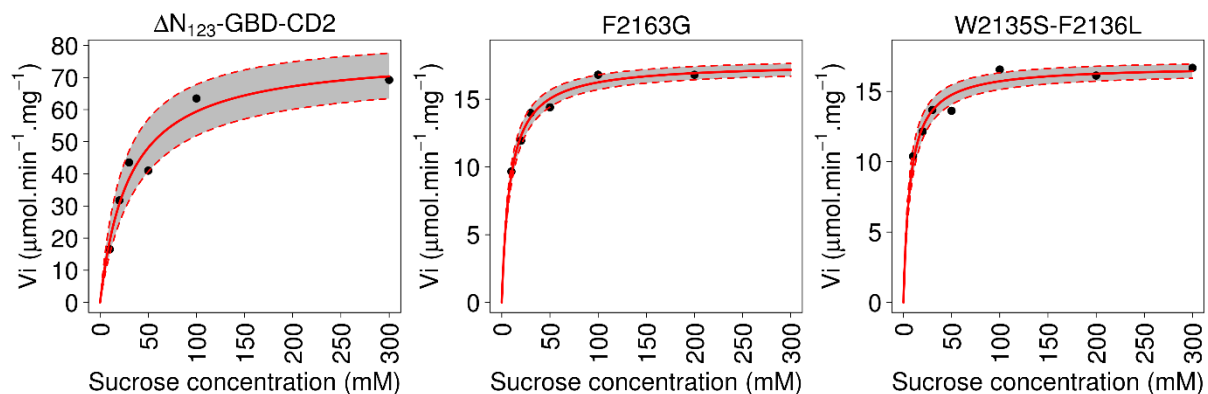


Figure S9: Initial velocities of sucrose hydrolysis activity of parental ΔN_{123} -GBD-CD2, and mutants F2163G and W2135S-F2136L. Determination of kinetic parameters was performed using increasing concentrations of sucrose (ranging from 10 to 300 mM), at pH 5.1 and 30°C with 0.25 U/mL of purified enzymes. The curves obtained by the least square fit are represented by the red continuous line and residual standard deviations are depicted by dashed red lines and the area between them is highlighted by gray color.

Supporting Tables

Table S1: The parameters extracted from the fitted Michaelis-Menten equation of N number experimental data having R^2 correlation coefficient with fit P-value. The K_m and V_{max} are given with them standard deviation. The fisher values versus the hypothesis of linear fits calculated were significantly higher than the F-table values for $v_1=1$ and $v_2=4$ or 3 from F-table provided between brackets for $\alpha=0.05$

Enzyme	N	R^2	K_m	V_{max}	P-value	F-stat
ΔN_{123} -GBD-CD2	6	0.95	30.47 ± 7.22	77.50 ± 6.09	0.001	70.95 (7.71)
F2163G	5	0.97	8.75 ± 1.03	17.67 ± 0.42	0.0002	151.92 (10.13)
W2135S-F2136L	6	0.93	7.02 ± 1.15	16.82 ± 0.45	0.0004	67.75 (7.71)

Complementary Results

This complementary part presents additional results about the biophysical and biochemical characterization of ΔN_{123} -GBD-CD2 and its mutants. Stability studies were performed, in particular the difference in melting temperature between the mutants and the parental enzyme was determined. Stabilizing effect of sucrose was demonstrated using the inactive mutant E2248Q. Finally, branching reactions using the natural acceptor of these enzymes, dextran, were performed and the branching ratio was calculated and compared to that of the parental enzyme.

Biophysical characterization: determination of the melting temperatures by Differential Scanning Fluorimetry

Stability of the enzymes ΔN_{123} -GBD-CD2, ΔN_{123} -GBD-CD2 W2135S-F2136L and ΔN_{123} -GBD-CD2 F2163G as well as the inactive mutant ΔN_{123} -GBD-CD2 E2248Q mutated on the catalytic acid/base residue were assessed by determining their melting temperature (T_m) in differential scanning fluorimetry (DSF) experiments. 15 μ L enzyme (final concentration of 5 μ mol.L⁻¹) was mixed with 5 μ L of SYPRO Orange dye 40X. Melting curves were collected using a real-time qPCR apparatus. The thermal shift was evaluated from melting curves. Effect of sucrose on the stability of the enzyme was assessed using an increasing concentration of 1, 10 and 100 mM sucrose with the inactive mutant ΔN_{123} -GBD-CD2 E2248Q.

The mutants F2163G and W2135S-F2136L had a first T_{m1} of 32.9°C and 34.7°C, respectively, while the parental enzyme and inactive mutant had a T_{m1} of 35.7°C and 35.9°C, respectively. They also both displayed a second T_{m2} of 42.2°C (Figure C1 and Table C1). The loss of the second melting temperature by the mutants could be explained by a different unfolding dynamic, where parental enzyme and inactive mutant could be partially folded up to 42°C, which could be the case for example if there was an upkeep of one particular domain of the GH70.

It should be noted that the parental enzyme quickly lost its activity at 37°C (data not shown), this loss of activity could be explained by the dynamical population shift of the loops surrounding the active site or by the unfolding of the domain A.

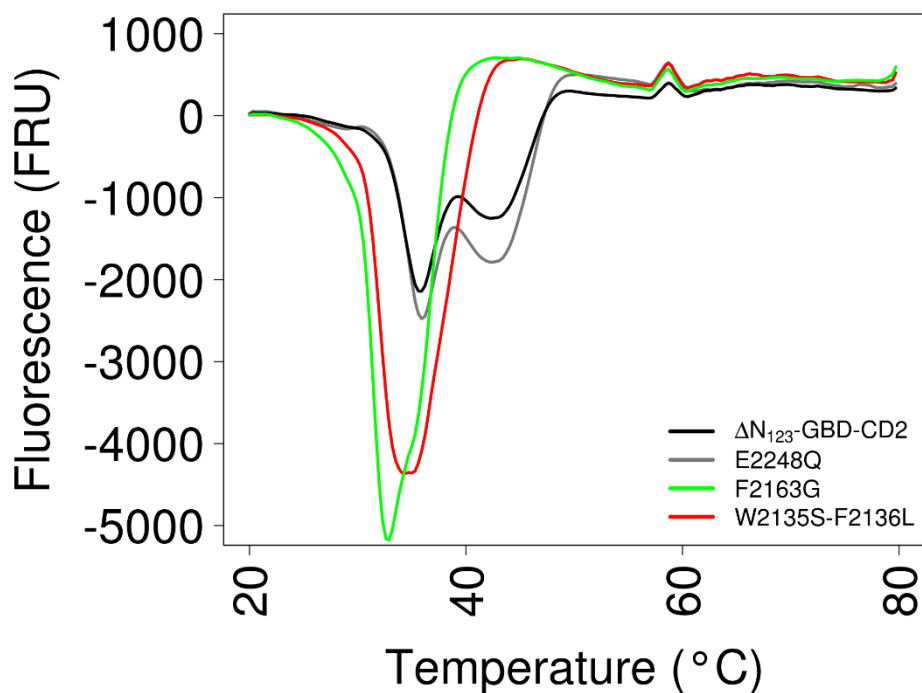


Figure C1: Melting temperature curves of ΔN_{123} -GBD-CD2, inactive mutant E2248Q and mutants W2135S-F2136L, F2163G determined by differential scan fluorimetry

Table C1: Melting temperature interval values of ΔN_{123} -GBD-CD2, inactive mutant E2248Q and mutants, F2163G, W2135S-F2136L determined by differential scan fluorimetry experiments repeated three times.

Enzyme	T_{m1} (°C)	T_{m2} (°C)
ΔN_{123} -GBD-CD2	35.7	42.2
ΔN_{123} -GBD-CD2 E2248Q	35.9	42.2
ΔN_{123} -GBD-CD2 F2163G	32.9	
ΔN_{123} -GBD-CD2 W2135S-F2136L	34.7	

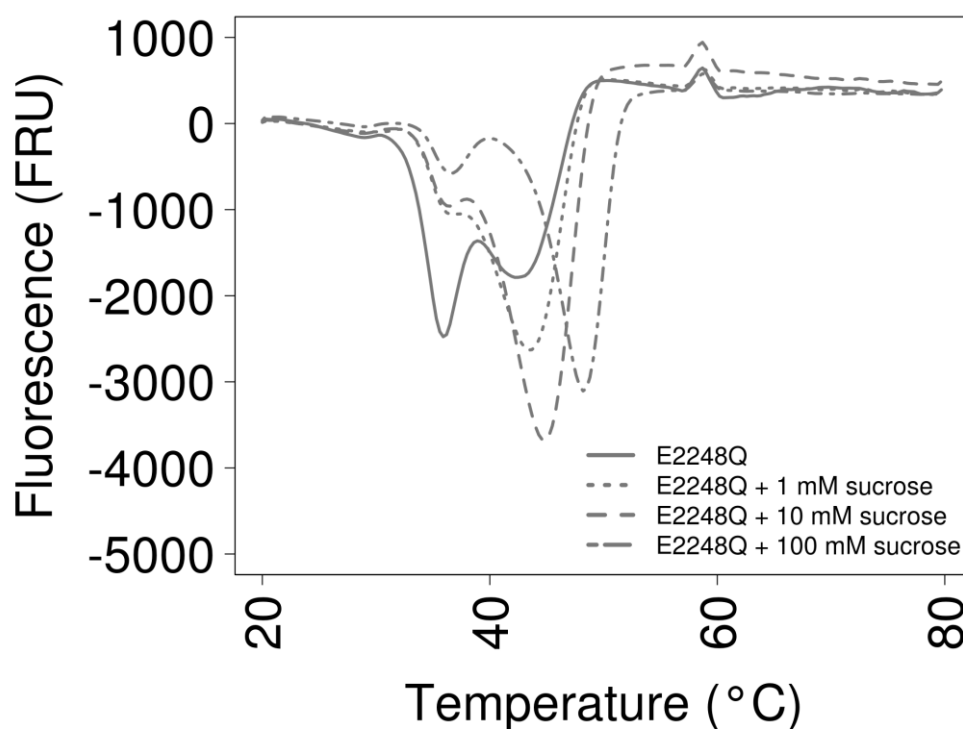


Figure C2: Melting temperature curves of ΔN_{123} -GBD-CD2 inactive mutant E2248Q determined in presence of increasing concentrations of sucrose by differential scan fluorimetry

The stabilizing effect of sucrose was determined using the inactive mutant E2248Q (Figure C2) as it is unable to hydrolyze sucrose. With increasing concentrations of sucrose (1, 10, 100 mM) the second T_{m2} was increased (+1.3°C, +2.5°C, +6°C, respectively). Surprisingly, the first T_{m1} was not affected by the increase of sucrose concentration. This stabilizing effect could be explained by the sucrose binding to the glucan binding domain V or other domains.

Table C2: Melting temperature interval values of ΔN_{123} -GBD-CD2 inactive mutant E2248Q determined in presence of increasing concentrations of sucrose (1, 10 and 100 mM) by differential scan fluorimetry experiments repeated three times.

Enzyme	T_{m1} (°C)	T_{m2} (°C)
ΔN_{123} -GBD-CD2 E2248Q (<i>inactive</i>)	35.9	42.2
ΔN_{123} -GBD-CD2 E2248Q + 1 mM sucrose	36.8	43.5
ΔN_{123} -GBD-CD2 E2248Q + 10 mM sucrose	36.5	44.7
ΔN_{123} -GBD-CD2 E2248Q + 100 mM sucrose	36.5	48.2

Biochemical characterization: determination of the branching ratio and product profiles using dextran as an acceptor

Branched dextran were synthesized using 1 U/mL of purified enzymes (ΔN_{123} -GBD-CD2, ΔN_{123} -GBD-CD2 F2163G and ΔN_{123} -GBD-CD2 W2135S-F2136L), 100 g/L sucrose and 50 g/L dextran either 1,500 Da, 70,000 Da or 2,000,000 Da, at 30°C, pH 5.1 during 24 h. Reactions were stopped by heating at 95°C for 5 min. Sucrose consumption was controlled by HPLC-CAD (Corona) on HPX-87C for branching reactions with dextran 70,000 Da and 2,000,000 Da. These samples were then dialyzed 3 times against 5000 volumes of dH₂O (14,000 Da cutoff), centrifuged, frozen at -80°C and lyophilized prior to re-suspension in D₂O for ¹H NMR analyses. The product profile and sucrose consumption of the reaction corresponding to synthesis of branched dextran using dextran 1,500 Da as an acceptor were determined by HPAEC-PAD on a Dionex Carbo-Pack PA100 column.

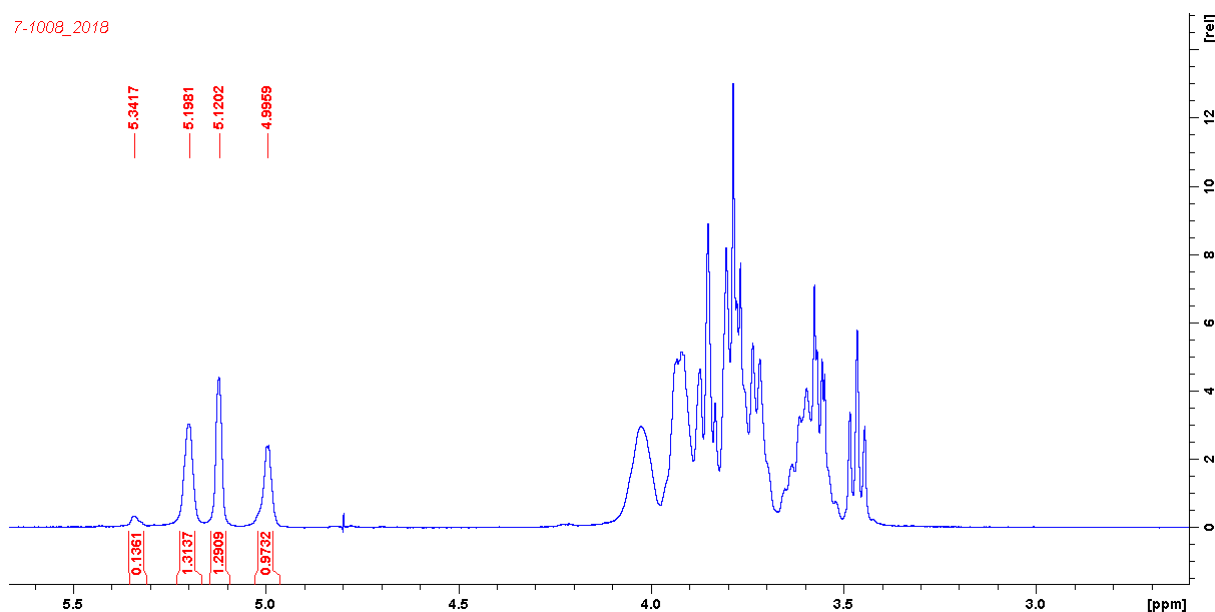


Figure C3: ¹H NMR spectrum of the α -1,2 branched dextran (2,000,000 Da) after a 24 h reaction using parental enzyme ΔN_{123} -GBD-CD2, 50 g.L⁻¹ of dextran and 100 g.L⁻¹ of sucrose. Peaks corresponding to anomeric protons signals of α -(1,3), α -(1,2) (first peak), α -(1,2) (second peak), and α -(1,6) linkages are at 5.34, 5.19, 5.12 and 4.99 ppm, respectively.

The branching was affected by the double mutation W2135S-F2136L, decreasing from 34.6% for the parental enzyme to 10.8% on dextran 70,000 Da. The same decrease was observed with longer dextran chains of 2,000,000 Da. This indicates that the mutations had a negative impact of the positioning of the dextran into the active site. As previously shown, these mutations allowed more flexibility of the loops surrounding the active site, which could cause closure of the entry point into the active site of dextran chains.

Interestingly, the product profile after branching of short dextran chains determined by HPAEC-PAD (data not shown) and compared to a standard of dextran 1,500 Da was strictly similar between the parental enzyme and the mutant F2163G, consistent with the fact that they have

the same branching ratio on long dextran chains.; while mutant W2135S-F2136L and the parental enzyme which have notable difference in the branching ratio with longer dextran chains also have a different product profile on short dextran chains (data not shown). This confirms that the two mutations W2135S and F2136L while being located in the active site must have an impact in the positioning of the dextran chains. The bulky W2135 and F2136 were probably stacking the growing dextran chain and the loss of this constraint may have impacted the good positioning of the glucose in the subsite +1. Modulation of the branching ratio by simply mutating a few key amino acids in the active site is of particular interest as dextrans presenting more or less branched chains may have different biological effect.

Table C3: α 1,2 branching ratio (in %) on dextran 70,000 and dextran 2,000,000 with the enzymes Δ N₁₂₃-GBD-CD2 and mutants F2163G, W2135S-F2136L. Branching ratio were calculated as follows:

Branching ratio (%) = $100 * \frac{\frac{\alpha_{1,2(1)} + \alpha_{1,2(2)}}{2}}{\alpha_{1,6} + \alpha_{1,2(1)} + \alpha_{1,2(2)} + \alpha_{1,3}}$, with the relative intensity values of anomeric protons signals integrated on the ¹H NMR spectra.

Enzyme	α 1,2 linkages on dextran 70,000 (%)	α 1,2 linkages on dextran 2,000,000 (%)
Δ N ₁₂₃ -GBD-CD2	34.6	35.1
Δ N ₁₂₃ -GBD-CD2 F2163G	32.3	33.4
Δ N ₁₂₃ -GBD-CD2 W2135S-F2136L	10.8	10.7

CHAPTER IV

Summary: *This chapter presents the computer-aided design of BRS-B $\Delta 2$ branching sucrose toward the regio-selective glucosylation of the target tetrasaccharide molecule. A set of 49 branching sucroses containing up to 15 mutations was evaluated, giving access to novel molecular diversity.*

Main achievements: *Computational enzyme design framework was implemented to adapt BRS-B $\Delta 2$ active site to the site-selective glucosylation of the target acceptor. Out of 49 designed mutants, 12 were found able to glucosylate the tetrasaccharide, giving access to 6 out of the 8 possible mono-glucosylated products whose structures were elucidated either by NMR spectroscopy or sensitive MS/MS techniques. In particular, a novel molecule **P3'** representative of S. flexneri 5a serotype was identified*

Contributions to the work: *Tetrasaccharide was chemically synthesized by L-A. Barel at Pasteur Institute.*

Computer-aided re-design of the enzyme was performed by A. Ben Imeddougne. I was responsible for the production of the mutant enzymes and optimization of their soluble expression conditions, screening of their activity, scale-up of the reaction, synthesis and purification of the products of interest.

The products were characterized by NMR by A. Ben Imeddougne. High resolution mass spectrometry was performed by H. Barbier. MS/MS studies were performed by D. Ropartz, M. Fanuel and H. Rogniaux from INRA Nantes.

Computer-aided engineering of a branching sucrase for the glucodiversification of a tetrasaccharide precursor of *S. flexneri* antigenic oligosaccharides

Introduction

In recent years, enzyme engineering has enabled unprecedented expansion of the enzyme repertoire, notably endowed with desirable physical and catalytic properties, and broadened the horizons of enzyme-based process development (Fuchs et al. 2015; France et al. 2017; Green and Turner 2016). In particular, progresses in the development of bioinformatics tools and computational methods have considerably contributed to better understand natural evolution of enzymes, rationalize mutations to acquire a given function or property and *in fine* accelerate conception of tailored catalysts for synthetic purposes. The strength of computational protein design has been demonstrated for engineering enzymes with improved stability (Malakauskas and Mayo 1998; Joo et al. 2010), enhanced catalytic performances (Ashworth et al. 2006), altered substrate selectivity and specificity (Murphy et al. 2009), or even able to catalyze new-to-Nature reactions (Jiang et al. 2008; Reeve et al. 2015; Röthlisberger et al. 2008; Siegel et al. 2010). Importantly, these computer-aided engineering methods enable to fine-tune the active site toward recognition and conversion of exogenous substrates by few specific amino acid mutations (Chen et al. 2009; Mak et al. 2015; Reeve et al. 2015; Siegel et al. 2015; Vergès et al. 2015), thereby avoiding the extensive screening of libraries required in directed evolution approaches. This is particularly advantageous when the amount of available substrate is limited, in particular when targeting rare, structurally complex or synthetic molecules, or when no high-throughput assay system is available. For many years, our research group has applied such rational and semi-rational engineering approaches to engineer sucrose-active α -transglucosylases from Glycoside Hydrolase families 13 and 70 in order to produce a variety of glyco-conjugates, carbohydrate derivatives, including chemically modified oligosaccharide precursors (Champion et al. 2009b, 2012; Vergès et al. 2015; Malbert et al. 2018). Complexity of carbohydrate structures renders particularly challenging differentiation and characterization of products obtained from catalytic reaction. Therefore, low screening throughput methods based on chromatography, mass spectrometry, and NMR spectroscopy are usually used.

In the continuation of our previous efforts (Champion et al. 2012; Salamone et al. 2015a; Vergès et al. 2015) to fashion the polysaccharide principal components of the *Shigella* membrane surface (Perepelov et al. 2012; Shashkov et al. 2013), the work disclosed herein aims at developing novel chemo-enzymatic routes to access well-defined O-Ag fragments that could enter in the composition of broad coverage vaccines (Chapter II). *Shigella* are gram negative bacteria responsible for shigellosis, a bacillary dysentery that represents a major burden in low and middle income countries (Livio et al. 2014), and for which there is no vaccine

yet. Among the prevalent isolates found in *Shigella* infection are found *S. flexneri* serotypes which present a tremendous O-Ag structural diversity. While most *S. flexneri* O-Ags share the same linear backbone composed of 1,2-*trans* linked L-rhamnosyl (A, B, C) and N-acetyl-D-glucosamine (D) residues, the repeating units of the different serotype-specific O-Ags differ by their α -D-glucosylation (E) pattern, as well as by their O-acetyl and phosphoethanolamine substitutions. One main synthetic roadblock to access the *Shigella flexneri* serotype-specific haptens resides in the regio- and stereo-specific control of the 1,2-*cis* glucosylation step.

With the aim of proposing novel and easy-to-implement synthetic alternatives, we came up with the idea of developing a highly convergent chemo-enzymatic strategy that will target the enzymatic regio-selective glucosylation of a lightly protected **ABC'D'** tetrasaccharide (Allyl α -L-rhamnopyranosyl-(1 \rightarrow 2)- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-O-chloroacetyl- α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-deoxy-2-trichloroacetamido- β -D-glucopyranoside) resembling the common backbone repeating unit of most *S. flexneri* O-Ags (Chapter II). Using native branching sucrases from GH70 family, we previously reported successful enzymatic glucosylation of **ABC'D'** for the production of a pentasaccharide **P1** characterized to be ABC'[E(1 \rightarrow 6)]D', characteristic of serotypes 4a/4b (Chapter II). Among the six tested branching sucrases, the best conversion of **ABC'D'** (31.3%) was obtained with BRS-B Δ 1. Using a pre-existing library of mutants of another branching sucrose, namely Δ N₁₂₃-GBD-CD2, we later obtained three additional pentasaccharides (named **P2**, **P2'** and **P3**), of which **P2** and **P2'** were characterized to be ([E(1 \rightarrow 3)]ABC'D' and [E(1 \rightarrow 4)]ABC'D', respectively (Chapter III). Of note, **P2** was found to be representative of prevalent *S. flexneri* serotype 3a.

In order to further extend the diversity of pentasaccharides, we undertook the challenge of re-designing the active site of a branching sucrose, BRS-B Δ 2, found so far to be the most efficient enzyme for the glucosylation of **ABC'D'** (Chapter II). Although the crystallographic structure of this enzyme was not available, the construction of a reliable 3D model by comparative modelling enabled the use of structure-based engineering strategies. Using a combination of molecular modelling methods and RosettaDesign-based approach, we explored amino acid mutations in the active site that could favor catalytically productive binding of **ABC'D'** in various orientations to enable regioselective glucosylation of units **A**, **C'** or **D'** of the tetrasaccharide, targeting in priority pentasaccharide products characteristic of prevalent *S. flexneri* serotypes 3a, 2a and 1a/1b. Evidently, conservation of the original specificity for sucrose donor substrate was also one of the main concerns during the redesign of the enzyme active site. The strategy disclosed herein led to the selection of 49 mutants containing up to 15 mutations scattered over 27 amino acid positions of the acceptor binding subsites. Mutants were then constructed and assayed for glucosylation of **ABC'D'**. Overall, our results demonstrate the versatility of the BRS-B Δ 2 scaffold, which by the introduction of specific combinations of mutations in the active site gave access to broader range of 1,2-*cis* glucosylation patterns from a common acceptor molecule.

Results & Discussion

Computer-aided re-design of BRS-B Δ 2 branching sucrase active site

With the aim of performing site-selective glucosylation of tetrasaccharide **ABC'D'** and thus enlarging accessible pentasaccharide diversity, we focused our work on the redesign of the active site of a branching sucrase named BRS-B Δ 2, truncated of 153 amino acids from the N-terminal end compared to native BRS-B (Genbank accession number CDX65123.1). Selection of this enzyme was based on a prior work (Chapter II), which revealed that it was the best performing native branching sucrase for the glucosylation of tetrasaccharide **ABC'D'**, yielding largest amounts of pentasaccharide **P1**. Deletion at the N-terminal extremity allowed a better recombinant production of the enzyme in soluble form, greater easiness of purification and concentration compared to the parental enzyme or BRS-B Δ 1 form (Chapter II) without altering the enzyme activity. As the three-dimensional structure of this enzyme was unknown, we decided to construct of a 3D-model of BRS-B Δ 2 using Δ N₁₂₃-GBD-CD2, for which an X-ray structure is available (PDB ID: 3TTQ), as template. BRS-B Δ 2 contains 1053 amino acid residues and shares 49% identity with Δ N₁₂₃-GBD-CD2. When taking into account only the active site of the enzymes, the sequence identity increases to 60%, indicating a high conservation of active site residues. This allowed construction of a reliable 3D-model of BRS-B Δ 2 by comparative modelling and opened the way to computer-aided design approaches illustrated in the framework of Figure 1A. Using the 3D model, pentasaccharide products characteristic of prevalent *S. flexneri* serotypes, ABC'[E(1→4)]D' (*S. flexneri* 1a/1b), AB[E(1→4)]C'D' (*S. flexneri* 2a) and E(1→3)ABC'D' (*S. flexneri* 3a), were docked in the active site (Figure 2). The crystallographic structure of homologous GTF180 glucansucrase (PDB ID: 3HZ3) in complex with sucrose bound in the active site (Vujicic-Zagar et al. 2010) was used as template to guide docking of the different pentasaccharides. More particularly, pentasaccharides were initially constructed using 3D coordinates of the sucrose glucosyl moiety from the crystallographic complex. Systems were subsequently subjected to simulated annealing (from 0 to 350 K in 100 ps and vice versa) in vacuum with harmonic positional restraints of 50.0 kcal/mol/Å² on the enzyme, the glucosyl unit, and the sugar pucker rings of **ABC'D'**. The lowest energy systems from each simulated annealing were then selected as starting points for the computational enzyme design procedure. After excluding the catalytic residues (the nucleophilic D671, the acid/base E709 and the transition state stabilizer D1136) and other amino acid residues described as important for either catalysis or sucrose recognition (R669, H787, Y144 and D1183 identified by homology with GTF-180) (Vujicic-Zagar et al. 2010), we selected 27 mutable (or designable) positions in total: 24 in catalytic domain A and 3 in domain B - (G594, W595 and F596 residues) (Figure 1B) on the basis of Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) calculation (Figure S1) and careful visual inspection.

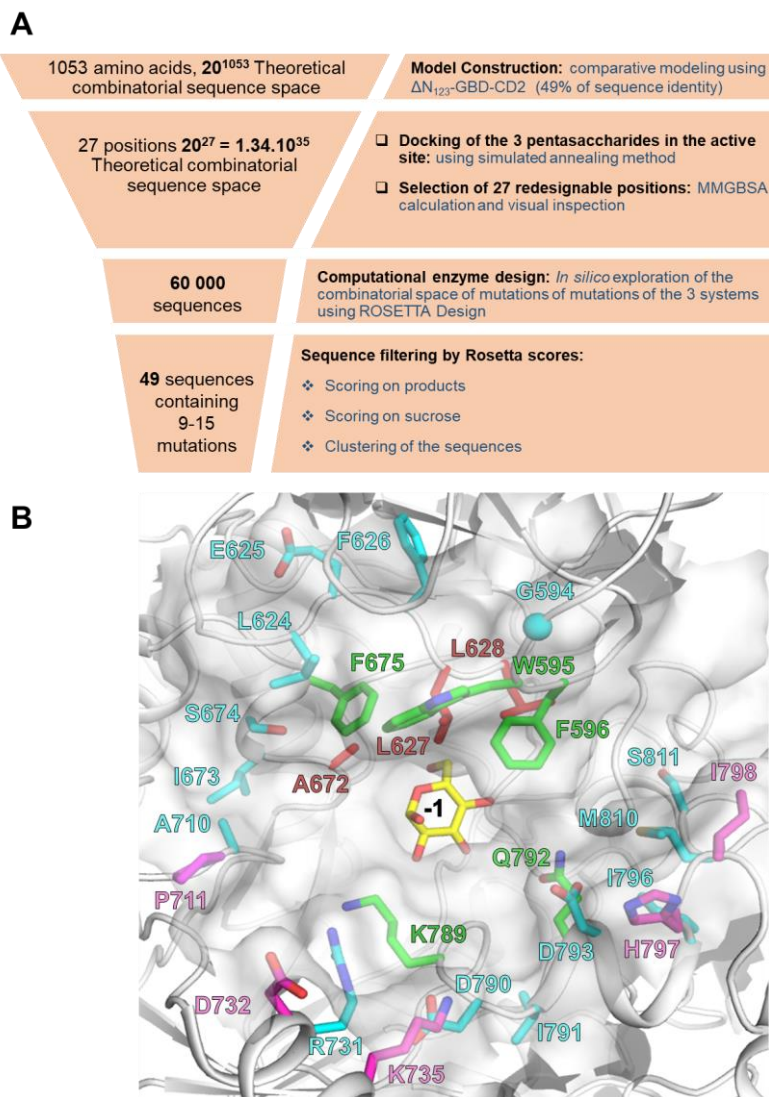


Figure 1: A. Computer-aided approach for BRS-B $\Delta 2$ active site re-design. After successive steps of the framework, the final designed library contained 49 sequences with between 9 and 15 mutations in the active site. **B.** View of the active site of BRS-B $\Delta 2$ model, The 27 redesignable amino acids are showed in red, green, cyan and magenta according to their belonging to the first, second, third and fourth contact shell respectively from the glucosyl (yellow) in the -1 position.

Most of these positions belong to loops with the exception of R731, D732, K735, D793 and H797, which are located in α -helices. Overall, these 27 residues are scattered over the catalytic site along four successive contact shells, due to the sliding of the **ABC'D'** core with respect to -1 subsite where glucosyl is bound (Figure 1B). Moreover, all selected residues are found highly variable among all known GH70 enzymes (CAZy database) with the exception however of residues L627 and A672 from the first shell which are strictly conserved (Table S1). It should be noticed that, if all 27 positions were considered to be mutable by one of the 20 possible amino acid residues, the theoretical combinatorial sequence space would have been as large as 1.34×10^{35} , clearly out of reach of currently available experimental screening methods and computational approaches. To explore and reduce this combinatorial space, we thus considered an enzyme design protocol based on the sampling of mutations at each of the 27 positions by performing 60 000 independent runs (20 000 per system) of Rosetta

Enzyme_Design, which takes into account the backbone and side chain flexibility of residues. The designed sequences were subsequently filtered using the Rosetta scores corresponding to enzyme:pentasaccharide binding interaction for each of the three systems, followed by a second round of sequence filtering based on the docking of sucrose donor in each mutant and estimation of its binding interaction (Figure S2). Finally, in order to limit sequence redundancy and enhance sequence variability in the final set of sequences, all designed sequences were clustered based on the percentage of sequence identity. The best Rosetta scores (with respect to pentasaccharides and sucrose) of each cluster (total of 49) were then selected for experimental evaluation. The set of 49 sequences provided in Table S2 contained between 9 and 15 mutations located between residues 595 and 811. These sequences were classified in three groups depending on the targeted pentasaccharide: group I (M1 to M16) for ABC'[E(1→4)]D' (*S. flexneri* 1a/1b), group II (M17 to M34) for AB[E(1→4)]C'D' (*S. flexneri* 2a) and group III (M35 to M49) for [E(1→3)]ABC'D' (*S. flexneri* 3a) (Figure 2).

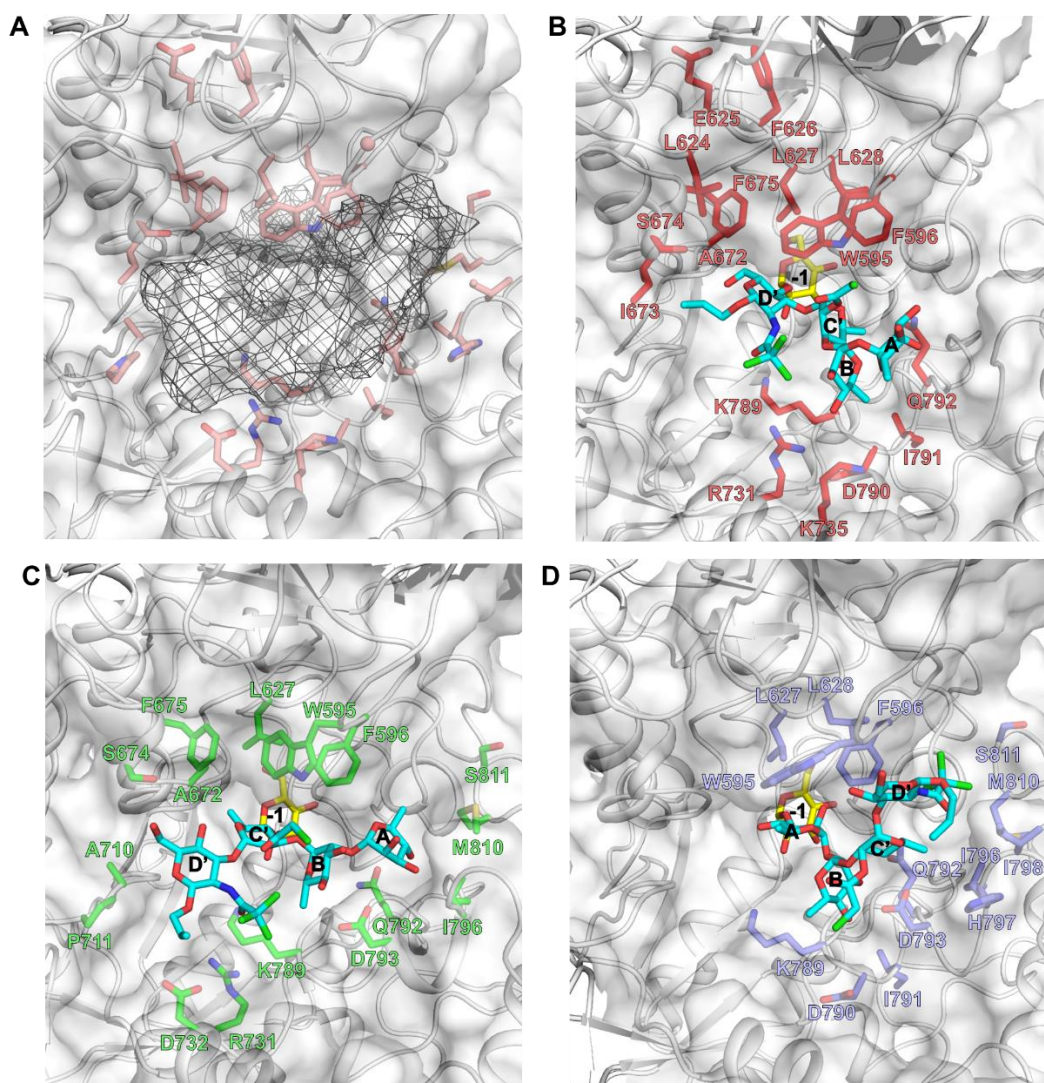


Figure 2: View of the space occupied by the three gathered pentasaccharides targeted by the design bound in the active site of parental BRS-B $\Delta 2$ (panel A) as well as targeted amino acid residues in each of the designs: ABC'[E(1→4)]D' (*S. flexneri* 1a/1b) targeted by mutants M1-M16 from Group I is shown in panel B; AB[E(1→4)]C'D' (*S. flexneri* 2a) targeted by mutants M17-M34 from Group II is shown in panel C; [E(1→3)]ABC'D' (*S. flexneri* 3a) targeted by mutants M35-M49 from Group III is shown in panel D.

Recombinant production of mutants

Unlike parental BRS-B $\Delta 2$ enzyme, recombinant expression of the 49 mutants in *E. coli* led to the formation of inclusion bodies and no soluble protein could be recovered. Therefore, optimization of the production conditions was carried out, notably by testing the influence of culture media and various chaperone proteins which have been described to assist protein folding (Thomas et al. 1997) and to help to prevent aggregation (Nishihara et al. 2000). We selected mutants M14 and M34, which contained respectively 15 and 11 mutations, and attempted their co-production in the presence of different combinations of chaperone proteins (groES, groEL, tig, dnaK, dnaJ, grpE) (Table S3) to enhance solubilization and prevent aggregation. Using co-expression with plasmids pG-KJE8 (coding for dnaK, dnaJ, grpE, groES, groEL) and pTf16 (coding for tig), mutant proteins were recovered in the soluble fraction. The production was further improved by growing the cells at 21°C using a culture medium composition previously optimized for the production of another branching sucrase, ΔN_{123} -GBD-CD2 (Vuillemin et al. 2014). Finally, we monitored the production level during the culture and showed that the highest values were reached after 24 h when using plasmid pTf16 and 32 h when using plasmid pG-KJE8 for chaperon production. Based on these results, the 49 mutants were successfully produced using the optimized culture conditions and chaperones encoded by either plasmid pTf16 or pG-KJE8. The optimization of the soluble expression of mutant M21 is presented in Figure S3 as an example.

Screening of the mutant library for the ability to use sucrose as donor substrate

In order to assess the capacity of the mutants to utilize sucrose and release reducing sugars, an “ON/OFF” and non-quantitative assay was set up. The soluble fractions were incubated with sucrose for 70 h and reducing sugar production was determined using 3,5-dinitrosalicylic acid (DNS) (Miller 1959). The variation of absorbance at 540 nm obtained for the parental enzyme and the mutants is reported in Figure 3, panels A, C, and E. This end-point assay showed that all the mutants were strongly impacted for sucrose consumption compared to the parental enzyme.

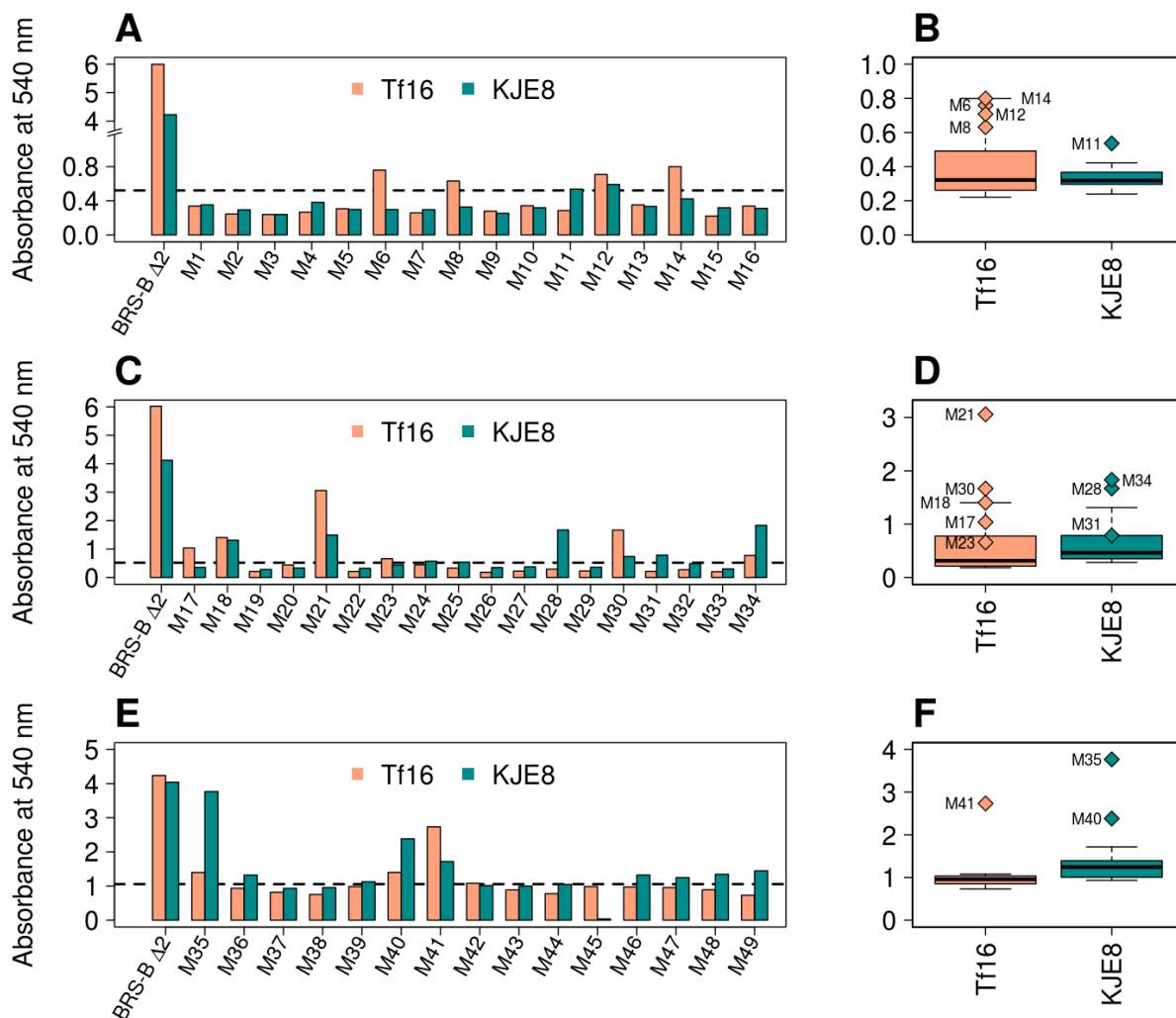


Figure 3: A, C, E: Relative absorbance at 540 nm (in % compared to the parental enzyme) determined after the DNS–based “ON/OFF” assay using crude enzyme extracts of the 49 mutants co-produced with either *tig* (Tf16, lightsalmon) or *dnaK, dnaJ, grpE, groES, groEL* chaperone proteins (KJE8, darkcyan) in presence of 100 g/L sucrose during 70 h incubation. The parental enzyme BRS-B Δ2 is indicated as a reference. Dashed lines indicate absorbance level of the inactive mutant BRS-B Δ2 E709Q. **B, D, F:** boxplot representing the absorbance data. Parental enzyme BRS-B Δ2 was excluded from the analysis. Data points (diamonds) pinpoint the 16 mutants selected for further evaluation.

Out of the 49 mutants, the boxplot analysis enabled us to retain 16 mutants found in the top tertile (and above) of the absorbance boxplot (Figure 3 panels B, D and F). These 16 mutants were purified to homogeneity, and among them, four mutants quickly aggregated after purification (M8, M12 and M17 produced with Tf16 and M11 produced with KJE8), indicating that their stability was impacted by the mutations. Active mutants (12 out of 49 screened, 24.5% active mutants in total) belonged to all three groups, with 2 from group I (12.5% of active mutants of this group), 7 from group II (38.8% of active mutants of this group), and 3 from group III (20% of active mutants of this group).

The 12 resulting mutants (M6, M14, M18, M21, M23, M30 and M41 produced with Tf16 and M28, M31, M34, M35 and M40 produced with KJE8) were selected for further evaluation of their ability to glucosylate **ABC'D'**. Meanwhile, specific activity toward sucrose was determined only for mutants M21, M23, M30, M34 and M35 (Table 1). It was found to be only 1.92 % of

the parental enzyme for M21, 0.23 % for M23 1.23 % for M30, 0.1 % for M34, and 1% for M35, indicating a tremendous loss of activity of mutants toward sucrose donor substrate.

Table 1. Relative absorbance at 540 nm (in % compared to the parental enzyme) determined after the DNS–based “ON/OFF” assay using crude enzyme extracts of the 12 selected mutants in presence of 100 g/L sucrose during 70 h incubation. Specific activity on sucrose (in U.mg⁻¹ of purified enzyme) was determined for mutants M21, M23, M30, M34 and M35. ^a Unpublished data from Marion Claverie (2014). *n.a.*: not available

	Relative absorbance compared to parental enzyme (%)	Specific activity on sucrose (U.mg ⁻¹ of purified enzyme)
BRS-B Δ2	100	26 ^a
M6	13.7	<i>n.a.</i>
M14	19.4	<i>n.a.</i>
M18	31.7	<i>n.a.</i>
M21	55	0.46
M23	11.9	0.06
M28	40.5	<i>n.a.</i>
M30	30	0.30
M31	19.1	<i>n.a.</i>
M34	44.3	0.02
M35	91.3	0.24
M40	57.7	<i>n.a.</i>
M41	49.1	<i>n.a.</i>

Transglucosylation of tetrasaccharide **ABC'D'**

All 12 selected mutants were tested for their ability to glucosylate tetrasaccharide **ABC'D'** and their products were further analyzed by LC-MS. Unlike previous assays (chapter II and III), reaction pH was set to 5.75 despite the **ABC'D'** stability decrease at this pH, due to the quick aggregation of mutants when lowering the pH. All selected mutants were found able to transfer glucosyl moieties onto **ABC'D'**, yielding at least one and up to four mono-glucosylated products (Figure 4 and S4). Overall, six different pentasaccharide products were detected which were named **P1**, **P2**, **P2'**, **P2''**, **P3** and **P3'** based on their retention time (t_R) and molecular mass. Mutant M6 showed similar profile to the parental enzyme BRS-B Δ2 and produced mainly **P1** (t_R = 21.6 min) together with a small amount of **P2** (t_R = 21.9 min). Conversely, mutants M14, and M34 revealed an opposite trend, producing mainly **P2** with significantly lowered amounts of **P1**. Interestingly, mutant M18 lost completely its ability to form **P1** in favor of the exclusive production of **P2**.

Another group of mutants composed of M21, M23, M28, M30, M31 and M41 shared a common product profile. While they all produced various ratio of products **P1** and **P2** or **P2'**, they also formed a novel pentasaccharide named **P3**, not synthesized by the parental enzyme. Mutant M21 was found to produce **P3** in larger amount. Interestingly, based on its retention time (t_R =

16.3 min), this product **P3** was found to be also produced by mutant W2135S-F2136L of the enzyme ΔN_{123} -GBD-CD2 in our earlier report (Chapter III).

Together with trace amount of **P3**, mutant M23 also produced two co-eluted products ($t_R = 21.9$ min) after UV analysis, **P2** and **P2'**. They revealed however distinct signals in the more sensitive HRMS analysis, suggesting thus structural differences (data not shown). Of particular interest, M35 and M40 both produced a novel distinct product, not identified before, named respectively **P2''** ($t_R = 22.4$ min) and **P3'** ($t_R = 15.3$ min), although in trace quantities that did not allow production and purification of enough amount for NMR characterization.

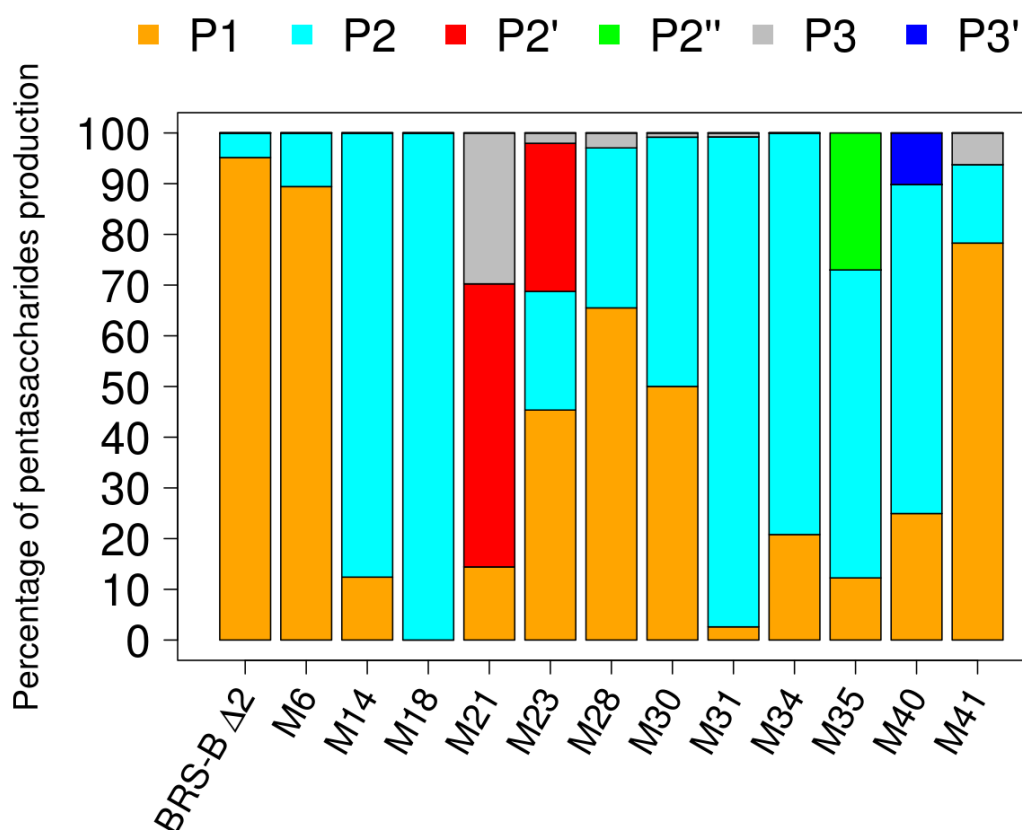


Figure 4: Representation of the ratio of the products formed by BRS-B $\Delta 2$ parental enzyme and its mutants, detected by HPLC-UV, after 16 h of reaction at 50 μ L scale in presence of 50 mM **ABC'D'** and 1 M sucrose at pH 5.75, using variable amounts of purified enzymes. High Resolution Mass Spectrometry revealed mass increase of 162 $\text{g}\cdot\text{mol}^{-1}$ compared to **ABC'D'**, corresponding to mono-glucosylated tetrasaccharides.

Structural characterization of the formed products

Mutant M21 forming products **P1**, **P2'** and **P3** was selected to produce sufficient amounts of pentasaccharides for structural characterization using NMR spectroscopy. M21 was produced and purified to homogeneity and used to carry out a 1 mL-scale reaction in presence of **ABC'D'** and sucrose. Structure of **P1** was previously determined without ambiguity (Chapter II). Only, products **P2'** (5 mg) and **P3** (1.6 mg) were isolated.

The 1D ^1H NMR spectrum obtained for **P2'** isolated using M21 was perfectly superposed to the one previously obtained using mutant F2163G of ΔN_{123} -GBD-CD2 (Chapter III) (Figure S5). This confirmed that glucosylation in **P2'** occurred on OH-4_A, and that **P2'** corresponds to the

co-eluted product of **P2** identified in chapter III. However, in the case of M21, **P2'** ($t_R = 21.9$ min) was produced alone without any trace of a distinct mass that could indicate production of **P2** ($t_R = 21.9$ min). The analysis of HSQC spectra of **P3** showed shifted resonances of the glucosylated positions and adjacent atoms (Figure S6); the glucosylated 4B carbon was high frequency shifted from 72.17 ppm to 80.85 ppm, while the adjacent 3B and 5B were shifted to lower frequency, from 69.8 ppm and 69.57 ppm to 68.55 ppm and 68.45 ppm, respectively. The cross-correlation peaks of the shifted **B** and **E** unit resonances were assigned using Double Quantum Filtered COrrrelation (QDF COSY) (Figure S7).

Given the very low amounts of **P2''** and **P3'** pentasaccharides produced by mutant M35 and M40, respectively, we turned to MS/MS, which is a more sensitive analytical method requiring less quantity of material (few μg) and avoiding laborious production and purification steps. First, the method was validated with **P1** as its structure was known. On the basis of its specific fragmentations, the glucosylation on OH-6_{D'} was confirmed (Figure S8). Next, we characterized the structures of **P2''** and **P3'** (Figure 5). The results revealed that glucosylation had occurred on OH-3_B, representative of *S. flexneri* serotype 5a, for **P3'**, and on OH-2_A for **P2''**.

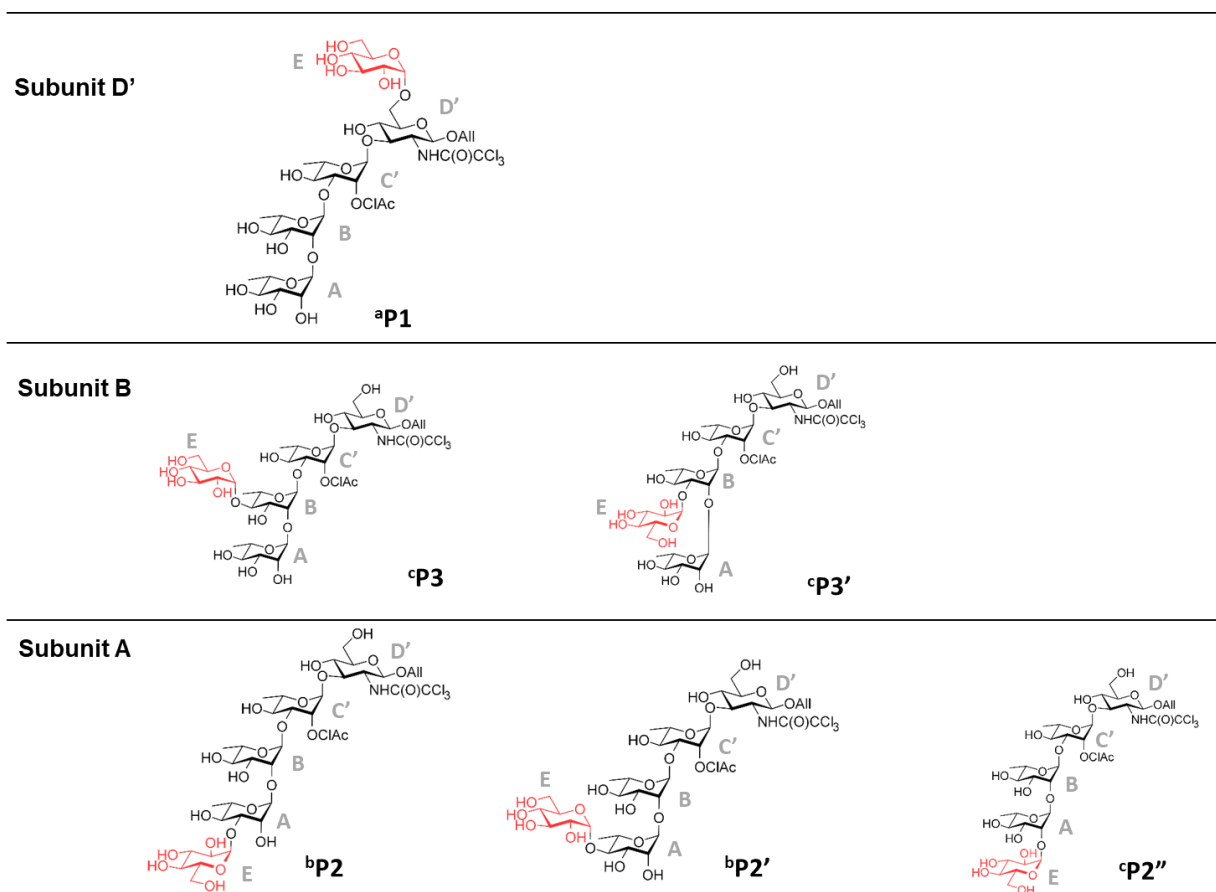


Figure 5: Diversity of pentasaccharides obtained with the mutants. The various products are shown on the basis of the glucosylated subunit (**A**, **B** or **D'**). **aP1** ABC'[E(1→6)]D', characterized in chapter II. **bP2** [E(1→3)]ABC'D' and **P2'** [E(1→4)]ABC'D' characterized in chapter III. **P3**, **P3'** and **P2''** were characterized in this study to be A[E(1→4)]BC'D' and A[E(1→3)]BC'D', and E[(1→2)]ABC'D', respectively. **P3** was characterized by NMR spectroscopy and **P3'** and **P2''** by MS/MS.

Insight on the impact of mutations on product profiles

BRS-B $\Delta 2$ naturally favors glucosylation of primary hydroxyl group OH-6_D from **ABC'D'**. Group I comprising 16 mutants (M1 to M16) aimed at introducing beneficial mutations that could shift binding of **D'** at +1 subsite in order to re-orient OH-4_D in catalytically productive manner with respect to anomeric C1 of glucosyl moiety at -1 subsite (Figure 3, panel A). Redesigning **D'** binding site required introduction of 12-15 mutations to unclutter the active site and enable steric access of hindered secondary OH-4_D to the glucosyl moiety. However, only two mutants (M6 and M14) out of the 16 members of group I turned out to be active on sucrose, indicating a highly detrimental impact of mutations on first step of the reaction. In the presence of **ABC'D'**, these mutants kept the innate specificity of their parental enzyme, producing only pentasaccharides **P1** and **P2** but in different proportions. In spite of its 13 mutations and similarly to the parental enzyme, mutant M6 revealed a preference for **D'** subunit glucosylation to produce **P1**. Conversely, mutant M14 containing 15 mutations of which 10 differed from M6, shifted its glucosylation efficiency toward **A** subunit, favoring production of **P2**. These results show that the selected combination of mutations did not force binding of **D'** at +1 acceptor subsite, probably due to major steric constraints related with the presence of protecting groups and size of the acceptor.

Group II encompassing 18 mutants (M17 to M34) aimed originally at glucosylating OH-4_C, the sole available hydroxyl function of **C'** moiety (Figure 3, panel C). Accessibility of this secondary hydroxyl group was found highly hindered due to the constrained β -1,3 linkage between rhamnosyl residues **B** and **C'**, and the presence of the three protecting groups at **C'** and **D'** subunits. Introduction of 10-13 mutations were proposed by the design. These mutations less severely affect sucrose recognition as 7 mutants were found relatively active toward sucrose and successfully glucosylated **ABC'D'**, producing up to 4 distinct pentasaccharides (**P1**, **P2**, **P2'** and **P3**). The versatility of these mutants to accommodate **ABC'D'** in different manners led to broader product specificity yielding novel molecular diversity (Figure 5) but they also induced considerable loss of glucosylation selectivity (Figure 4). With the exception of M28 that had a preference for glucosylation of **D'**, all 6 remaining mutants rather exhibited a preference for glucosylation at **A** and **B** moieties. Two mutants, M21 and M23 that both contained 10 mutations, were the only mutants producing **P2'**, glucosylated at OH-4_A, not synthesized by the parental enzyme. However as **P2** and **P2'** are co-eluted products, either the development of a new separation method or an accurate structural characterization would be required to confirm these results. Interestingly, most mutants revealed their ability for inner chain glucosylation at **B** moiety, producing **P3**, which was never reported to be synthesized by native branching sucrases.

Group III gathers 15 mutants (M35 to M49) containing between 9 and 12 mutations aiming at favoring end chain glucosylation, targeting OH-3_A to improve production of **P2**, only weakly yielded by parental enzyme (Figure 3, panel D). Here again, introduced mutations turned out to drastically affect sucrose recognition. However, all three mutants (M35, M40 and M41)

revealed their ability to form, in addition to **P1** and **P2**, a new product not synthesized by parental enzyme. Glucosylation of OH-2_A was observed for the first time with mutant M35, yielding pentasaccharide **P2** and enabling glucosylation of the third hydroxyl group from **A** moiety, although of less interest in *S. flexneri* context as this position is involved in chain elongation. Mutants M40 and M41 enabled glucosylation of inner chain **B** subunit, providing respectively access to **P3'** (glucosylated at OH-3_B) and **P3** (glucosylated at OH-4_B). Here again, with these mutants, we gained access to all possible glucosylation patterns of the **B** moiety.

In our previous study, we used a graph approach using Fruchterman-Reingold algorithm to compare the impact of the mutations on the topology of the active site of ΔN_{123} -GBD-CD2 branching sucrose and its mutants and possibly on their product profiles (Chapter III). We used herein the same technique in order to map clusters of amino acid interactions within the active site of BRS-B $\Delta 2$ enzyme used as template for computational design strategy (Figure S9). However, analysis of the mutants did not allow relating the mutation clusters with product specificity.

Computational protein design undertaken here was highly challenging as 4 subsites had to be re-designed to improve tetrasaccharide **ABC'D'** accommodation for each targeted pentasaccharide without losing the affinity for sucrose. The huge combinatorial sequence corresponding to the 27 selected positions of mutations was drastically reduced thanks to the design. By screening a very limited set of 49 sequences, containing high number of mutations (between 9 and 15 depending on the design), several mutants were successfully isolated. Impressively, after a challenging structural analysis of the products, we found out that these mutants enabled the glucosylation of six out of the eight hydroxyl functions of **ABC'D'**, three of these pentasaccharides being characteristic of *S. flexneri* serotypes (4a/4b, 5a and 3a) for which no synthetic route has been proposed yet.

Remarkably, two mutants (M40 and M41) showed a product profile in line with the design expectations and produce higher amounts of product **P2** than the parental enzyme (data not shown). Experimental deconvolution of the mutations on these mutants could help to better understand the contribution of the different amino acid mutations and their combinatorial effect. Such deconvolution could also be applied to the other mutants. However, given the high number of mutants and mutations, this would remain highly tedious and challenging.

These results highlight difficulties of redesigning such a large and exposed active site as that of branching sucroses for the selective glucosylation of a structurally complex and chemically modified molecule, presenting no similarity with the natural substrate. This required introduction of a significant load of amino acid mutations to target glucosylation at various hydroxyl positions but also made difficult to narrow down selectivity of the enzyme for the

production of only one product. The effect was further pronounced due to the highly exposed and flexible active site that made difficult to specialize the enzyme. Furthermore, re-engineering enzymes catalyzing multi-step reactions involving multiple substrates considerably enhanced complexity of the design, requiring multi-objective optimization to develop mutants still able to utilize sucrose donor but that had also gained novel specificity toward exogenous acceptor.

Another limitation in the computational redesign of the enzyme was the lack of crystallographic structure that led us to assume that reliable 3D-modelling could be performed given the high sequence identity between BRS-B $\Delta 2$ and GH70 branching sucrose of known structure (ΔN_{123} -GBD-CD2). Nonetheless, suggested flexibility of several loops surrounding the active site could drastically alter topology of the active site and recognition of acceptor **ABC'D'**. To circumvent the lack of structural data, we still modelled binding of highly complex pentasaccharide products in the active site and we had to make an arbitrary selection of a bound conformation to target by combinatorial space exploration. Undoubtedly, accuracy of the design would have benefited from the determination of crystallographic structures of the enzyme in complex with the acceptor or the products. However, successful outcome of the design would still not be warranted given the many limitations still faced by computational protein design methods such as the poor integration of molecular flexibility and conformational rearrangements (Chapter III), the under consideration in energy functions of entropy penalty and solvent effect, etc.

Conclusion

Computer-aided design was for the first time applied to the redesign of several acceptor subsites of a branching sucrose to redirect glycosylation regioselectivity toward various hydroxyl functions of a lightly protected tetrasaccharide, precursor in the synthesis of serotype-specific *S. flexneri* haptens. By predicting introduction of as many as 15 amino acid mutations in the active site, mutants were found to be able to produce up to six distinct pentasaccharides, whereas only two were synthesized by the parental enzyme. We demonstrated for the first time that mutants could perform branching reactions onto **ABC'D'**, and could advantageously open the way to the glycosylation of longer oligosaccharide chains, such as the *S. flexneri* oligosaccharide backbone. Given that no equivalent enzymatic activity has been identified in Nature, to our knowledge, these mutants are promising starting templates for further rounds of evolution and/or optimization. Yet, fine tuning of this enzyme's active site is still fraught with difficulty because of the lack of detailed understanding of the reaction mechanism and 3D structural information. Although objectives of the computational design were not fully met, the impressive malleability of acceptor binding site and the demonstrated ability to expand the tolerated productive binding modes for the **ABC'D'** acceptor still led to very unique and exciting achievements that could offer novel opportunities for the development of high convergent chemo-enzymatic routes toward *S. flexneri* haptens.

Material & Methods

Chemical synthesis of the tetrasaccharide **ABC'D'**

Chemical synthesis of the tetrasaccharide **ABC'D'** was performed as previously described (Chapter II)

Computational enzyme design

The 3D model of BRS-B $\Delta 2$ was constructed by comparative modelling using MODELLER software (Eswar et al. 2006), using the reference coordinates of the template ΔN_{123} -GBD-CD2 (PDB ID: 3TTQ) (Brison et al. 2012). The same strategy described in our previous study of the complexes ΔN_{123} -GBD-CD2: pentasaccharides was used for docking ABC'[E(1→4)]D', AB[E(1→4)]C'D' and [E(1→3)]ABC'D' pentasaccharides in the active site and for calculation of the force field parameters (Chapitre III). Between 12 and 16 designable positions were selected for the redesign based on our intuition after careful visual inspection of the enzyme:pentasaccharide interfaces using Pymol software (DeLano 2009). This graphical examination was helped by the MMGBSA calculations performed to evaluate the individual amino acid contributions to the free energy binding of BRS-B:pentasaccharides for each system and described in details in Supporting Information. The residues having C α within 10 Å of the C α of the redesignable residues were allowed to be repacked. 20 000 independent runs were carried out for each system using RosettaDesign (Richter et al. 2011) software and Beta_Nov15 energy function (Alford et al. 2017). The output sequences were filtered based on different Rosetta scoring and filtering schemes. The RosettaDesign protocol and sequence filtering are described in deep details in the Supporting Information and Methods.

Mutant construction

Plasmid pET55-BRS-B_ $\Delta 2$ _WT was obtained after 5' truncation of 459 base pairs in the gene *brsB* from *Leuconostoc citreum* NRRL B-742 using the method described by Wang and Malcom (Wang and Malcolm 1999). Briefly, the method comprised the use of a two-stage procedure, based on the QuicChange Site-directed Mutagenesis protocol (Stratagene, La Jolla, CA, USA) with a pre-PCR consisting in a single-primer extension stage before the standard protocol and allowing deletion of the sequence of interest. The PCR product was cloned into pET55 plasmid using Gateway[®] technology (Thermo Fisher Scientific) for expression in *E. coli* under the control of T7 promoter and with a gene for ampicillin resistance. Plasmid pET55-BRS-B_ $\Delta 2$ _E709Q was obtained after performing E709Q mutation in *brsB* gene. Gene *brsB* was amplified by inverse PCR using 5'-phosphorylated forward primer CAT-ATT-TCA-ATT-GTT-**CAG**-GCT-CCA-AAG-GGG-GAA-AC and 5'-phosphorylated reverse primer TAT-GTT-GAT-TGG-CAA-CTG-CCT-CAT-TGT-CAG that introduced mutation E709Q.

Parental plasmid was digested by DpnI enzyme (NEB) and ligation was performed using T4 DNA ligase (NEB). The reaction was purified using NucleoSpin® PCR Clean-up (Macherey-Nagel) before transformation into *E. coli* TOP10 (Invitrogen). Correct sequence was verified by Sanger sequencing (GATC Biotech). Plasmids pET55-BRS-B_Δ2_M1 to pET55-BRS-B_Δ2_M49 were purchased from GenScript.

Production of a library of 49 BRS-B Δ2 mutants

Plasmids pTf16 containing chaperone gene *tig* under the control of promoter *araB*, and pG-KJE8 containing chaperone genes *dnaK*, *dnaJ*, *grpE*, *groES*, *groEL* under the control of promoters *araB* and *Pzt-1* (Takara) were chemically transformed into *E. coli* BL21 Star™ (DE3) (Life Technologies). Strains were rendered competent again and chemically transformed using plasmids pET55-BRS-B_Δ2_WT, pET55-BRS-B_Δ2_E709Q and pET55-BRS-B_Δ2_M1 to pET55-BRS-B_Δ2_M49 encoding for parental BRS-B Δ2, inactive mutant of BRS-B Δ2 and the 49 BRS-B Δ2 mutants, respectively. After overnight preculture at 37°C in LB medium (10 g.L⁻¹ tryptone, 5 g.L⁻¹ yeast extract, 10 g.L⁻¹ NaCl) containing 100 μg.mL⁻¹ ampicillin (Euromedex) and 12.5 μg.mL⁻¹ chloramphenicol (Sigma), the main culture was inoculated at optical density at 600 nm (OD_{600nm}) = 0.05 and used (i) modified ZYM-5052 medium with 0.1% D-lactose, 1% glycerol ((Studier 2005), Chapter II), (ii) 2x YT medium (20 g.L⁻¹ tryptone, 10 g.L⁻¹ yeast extract, 10 g.L⁻¹ NaCl) or (iii) optimized ZYM-5052 with 0.05% D-glucose, 0.75% D-lactose, 1.5% glycerol (Vuillemin et al. 2014). Medium were supplemented with 100 μg.mL⁻¹ ampicillin and 12.5 μg.mL⁻¹ chloramphenicol as selection markers. When using medium (ii) and (iii), chaperone protein expression was induced immediately after inoculation by adding 4 g.L⁻¹ of L-arabinose (ACROS Organics) for pTf16 or 4 g.L⁻¹ of L-arabinose and 10 ng.mL⁻¹ tetracycline (Sigma) for pG-KJE8. Cultures (i) and (iii) were grown at 21°C in either 50 mL, 200 mL or 1 L scale in baffled flasks for 24 h (pTf16) or 32 h (pG-KJE8). When using medium (ii), production of the recombinant enzymes was induced using isopropyl β-D-1-thiogalactopyranoside (IPTG) when OD_{600nm} reached 0.4 after growth at 37°C, and the temperature was lowered to 21°C for another 24 h culture. Cells were harvested by centrifugation and re-suspended in appropriate volumes of NaAc buffer 50 mM pH 5.75 to concentrate the cells at OD_{600nm} = 15. Cells were then disrupted by sonication and the cellular debris were removed by centrifugation at 13,000 g for 30 minutes at 8°C to recover the soluble fraction. Both insoluble and soluble fractions were run on SDS-PAGE to assess the production of the enzymes.

Library screening on sucrose in liquid DNS-based ON/OFF assays

Hydrolytic activity of the enzymes were assessed at 30°C in presence of 100 g.L⁻¹ sucrose in NaAc buffer 50 mM pH 5.75. After 70 h incubation, reactions were mixed with the same volume of DNS (3,5-dinitrosalicylic acid) solution (Miller 1959) and heated to 95°C for 5 minutes.

Absorbance was read at a wavelength of 540 nm and active enzymes were selected by comparing the absorbance value with the one from inactive mutant E709Q that constituted the baseline.

Purification and concentration of the best mutants

Active mutants on sucrose were produced as previously described with the following exception: after sonication the soluble fraction was recovered by centrifugation at 50,000 *g* (instead of 13,000 *g*) for 30 minutes at 8°C. Purification was performed as previously described (Chapter II, Chapter III). The mutant enzymes were then concentrated using AmiconUltra-15 (Merck Millipore) with a cutoff of 50 kDa. Concentration was possible to different extents before aggregation depending on the mutant. Final concentrations were measured using Nanodrop instrument (Thermo Scientific) and taking into account the theoretical molecular weight (MW = 132 480 g.mol⁻¹) and molar extinction coefficient ($\epsilon = 196\,770\text{ M}^{-1}.\text{cm}^{-1}$) of BRS-B $\Delta 2$ calculated with ProtParam tool from ExPASy server. Concentrations were 12.5 g.L⁻¹ for M6, 1.3 g.L⁻¹ for M14, 3.1 g.L⁻¹ for M18, 2.9 g.L⁻¹ for M21, 7.1 g.L⁻¹ for M23 and 3.9 g.L⁻¹ for M30, 0.2 g.L⁻¹ for M28, 0.2 g.L⁻¹ for M31, 3.7 g.L⁻¹ for M34, 0.7 g.L⁻¹ for M35, 1.9 g.L⁻¹ for M40 and 1.0 g.L⁻¹ for M41.

Specific activity determination on sucrose

Specific activities of mutants M21, M23, M30, M34 and M35 were determined at 30°C in NaAc buffer 50 mM pH 5.75 over the course of 50 minutes with sampling every 10 minutes, using 100 g.L⁻¹ of sucrose and purified concentrated enzymes.

Reaction assays with sucrose and **ABC'D'** acceptor

Reactions were performed in presence of tetrasaccharide acceptor **ABC'D'** (50 mM) and sucrose (1 M) in miniaturized assay at 50 μL scale, in NaAc buffer 50 mM pH 5.75. After 16 h incubation, reactions were stopped using a solution of acetonitrile at 30% supplemented with 0.08% TFA.

HPLC-UV-MS / HRMS for tetrasaccharide and product profiles

The same analytical methods as in Chapter III were used.

Determination of pentasaccharide structures

NMR experiments

The samples were dissolved in DCI-containing D₂O at pH 5.1. For NMR studies, the samples were lyophilized three times and dissolved in 180 μ L of 99.9% DCI-containing D₂O. All NMR spectra were recorded on a Bruker Avance spectrometer operating at a proton frequency of 950 MHz (TGIR- RMN-THC Fr3050 CNRS, Gif-sur-Yvette) and at a carbon frequency of 238 MHz with a 5-mm gradient indirect cryoprobe. All spectra were processed and analyzed with Topspin software (Bruker) and Sparky software (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

¹H and ¹³C 1D NMR spectra were accumulated at 25 °C, 65536 data points were acquired with 32 and 2048 scans respectively for proton and carbon experiments. ¹H-¹³C HSQC (Heteronuclear Single Quantum Coherence spectroscopy) and Double Quantum Filtered COrrrelation SpectroscopY (QDF COSY) (Rance et al. 1983) experiments were performed at 25 °C. Homo and heteronuclear spectra were recorded under the following experimental conditions: 512 increments of 2048 complex points are acquired with an accumulation of 16 scans. Spectral widths were 16025 Hz for protons dimension and 44267 Hz for carbon dimension.

MS/MS

MS/MS studies were performed by D. Ropartz, M. Fanuel and H. Rogniaux on BIBS facility at INRA Nantes. The analyses were carried out on a LC-ESI-Q-ToF platform composed of an Acquity UPLC H-class system, coupled to a Synapt G2Si HD mass spectrometer (Waters Corp., Manchester, UK). Reaction media were diluted 250 times in H₂O/acetonitrile (95.5:4.5, v/v). Separation of glucosylated products was performed by LC using a Porous Graphitized Carbon (PGC) (Hypercarb (2,1 mm x 100 mm, 3 μ m)) analytical column placed in an oven at 80 °C. The injected sample amount was 10 μ L and the flow rate was set to 400 μ L.min⁻¹. A binary gradient was used (A: pure water, B: pure acetonitrile): from 2% to 25% of solvent B in 10 min, then up to 73% at 23.5 min and maintained at 73% for 4 min.

Mass detection was carried out with the instrument operating in negative polarity over mass ranges of m/z 350 – 2000 in MS mode and m/z 150–1100 in MS/MS mode. The spray voltage was set at 3.5 kV. Ions of interest (1038.21 m/z) were fragmented by collision-induced dissociation in the transfer cell of the instrument with argon used as the collision gas (collision energy of 50 V). Data acquisition was carried out using MassLynx software (V4.1).

Acknowledgements

This work was funded by the French National Research Agency (ANR Project CARBUNIVAX, ANR-15-CE07-0019-01). Authors thank N. Monties and G. Cioci from the PICT-ICEO screening facility from TBI for providing access to analytical and protein purification facilities. D. Lefebvre is acknowledged for his precious contribution to the optimization and production of mutants during his Master internship. The authors also thank F. Bellvert and H. Barbier for granting access to HRMS facility of the MetaToul platform of the Toulouse Biotechnology Institute (Toulouse, France), and performing the product analysis. Financial support from the TGIR- RMN-THC Fr3050 CNRS (Gif-sur-Yvette) for the NMR experiments is gratefully acknowledged as well as N. Morellet for helpful advice on NMR experiments. This work was granted access to the HPC resources on the TGCC-Curie, CINES-Occigen supercomputers and the Computing mesocenter of Région Midi-Pyrénées (CALMIP, Toulouse, France).

Supplementary Information

Supporting Methods

MMGBSA calculation

In order to estimate per residue enthalpy contribution to the enzyme: pentasaccharide free energy binding, the Molecular Mechanics (MM) energies combined with the Generalized Born (GB) and Surface Area (SA) continuum solvation calculation were carried out on enzyme-pentasaccharide complexes, using MMPBSA.py.MPI module of AMBER program. The generalized Born calculation was performed with the model developed by the Case group (Onufriev et al. 2004) using `igb=5` control command and a salt concentration of 0.1 M. The surface area for the nonpolar solvation term was calculated using Linear Combination of Pairwise Overlaps algorithm (Weiser et al. 1999), and the surface tension was set to default value *i.e.* 0.0072 kcal/mol·Å².

Shannon entropy calculation

Position-dependent amino acid residue variation in available GH70 sequences aligned by CLUSTALW, were analyzed using the Shannon information entropy measure (H_x), calculated using our in house SEQUESTER software (Daudé et al. 2013b).

Rosetta Design protocol

The Rosetta python script `molefile_to_params.py` was used to convert mol2 files to Rosetta parameter files and to generate ligands (sucrose and pentasaccharides) coordinates fitting the Rosetta atom names. Before the design, the enzyme and pentasaccharide interactions were optimized by a gradient-based minimization without restraints on the ligands. After that, 3

cycles of sequence design/packing and minimization were performed. Throughout the packing steps, the soft-repulsive force field was used, the input structure coordinates was included in the rotamer set, and the sub-rotamer Chi1 and Chi2 were taken into account. During all the computational design protocol steps (minimization/design/packing) the flexibility of the backbone and the side chain dihedrals of designable/repackable region was allowed, while the catalytic residues were prevented from being repacked or minimized.

Output Sequences filtering

In order to compare the parental enzyme and the designed sequences scores, the parental enzyme: pentasaccharide complexes (starting points of the computational design) were scored using the same steps of interface optimization and 3 cycles of repack minimization except for the catalytic residues. Then, the first sequences selection was performed by comparing the total and enzyme-ligand interface scores of the mutants to parental enzyme total and interface scores (Figure S2). The second round of data processing was the enzyme:sucrose complexes scoring, after the docking of the sucrose in the active site of the mutants having total and interface scores less than the parental enzyme. The docking was carried out by tacking the coordinates of the glucosyl moiety of the pentasaccharides for sucrose glucosyl unit. Afterwards the enzymes:sucrose complexes were scored by Rosetta by the same protocol used for the parental enzyme:pentasaccharides and described above. The resulting sequences were then clustered by the *Decrease Redundancy* module of *ExPASy* webserver (Gasteiger et al. 2003). The best score sequences of each cluster was selected for experimental testing (Table S1).

Supporting Figures

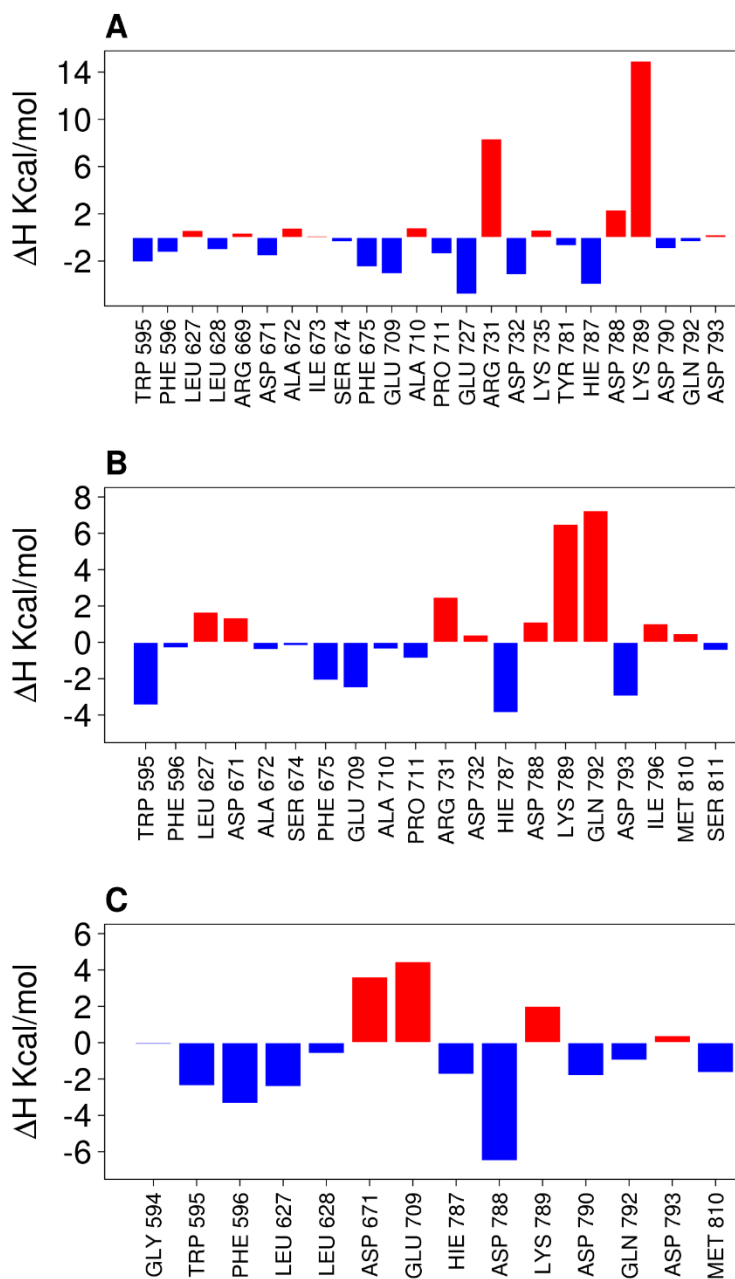


Figure S1: Graphics represent the amino acid residue enthalpy contribution to the free energy binding of the BRS-B $\Delta 2$ and ABC'[E(1 \rightarrow 4)]D' (*S. flexneri* 1a/1b) (A), AB[E(1 \rightarrow 4)]C'D' (*S. flexneri* 2a) (B) and E(1 \rightarrow 3)ABC'D' (*S. flexneri* 3a) (C) complexes. The enthalpies were calculated from the redesign starting point of each system, the negative and positive values are represented by blue and red color.

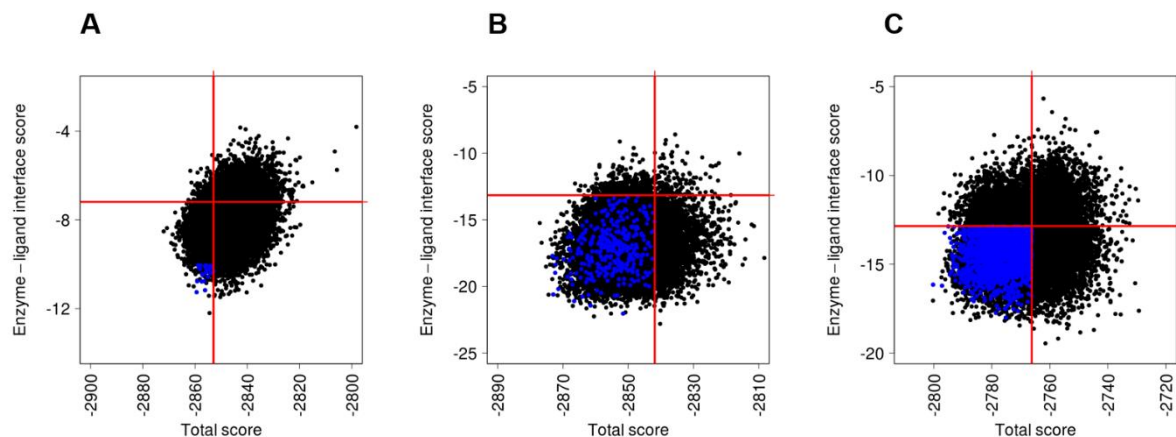


Figure S2: Graphics represent, the total score (x-axis) against the enzyme:pentasaccharide interface score (y-axis) of the BRS-B:ABC'[E(1→4)]D' (A), BRS-B:AB[E(1→4)]C'D' (B) and BRS-B:[E(1→3)]ABC'D' (C) complexes represented in black solid circles and their enzyme:sucrose complexes from the first selection described in the text are illustrated by blue solid circles. The starting parental complex (BRS-B Δ 2) total score and enzyme:pentasaccharide interface score are illustrated by the red lines.

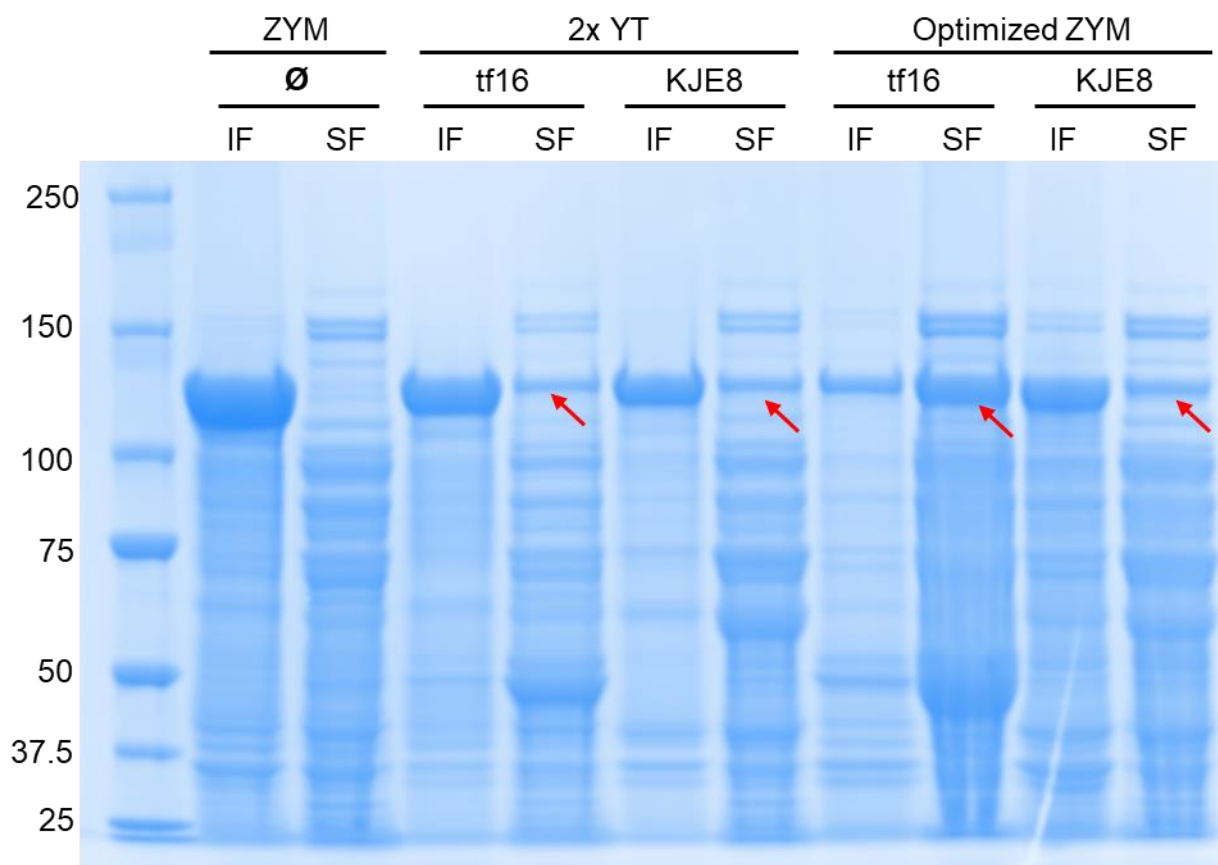


Figure S3: Optimization of the soluble expression of mutant M21. Size marker is present in lane 1 (sizes indicated in kDa). Using ZYM medium and without chaperone proteins (\emptyset), all the enzyme was produced in the insoluble fraction (IF, lane 2), no band was detected at the expected size of 132 kDa in the soluble fraction (SF, lane 3). When chaperone proteins (tf16 or KJE8) were co-expressed with M21 in 2x YT medium (lane 4 to 7), a band was detected at the expected size in soluble fractions (red arrows). The size of the band was further increased after optimization of the ZYM medium (lanes 8 to 11), especially when using tf16 chaperone proteins in the present case.

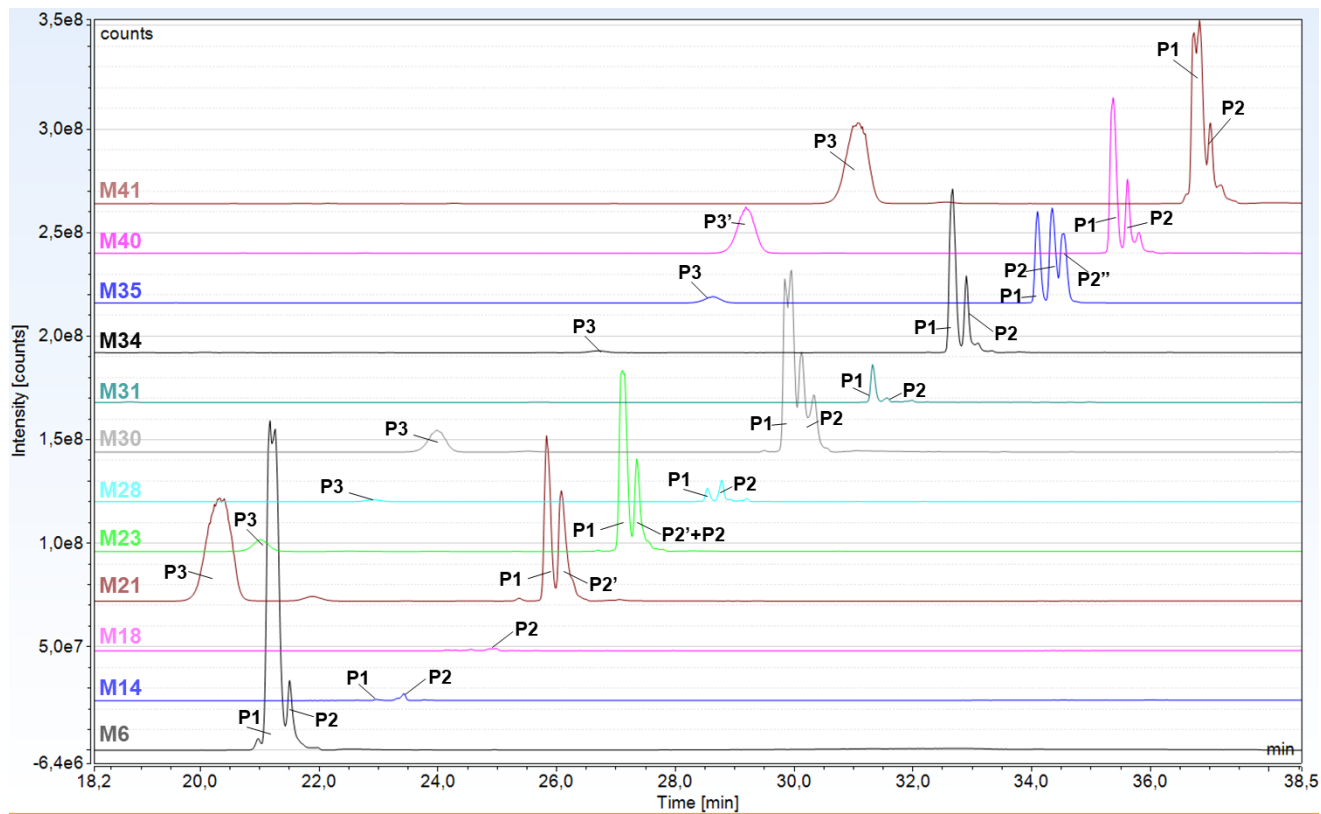


Figure S4: Comparison of product profiles obtained with BRS-B $\Delta 2$ mutants M6, M14, M18, M21, M23, M28, M30, M31, M34, M35, M40 and M41, after 16 h of reaction in presence of sucrose (1 M) and tetrasaccharide **ABC'D'** (50 mM) at pH 5.75 and 30°C. Mass spectrum (represented with time and signal shifted for clarity purpose) corresponding to ions of m/z 1057-1064 after separation by HPLC. **P1** ($t_R = 21.6$ min), **P2** ($t_R = 21.9$ min), **P2'** ($t_R = 21.9$ min), **P2''** ($t_R = 22.4$ min), **P3** ($t_R = 16.3$ min) and **P3'** ($t_R = 15.3$ min): products of the enzymatic glucosylation. A shift in molecular mass by 162 Da, characteristic of the glucosylation, was observed for these products.

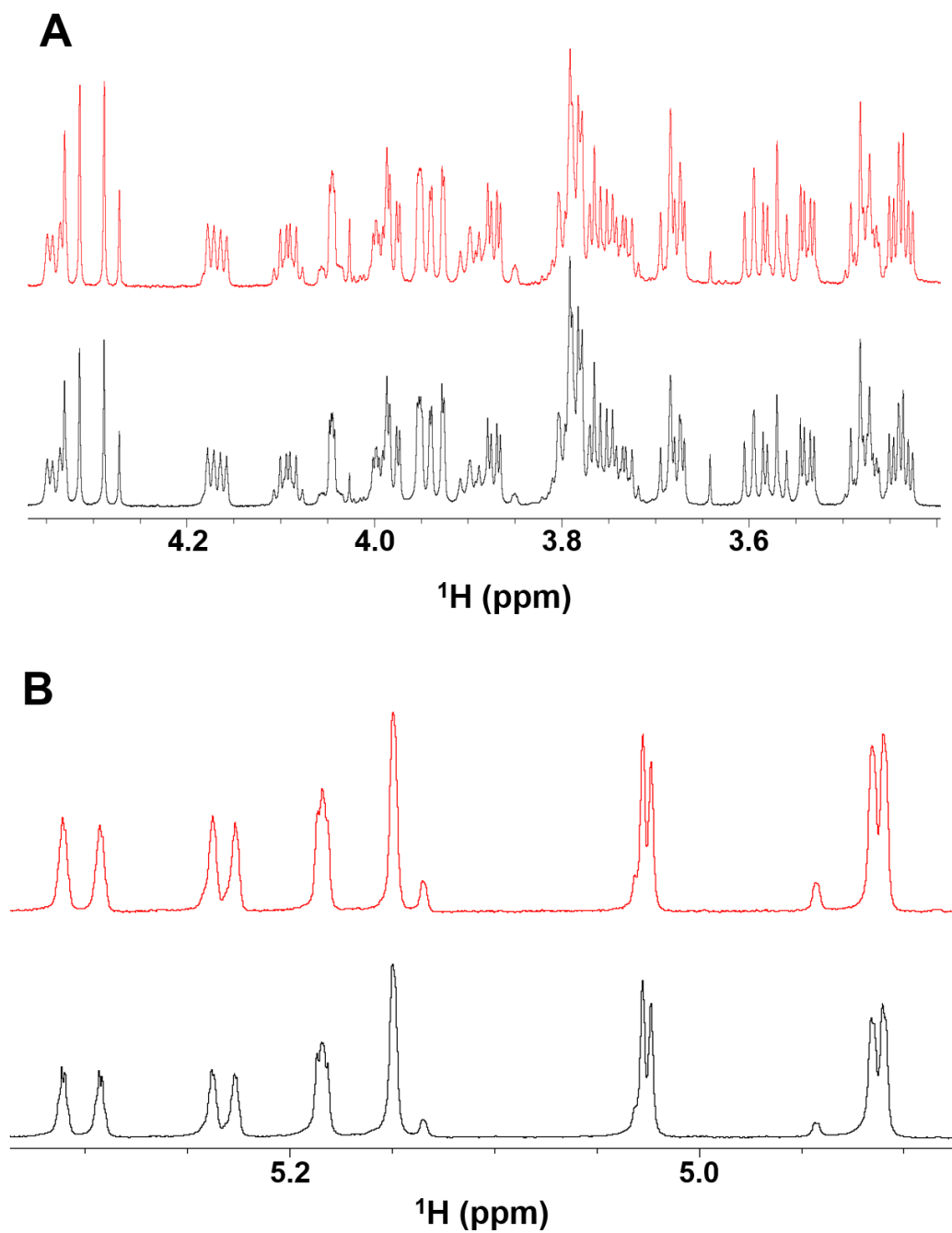


Figure S5: 1D spectra of the H2, H3, H4 and H5 region (A) and the anomeric region (B) of the pentasaccharides **P2'** obtained by the F2163G of ΔN_{123} -GBD-CD2 (black) and by the M21 of BRS-B $\Delta 2$ at 25°C in D_2O .

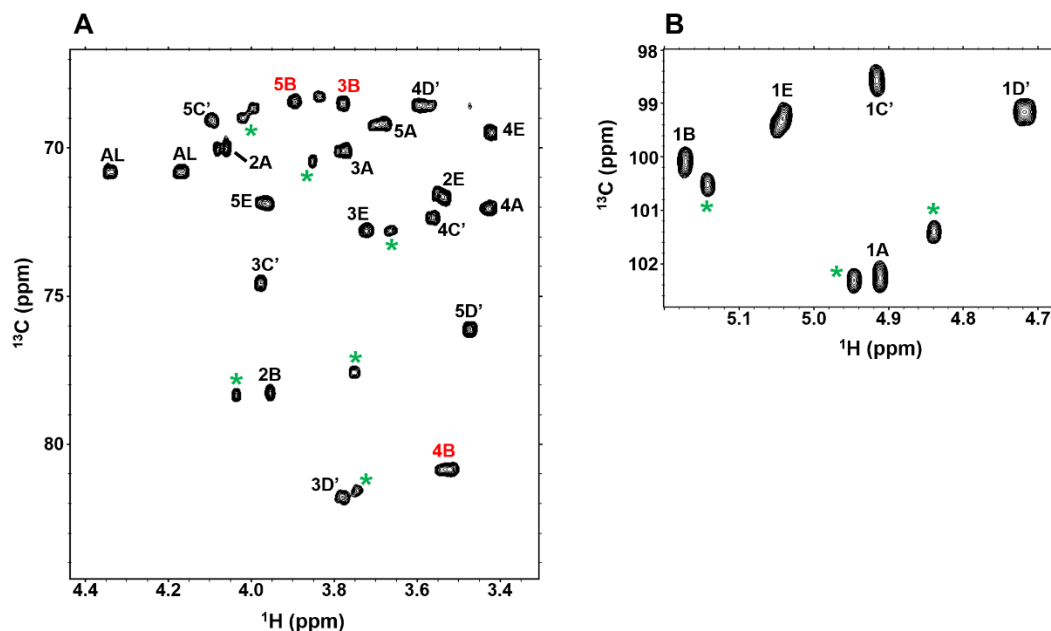


Figure S6: ^1H - ^{13}C HSQC spectra of the C2, C3, C4 and C5 resonances (A) and the anomeric region (B) of the mixture **P3** and its dechloroacetyled form. The shifted resonances of the **P3**, compared to the tetrasaccharide, were labeled in red. The peaks of the dechloroacetyled form **ABCD'** are labelled by green stars. The spectrum was acquired at 950MHz.

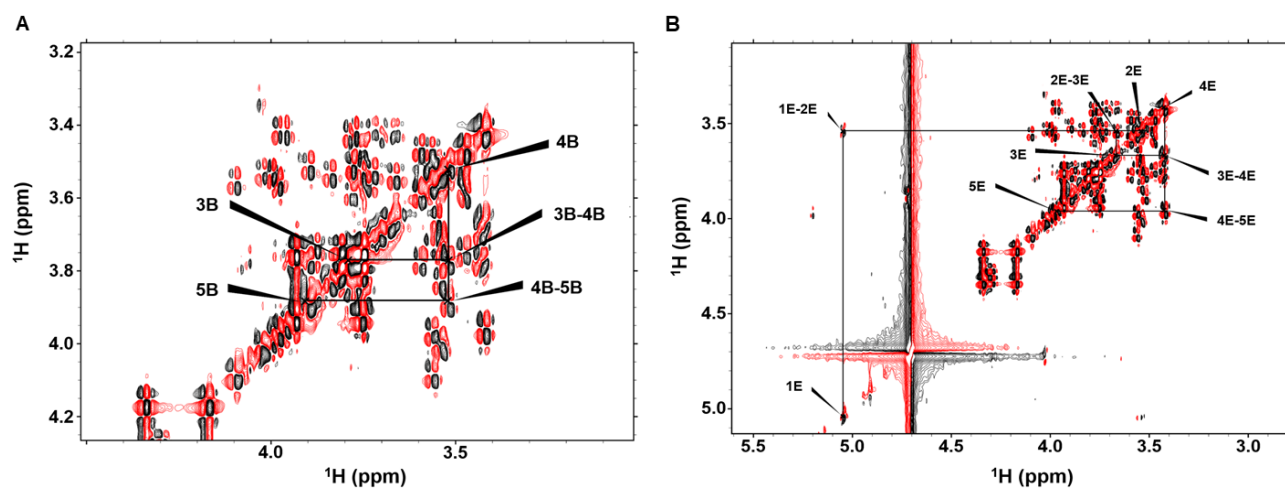


Figure S7: ^1H - ^1H QDF-COSY spectra at 950MHz of the mixture of the **P3** and its dechloroacetyled form. Positive and negative peaks are showed in black and red respectively. The B (panel A) and E (panel B) units belonging to the **P3** connectivities are plotted and the corresponding peaks are labeled.

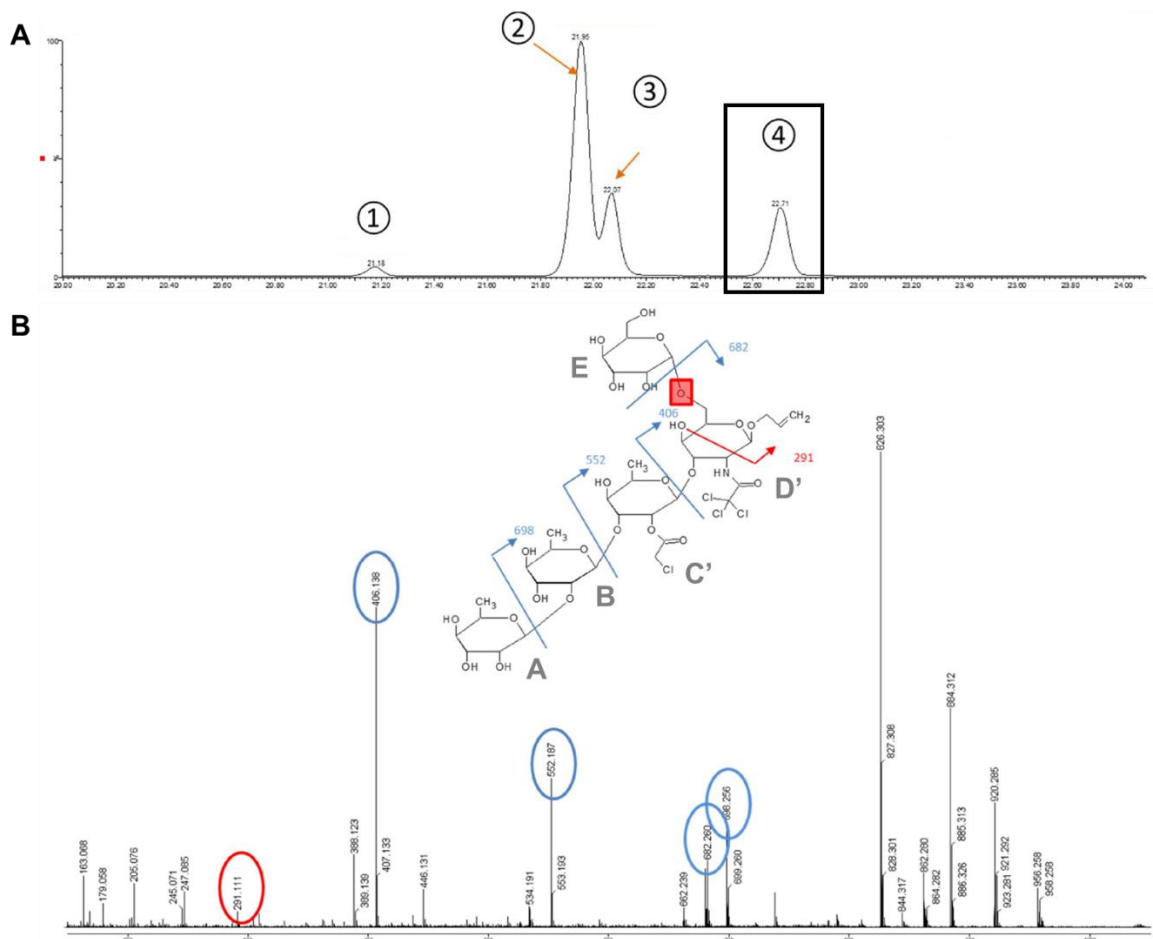


Figure S8: **A/** TOF MS ES- spectrum of M35 enzymatic reaction obtained for mass ranges of 1037.828-1043.597 after separation by HPLC. Black frame: peak ($t_R = 22.71$ min) corresponding to the ion selected for re-fragmentation by MS/MS **B/** MS/MS spectrum obtained after re-fragmentation of the [ABC'D'] ion of m/z 1038.21 at $t_R = 22.71$ min (peak ④ in panel A). In blue: specific ions obtained after re-fragmentation of the ion between the moieties **A/B**, **B/C'**, **C'/D'** and **E/D'**, respectively. In red: specific re-fragmentation of the **D'** moiety. Fragmentations are represented on the structure of ABC'[E(1→6)]D' with arrows. Red frame: position glycosylated on the **D'** moiety.

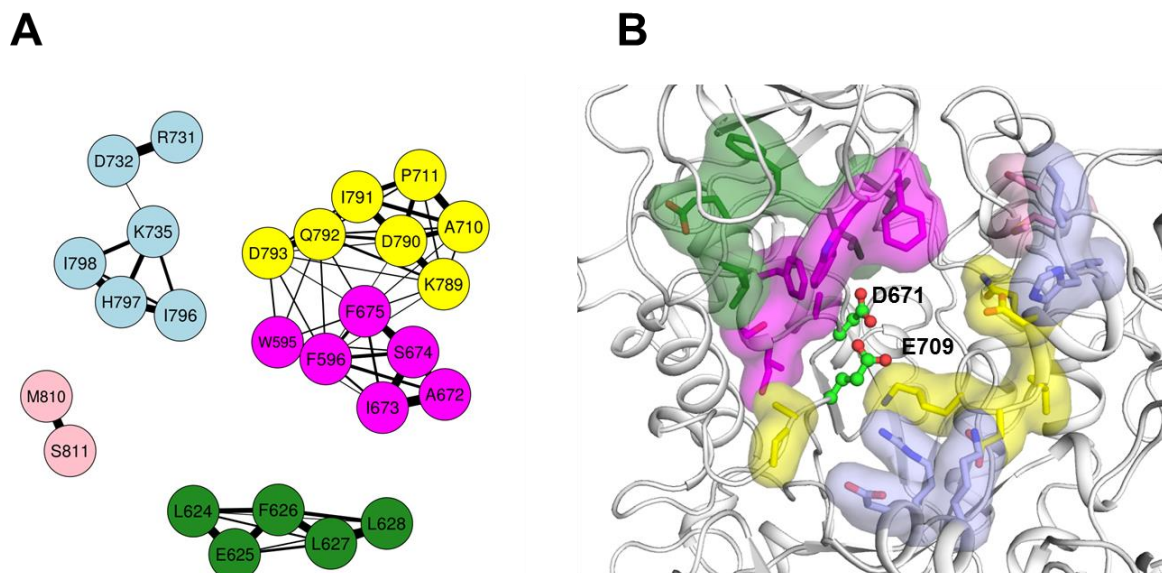


Figure S9: Comparison of network communities (left) and structural representation (right) derived from BRS-B $\Delta 2$ model. Left: Mutated residues network presented by spring layout using Fruchterman-Reingold algorithm from BRS-B $\Delta 2$ model distance matrixes. Right: View of Active site residues clusters of BRS-B $\Delta 2$. The 5 communities from the spin-glass algorithm are presented by I (magenta), II (dark green), III (yellow), IV (light blue) and V (pink).

Supporting Tables

Table S1: The table shows the Shannon entropy values for the redesignable positions, the conserved L627 and A672 are labelled in red color.

Residue	$H(x)$
G594	0.60
W595	0.53
F596	0.18
L624	0.3
E625	0.15
F626	0.38
L627	0
L628	0.10
A672	0

I673	0.15
S674	0.02
F675	0.06
A710	0.33
P711	0.33
R731	0.04
D732	0.4
K735	0.37
K789	0.28
D790	0.26

I791	0.18
Q792	0.09
D793	0.36
I796	0.39
H797	0.55
I798	0.10
M810	0.32
S811	0.57

Table S2. Summary of the mutations introduced in mutants of each group at the 27 re-designable residues. A blank indicates that the amino acid of the parental wild-type enzyme was conserved.

Targeted positions		594	595	596	624	625	626	627	628	672	673	674	675	710	711	731	732	735	789	790	791	792	793	796	797	798	810	811	Number of mutations	
Wild-type amino acid		G	W	F	L	E	F	L	L	A	I	S	F	A	P	R	D	K	K	D	I	Q	D	I	H	I	M	S		
Group I	M1		E			L	H			G		D	R			I		L	M	N	L	V							12	
	M2		L	Y		Y	W	M		G	L		Y			I		L	L	L	L	V							14	
	M3		A		P	K	M			G	M	N				L		L	L	Q	L	I							13	
	M4		T	Y	P	I				G	L	T				I		L	L	M	L	I							13	
	M5		V			K	L			G	L		Y			I		L	L	L	L	V							12	
	M6		A			I	K	V		G	L	A				I		L	L	L	L	V							13	
	M7		A		T	L	L			G	L	A				I		L	L	L	L	V							12	
	M8		T		P	L				G	L	M	L			I		L	L	L	L	V							13	
	M9		T	Y		Y	W			G	M		Y			I		L	M	N	L	V							13	
	M10		T		T	K	V			G	L	I				V		L	L	L	L	V							13	
	M11		A		M	K	H			G	I	A	F			I		L	L	L	L	I							14	
	M12		A		I	L	Y			G	L	N	Y			F		V	L	N	L	I							13	
	M13		H		T	I	Y			G	L	F	L			I		L	M	N	L	V							14	
	M14		M	M	V	K			M	G	M	W	L			V		L	L	N	L	V							15	
	M15		S			K	A			G	L		T			V		L	L	L	L	V							13	
	M16		S	M	V	L				G	L	N	T			I		L	L	L	N	L	V						14	
Group II	M17		L		K	V				V	D	W			L		W	I			L								10	
	M18		L		I	R	Y			P	D	W			I		L	I		N	V	L							13	
	M19		L		I	K				V	H				V		L	V	N	V	L								11	
	M20		M	M							M			D	S	L	L		I			V	E				W	T	13	
	M21		L								D			D	S	W	L	L	L				E	V			W	T	11	
	M22		M								M			D	S	W	L	L				L	E	V			W	T	12	
	M23		P								M				D	L	L	L				L	E	T			D	T	10	
	M24		L										W		D	M	L	L		L		N	L	W			A	T	11	
	M25		L								A	W	D	D	L	I	M	L		L			E	L			T	T	12	
	M26		L								A	W	D	D	T	I	M	L		L			E	V			W	T	12	
	M27		V								A	W	D		S	I	L	L		L			E	L			T	T	12	
	M28		E								D				S	I	L	L		L		W	E	W			Y		10	
	M29		L									D		W	D	I	M	L		L		W	E	V			W	T	12	
	M30		L											W		L	M	Q		V		V	E	T			M	T	11	
	M31		L											W		L	M	E		I		I	L	W			V	T	11	
	M32		E	M										W	D	T	M	L		Q		L	E	W			A	T	13	
	M33		M	M										W	D	T	I	E		M		L	Q	W			A	T	13	
	M34		L											W		L	M	A		V		L	E	V			P	T	11	
Group III	M35		A					M										A		V	V				G	T	H	T	9	
	M36		L						M									L		L	Y	I	V	G	L			T	10	
	M37		L	L														L		L	Y	V	V	S	L			T	10	
	M38		D					M										A		L	F	I	V	G	T				10	
	M39		L	T														A		L	Y	V	G	L				T	10	
	M40		D															V		T	I	V	S	L		S		T	11	
	M41		L	M														L		L	E	V	V	S	L			T	9	
	M42		L	M														M		V	L	L	V	S	L			T	10	
	M43		M	T																L	Y	I	V	G	L			T	10	
	M44		L	N															M		L	Y	I	V	S	L		T	11	
	M45		F	A																	L	Y	I	W	L			T	9	
	M46		F	A																	L	W	I	V	G	L			T	9
	M47		L	T																L	W	L	G	L		F		T	10	
	M48		M	V					M										M	N	L	F	I	V	S	L		A	12	
	M49		A	H															H			L	L	V	S	L	S		T	10

Table S3. The five plasmids encoding various sets of chaperone proteins tested in this study. Three chaperone sets co-expressed with BRS-B $\Delta 2$ mutants enabled soluble expression.

Plasmid	Chaperone proteins encoded	Soluble expression detected
pGRO7	groES, groEL	No
pG-Tf2	groES, groEL, tig	No
pTf16	tig	Yes
pKJE7	dnaK, dnaJ, grpE	Yes
pG-KJE8	dnaK, dnaJ, grpE, groES, groEL	Yes

CONCLUSIONS AND PERSPECTIVES

General conclusion

Enzymatic glycosylation, using either native or engineered glycoenzymes, and combined with chemical synthesis in chemo-enzymatic pathways, has the potential to give access to a new diversity of carbohydrate molecules that could be used in various applications. Such carbohydrates indeed can present a wide range of properties of interest for applications in the food or feed industries, cosmetics, pharmaceuticals, etc. In particular for the latter, cell-surface carbohydrates such as mammalian glycans, Gram negative bacteria lipopolysaccharides or capsular polysaccharides are gaining interest as they could open the way to new diagnostic tools and curative or preventive treatments. For example, cell-surface carbohydrates of Gram negative bacteria such as the polysaccharide components of the lipopolysaccharides often display a large structural diversity that prevents development of efficient anti-bacterial vaccines. New vaccines would greatly benefit from access to these polysaccharides (O-antigens), but their development remains limited mostly due to synthesis difficulties. In the context of this challenging field of research, the group of Laurence A. Mulard at Pasteur Institute has investigated novel routes to access haptens representative of the Gram negative bacteria *Shigella flexneri* -responsible for shigellosis, a form of bacillary dysentery. This work lies the foundation for access to a vaccine for low and middle income countries, which suffer heavy burden from the disease.

Upon 10 years of collaboration with our team at TBI and successive projects of increasing complexity, a new chapter unfolded with the ANR project “CarbUniVax” (2015-2019), in which was integrated my PhD thesis. The main objective of the work aimed first at designing and synthesizing a conveniently protected tetrasaccharide (**ABC'D'**) compatible with subsequent elongation into longer haptens, and that mimicked the linear tetrasaccharide backbone present in the LPS repeating unit of *Shigella flexneri*. Next, my thesis aimed at identifying and engineering enzymes able to selectively glucosylate **ABC'D'** to produce a variety of pentasaccharides, so-called glyco-bricks, characteristic of prevalent serotypes of *S. flexneri* 1b, 2a and 3a targeted in priority in the project as they are currently responsible for the most cases of shigellosis. Access to other less representative serotypes still constitute a secondary objective as they might very well become prevalent in the future. The ultimate goal foreseen in the project is the large-scale production of these serotype-representative glyco-bricks, which could be linked to a carrier protein, in order to further evaluate their biological effect and potential interest for broad coverage vaccine development.

In order to access these glyco-bricks, a highly convergent synthesis strategy was devised that takes advantage of the combination of chemical synthesis of the tetrasaccharide backbone and its one-step enzymatic glucosylation. The first challenge –for our colleagues at Pasteur Institute– was to design the protected tetrasaccharide **ABC'D'**. Several constraints were taken

into account when choosing the protecting groups, which had to (i) be compatible with subsequent chemical oligomerization of the glycobrick but also (ii) be compatible in size with the enzyme active site, therefore carrying small protecting groups, and finally (iii) yield a stable and soluble molecule in aqueous solvent compatible with enzyme usage. Sufficient quantities of the molecule had to be synthesized for method development and enzyme screening.

The very challenging chemical route comprised a total of 23 steps, and represented one of the main roadblocks of this project. Indeed, the molecule proved to be unstable at pH higher than 5 and was synthesized only in mg-amounts due to difficulties in scaling-up the reaction. This had to be taken into account when considering enzyme usage, and enzymatic assays had thus to be miniaturized to limit the consumption of the acceptor molecule.

In the absence of natural enzymes reported to catalyze wanted reactions, we decided to turn to branching sucrases from the GH70 family. These enzymes are biocatalysts of choice to attempt this glucosylation as they are able naturally to branch dextran molecules with glucosyl units, they are active at pH 5 and they display a broad acceptor promiscuity. Owing to their very open active site which features several flexible loops, branching sucrases have been shown to tolerate a broad range of hydroxylated molecules as acceptor, often differing drastically from the structure of their innate acceptor substrate. Moreover, these enzymes utilize sucrose, a readily-available and cheap agro-resource as a glucosyl donor, perfectly suited for integration in a large-scale production process. For example this would be of particular interest in a process that would aim at developing vaccines for low and middle income countries at a reduced cost. Finally, unlike family-related glucansucrases, branching sucrases do not produce polymers, limiting presence of unwanted co-products that could lower the yields and render downstream processes more difficult.

Glucosylation of the non-natural lightly-protected tetrasaccharide **ABC'D'** using promiscuous native branching sucrases

Using 6 native branching sucrases of our collection, we were able to glucosylate **ABC'D'** without the need of enzyme engineering, and only few optimizations of the reaction conditions allowed to improve the yields. This allowed to isolate and identify a first mono-glucosylated product, **P1** pentasaccharide, representative of *Shigella flexneri* serotypes 4a/4b, and produced with a conversion yield increased up to 31% using the BRS-B Δ 1 enzyme. A second mono-glucosylated product, **P2**, was produced but unfortunately in too low amount to allow its isolation for structural characterization by NMR spectroscopy. Added to a previous work where the branching sucrose Δ N₁₂₃-GBD-CD2 was used for glucosylation of a non-natural disaccharide, this new result further expanded the range of non-natural carbohydrates recognized by branching sucrases. More generally it constituted another evidence of the glucosylation of exogenous acceptors pointing out again the potential of these enzymes as

glucosylation tools. This preliminary work revealed however some limitations in the use of native branching sucrases that allowed to access only two out of the eight possible glucosylation patterns. To that extent, active site engineering was considered to generate wider molecular diversity. Mutagenesis could also advantageously help to improve the glucosylation yield for the two glyco-bricks (**P1** and **P2**) already produced by the parental enzyme. The utilization of BRS mutants was thus envisioned in a second stage.

Evaluation of pre-existing engineered branching sucrases to redirect preferential production toward two distinct pentasaccharides

We decided to evaluate a library of 22 pre-existing mutants of a branching sucrose, ΔN_{123} -GBD-CD2, which were previously designed to accommodate and glucosylate bulky flavonoids. Mutations targeted the first and the second-shell amino acid residues of the acceptor subsites. They mainly focused on positions W2135 and F2136 or F2163 and L2166. Despite using a lower pH than the optimal one, the mutant library could be assayed in a 10 μ L scale miniaturized format designed for lowering consumption of the acceptor molecule. Interestingly, with the exception of one mutant, all of them were found able to glucosylate **ABC'D'**, redirecting preferential production towards **P1** or **P2** depending on the introduced mutations. Six mutants forming preferentially either **P1** or **P2** were selected for further investigation at a larger 50 μ L scale. Three mutants clearly showed distinct features. The double mutants W2135S-F2136L or W2135I-F2136C led to (i) new specificities with new products identified: **P3** pentasaccharide and an hexasaccharide **H1**, and (ii) a better production of **P1** compared to the parental enzyme. On the other hand, the single mutant F2163G allowed a better production of **P2** with an 8% acceptor conversion yield, and the corresponding pentasaccharide was identified as being characteristic of *S. flexneri* prevalent serotype 3a.

Mutants W2135S-F2136L and F2163G were characterized in more details at biochemical and biophysical levels. First, their kinetic parameters in sole presence of sucrose were determined, indicating a better affinity for sucrose compared to the parental enzyme but decreased maximum velocities. Stability of the mutants was assessed by determination of Melting temperatures (T_m) by Differential Scanning Fluorimetry, showing two peaks. Results showed that T_m value was lowered by 3°C in its first T_m in the presence of the F2163G mutation. Both mutants lost their second melting temperature suggesting a different unfolding behavior compared to the parental enzyme. Finally, in the presence of the natural acceptor, dextran, mutant F2163G was found to keep similar branching ratio as the parental enzyme with introduction of around 33% α 1,2 linkages determined by NMR analysis. On the other hand, mutant W2135S-F2136L revealed a decreased branching ratio with only 12% of α 1,2 linkages introduced, and a slightly different branched dextran profile as shown by HPAEC-PAD analyses.

In order to better understand the effects of these mutations, modelling studies were carried out and provided insights on the tetrasaccharide binding mode in the branching sucrose active site. Of note, the interaction networks of active site residues were established after graph analysis of the simulations carried out for the three mutants. Key residues and interactions favoring glucosylation of the tetrasaccharide at one or the other extremity were identified. In light of experimental results, these analyses helped to improve our knowledge on key molecular and dynamical determinants from the active site onto product specificity which could help to guide further their engineering. This also demonstrates the potential of introducing selected mutations in branching sucroses to control the regio-selectivity and enhance the range of accessible glucosylated molecules.

Computer-aided engineering of the branching sucrose BRS-B $\Delta 2$ to access a whole new diversity of regio-selective glucosylation patterns of the tetrasaccharide

Building upon this achievement which indicated that mutations into the active site could lead to changes in the product profiles and give access to new mono-glucosylated **ABC'D'** molecules, we added another layer of complexity by applying a computer-aided enzyme re-design approach to BRS-B $\Delta 2$, the best **P1** producing branching sucrose. Using a combination of molecular modelling and computational protein design techniques, a small library of 49 mutants was designed, compatible with limited availability of **ABC'D'** acceptor. Mutants contained between 9 and 15 mutations in their active site. Although very challenging, we managed to produce all the mutants in soluble form after optimization of the culture conditions. This required the use of chaperone proteins that helped to fold the enzymes and prevented their aggregation, as well as the optimization of the culture medium. Despite detrimental impact of mutations on sucrose recognition, 16 mutants out the 49 tested, conserved an activity on the donor, representing around 33% of positive hits, far more than earlier reports that used random or semi-rational mutagenesis (Vergès et al. 2015). Mutants were subsequently evaluated for glucosylation of **ABC'D'**, providing access to 6 mono-glucosylated products and several double-glucosylated products (preliminary data not shown). Some mutants remained quite specific for the production of one pentasaccharide while others produced up to 4 mono-glucosylated products. Among them, M21 was the best enzyme for the production of **P3**, even in comparison to mutant W2135S-F2136L from ΔN_{123} -GBD-CD2 branching sucrose. **P3** was isolated and further characterized by NMR, revealing a glucosylation on the tetrasaccharide inner chain **B** subunit, never achieved using a native branching sucrose. The exact determinants responsible for re-orienting the tetrasaccharide in a specific binding mode compatible for branching within the chain could not be elucidated and probably result from synergetic effects of the 11 different mutations.

On the same note, M35 and M40 displayed a novel specificity, and while not yet fully understood on a molecular level, they allowed to access newly identified pentasaccharides, **P2''** and **P3'**, respectively. Produced in too low amounts, their characterization required a more sensitive technique than NMR spectroscopy that would avoid large scale production and purification of the molecule. MS/MS was then used to elucidate product structures, revealing more particularly that **P3'** was representative of *S. flexneri* serotype 5a. Further work is still necessary to understand in more depth the role of the mutations. Deconvolution could help to identify the critical ones and may lead to (i) remove mutations that were deleterious for enzyme soluble expression or loss of activity on sucrose and (ii) improve specific production of a given pentasaccharide, but given the high number of mutations and combinations, it would be quite challenging and would require higher quantities of acceptor molecule.

This unprecedented engineering of the recently discovered branching sucrase enzymes lays the ground for the construction of a glucosylation toolbox that could be used as a starting point for glucodiversification of various molecules, including non-natural carbohydrates. Of course, further engineering would still be required to improve enzyme activity and conversion yields or enzyme expression to make the mutants suited for larger-scale processes.

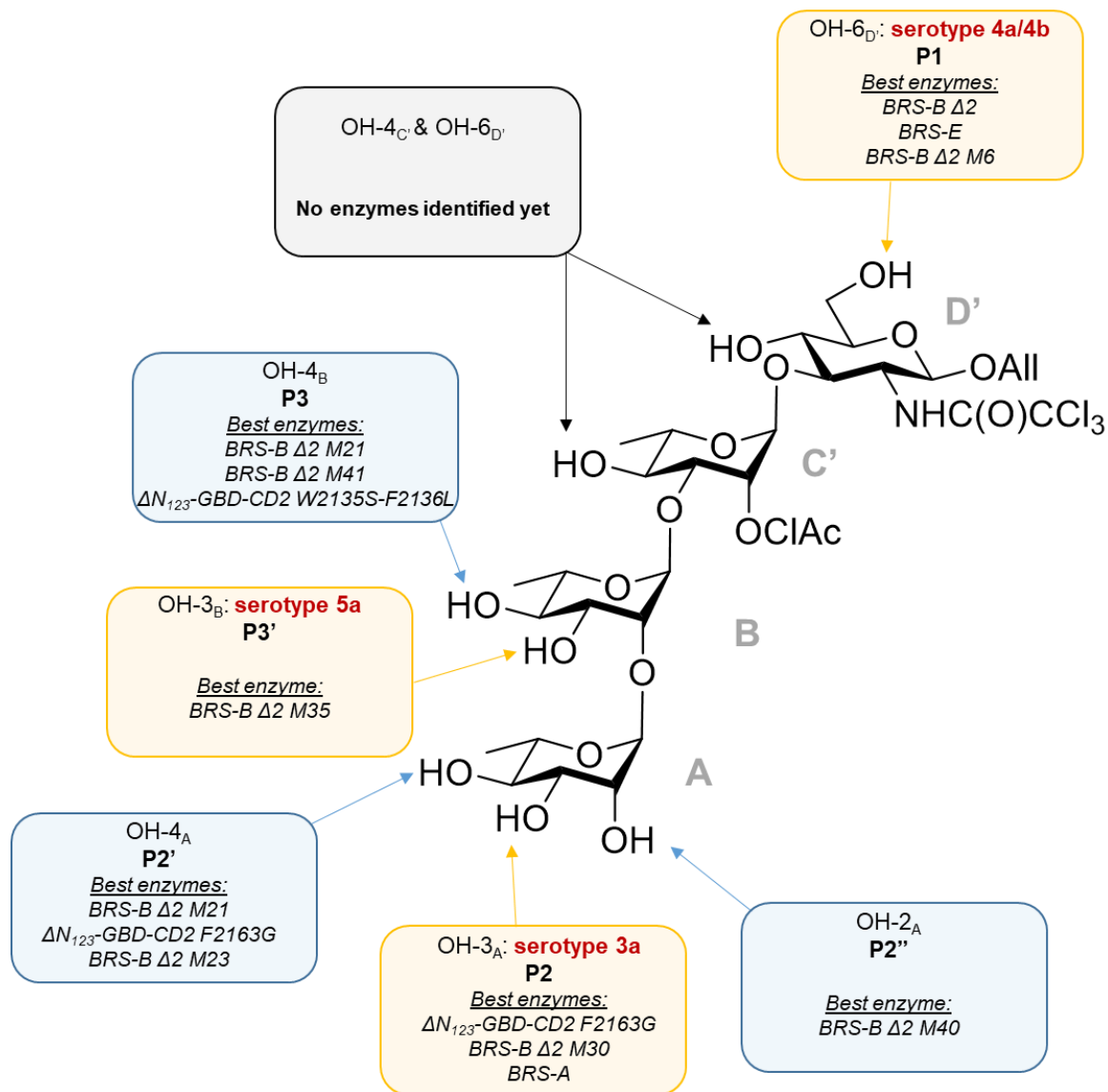


Figure 1: Summary of the different mono-glucosylated products obtained using native and engineered branching sucrases. The best enzymes identified for their synthesis are indicated for each targeted hydroxyl group. In red are indicated the glucosylation patterns characteristic of *S. flexneri* serotypes.

Perspectives

Altogether, our results set the ground for the development of chemical and enzyme-based processes working in combination to achieve the synthesis of carbohydrates relevant for vaccine development. Thanks to these synergistic routes, we were able to access for the first time 6 distinct mono-glucosylated products out of the 8 possible ones from **ABC'D'**, using both native and engineered branching sucrases. Three of these products turned out to be representative of *S. flexneri* serotypes 4a/4b, 3a/X, and 5a (Figure 1). Of course, product yields were often low and the enzymes not always suited for a larger scale process, requiring further optimization, but they still constitute a good starting point to address the following objectives:

- 1) To tackle the tetrasaccharide synthesis challenges for scaling-up its production, allowing to improve the reaction conditions and enabling screening of new mutants
- 2) To develop an appropriate methodology for higher screening throughput of BRS enzymes
- 3) To obtain crystallographic structures of BRS-B $\Delta 2$ and its mutants, if possible in complex with the products
- 4) To validate the concept proposed by Pasteur Institute by producing glyco-bricks at the gram-scale for chemical elongation

Efforts to address the synthesis challenges

Further optimization of the enzymes was limited due to the low available amount of tetrasaccharide acceptor. Each step of the tetrasaccharide chemical synthesis is being carefully optimized at Pasteur Institute to ease the scale-up of the molecule. Laboratory scale-up was difficult to control, especially the final deprotection step of the synthesis. Moreover, unexpected low stability of the chloroacetyl protecting group at pH > 5 was another obstacle to the synthesis of **ABC'D'**. Alternative synthetic routes are being explored at the Pasteur Institute that will hopefully lead to higher production of the tetrasaccharide.

This will enable to further improve the enzyme reaction conditions by thoroughly setting up a design of experiments that could not be done so far. Various parameters could be considered in the optimization such as the concentration of **ABC'D'** acceptor and sucrose donor, the enzyme quantity, the optimal pH and temperature, as well as the duration of the reaction. Moreover, kinetic parameters could be determined in order to better understand the enzyme behavior in presence of the acceptor. This would help to determine the best conditions for enzymatic reaction, required for a scale-up, especially if observing inhibitory effects of the acceptor.

Construction of a versatile enzymatic toolbox for the glucosylation of exogenous acceptors

In *S. flexneri*, at least 15 serotypes are currently described with several new (sub)serotypes emerging and displaying new modifications of the repeating linear tetrasaccharide backbone constituting the LPS such as introduction of phosphoethanolamine and O-acetyl groups or new glucosylation patterns. In particular, serotypes defined by either two distinct glucosylations (serotypes 1d or 2b) or one di-glucosylation (serotypes 7a and 7b) are already described, and new mono-glucosylation patterns might emerge in Nature after phage-mediated transfer of genes encoding new glycosyltransferases. Therefore, the additional glyco-bricks obtained with either atypical mono-glucosylation at OH-4_A and OH-4_B or di-glucosylations could very well be of interest in future episodes of emerging infections. The enzymatic toolbox developed in this study could thus allow to access to these structures and be used for vaccine formulation. This is why it would be quite interesting to determine the structures of all the remaining glucosylated products using the preliminary MS/MS method described in chapter IV. However, we have to keep in mind that yields were very low and except for **P1**, **P2** and **P3**, the products were often detected in trace quantities. The enzymes responsible for the synthesis of the novel products could however be further optimized to fulfil several objectives: improve the soluble expression, improve the yields, increase the activity toward sucrose donor, and achieve more specific synthesis of a given product. The latter could be the most challenging, due to the large and exposed active site of branching sucrases, surrounded by highly flexible loops, which allow the acceptor to bind in several productive modes, leading to various products. Larger libraries of mutants and high-throughput screening methods would then be needed to explore a larger diversity.

Branching sucrases reported herein display original specificity that could be opportunistically utilized to glucosylate additional acceptors, including carbohydrates representative of other bacteria serotypes featuring α -glucosyl on their cell-surface polysaccharides, such as *Candida difficile*, *E. coli* O25, *Enterococci faecalis*, *Klebsiella pneumoniae* K2, *Moraxella catarrhalis*, *Pseudomonas aeruginosa* O11 etc. (Micoli et al. 2018b)

Engineering a strain for the development of a high-throughput selection assay

As already stated, additional enzyme engineering is a prerequisite to improve further enzyme expression and catalytic performances. However, the low/medium LC-MS throughput-screening methods used herein are not appropriate for large libraries of mutants. New methodologies are emerging such as microfluidics and capillary analysis in Ultra High Performance Liquid Chromatography that could allow scaling down reaction volumes to a few nanoliters, avoiding consumption of rare acceptors, as well as increasing the analysis

throughput. However, pre-screening the library of mutants for active enzymes would still be a bottleneck in the workflow.

To address this limitation, we started the development of an engineered strain suited for the screening of sucrose-utilizing enzymes. We planned to integrate in the genome of *E. coli* BL21 strains, used for enzyme expression, a sucrose permease gene from the sucrose catabolic operon (*cscAKBR*) of *E. coli* W strain, based on prior work of Claudia E. Vickers group (Sabri et al. 2013b, 2013a). This operon contains initially the *cscB* gene, encoding for a sucrose permease, *cscA* (invertase), *cscK* (fructokinase) and *cscR* (*csc*-specific repressor). The rationale of this integration was to develop a growth selective assay that is described hereafter. After transformation into the modified *E. coli* strain, a library of variants will be plated onto minimal medium containing sucrose as the sole carbon source. Clones that integrated an active enzyme on sucrose will be able hydrolyze it into fructose and glucose, which will be then used in central carbon metabolism for colony growth. Conversely, inactive variant, deprived from a carbon source and not able to metabolize sucrose will not grow (Figure 2). Previous medium-throughput screening assays on solid medium developed in the team relied on colorimetric discrimination of active and inactive clones and were only compatible with the highly sucrose-active glucansucrases (Champion et al. 2010). This selection method offers the advantage to select only active variants, is compatible with branching sucrases as well, and could be automated using automated colony picking, enabling to access a high-throughput selection assay.

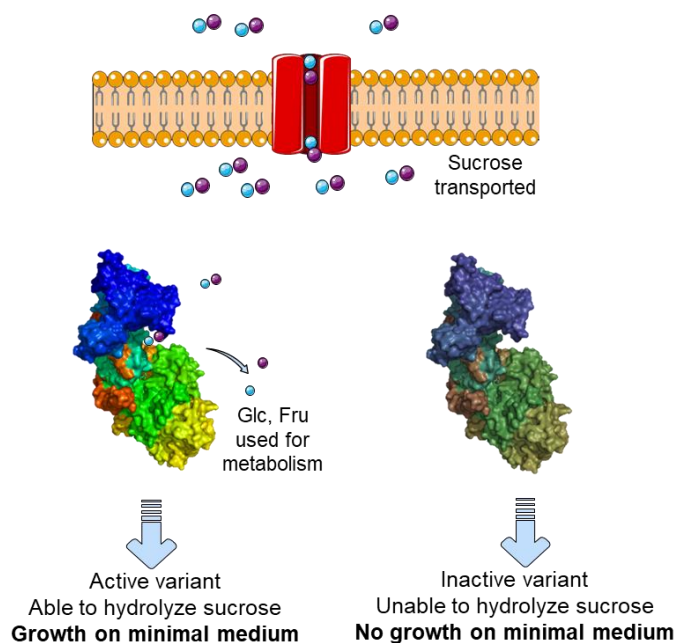


Figure 2: The new strain able to express a sucrose transporter would be suited for selection of sucrose-utilizing enzymes. After transformation of a library of variants, only clones having integrated an active enzyme on sucrose will be able to grow on a minimal medium supplemented in sucrose.

Preliminary work was performed, the sequence of part of the operon, *cscAKB*, was retrieved from the pCSCX plasmid available in Addgene public bank. The ribosome binding site (RBS) and the gene *cscB* encoding the sucrose permease were specifically amplified, as well as the promoter region of the operon. The “promoter-RBS-*cscB*” construct was assembled with multi-fragment Gibson assembly into pKIKO (knock-in / knock-out) plasmid developed by Vickers’ group. This plasmid carries homologous sequences allowing integration of genes into *E. coli* genome at the non-essential *ars* locus encoding for an arsenite transporter, along with a gene for kanamycin resistance used for selection. A mutation, Q353H, described to enhance permease activity and therefore sucrose transport (Jahreis et al. 2002) was also performed on *cscB* gene. Both constructs (transporter and mutated transporter) were linearized and transformed into *E. coli* BL21, previously transformed with the temperature-sensitive plasmid pKD46 containing λ *red* sequence promoting the homologous recombination system. Our first attempts did not allow to integrate the sequence and this work is to be continued. In parallel, another strategy for expression of the transporter on a plasmid was started. This work could ultimately lead to the development of a strain usable for the accelerated selection of sucrose-utilizing enzymes, such as amylosucrase from GH13, and glucansucrases, as well as branching sucrases from GH70.

Gaining structural insight on BRS-B Δ 2 branching sucrose to improve design accuracy and guide further engineering studies

Only one crystallographic structure of a branching sucrose (Δ N₁₂₃-GBD-CD2) is available to date. Thanks to this structure, it was possible to construct a 3D-model of BRS-B Δ 2 (49% identity with Δ N₁₂₃-GBD-CD2) by comparative modelling. However, the construction of a model presents many risks, as biases could be introduced due to structural rearrangements, mutation impact on 3D organization and dynamics, environment effect, oligomerization etc. In the case of branching sucrases, we investigated the impact of loop dynamics on active site topology of branching sucrases (Ben Imeddourene et al. 2018), suggesting that it could be at the origin of the large acceptor promiscuity. As seen in this study, it is difficult to re-direct the glycosylation towards one specific hydroxyl on the basis of current structural knowledge and information. Therefore, preliminary attempts were made to obtain the X-ray structure of BRS-B Δ 2.

Obtaining this structure would allow to better understand the molecular determinants involved in the natural α 1,3 linkage specificity when dextran is used as an acceptor and overall improve our knowledge on this recently discovered sub-class of enzymes. It could also help to guide future engineering projects involving this enzyme, which has already proven to be versatile and a very powerful tool for glycosylation of exogenous acceptors. However, crystallization of GH70 enzymes remains difficult, due to their size and flexibility, in particular at their N- and C-terminal extremities. The original purpose of the BRS-B Δ 2 truncation was to remove

disordered regions at N-terminal extremity, which was done in the team by M. Claverie (unpublished work, 2014). However, no crystals were obtained.

Continuing this work, we showed in this study both by molecular modelling studies and experimental work that the presence of sucrose and isomaltooligosaccharides could stabilize branching sucrases. Notably, DSF experiments showed an increase in the melting temperature correlated with an increase of sucrose concentration or an increase of the oligosaccharide size, indicating an enzyme stabilization (Figure 3). Active form of BRS-B Δ 2 however could not be used in presence of sucrose in crystallographic assays due to its hydrolytic activity. Therefore, I constructed the inactive mutant BRS-B Δ 2 E709Q, mutated on its catalytic acid/base residue. The mutant enzyme should still be able to bind sucrose but not utilize it. After production and purification set-up, the pure enzyme was assayed and a first set of 384 conditions was tested in presence of sucrose and isomaltooligosaccharides of different sizes (up to DP11). Luckily, I obtained small crystals (10 μ m) and the first low resolution diffraction data sets could be collected. Optimization of the crystallization conditions is still ongoing and will hopefully allow to obtain larger crystals and better diffraction data. This work is currently in progress with the help of G. Cioci, research engineer in the team, and will hopefully lead to exploitation and valorization of the crystallographic structure.

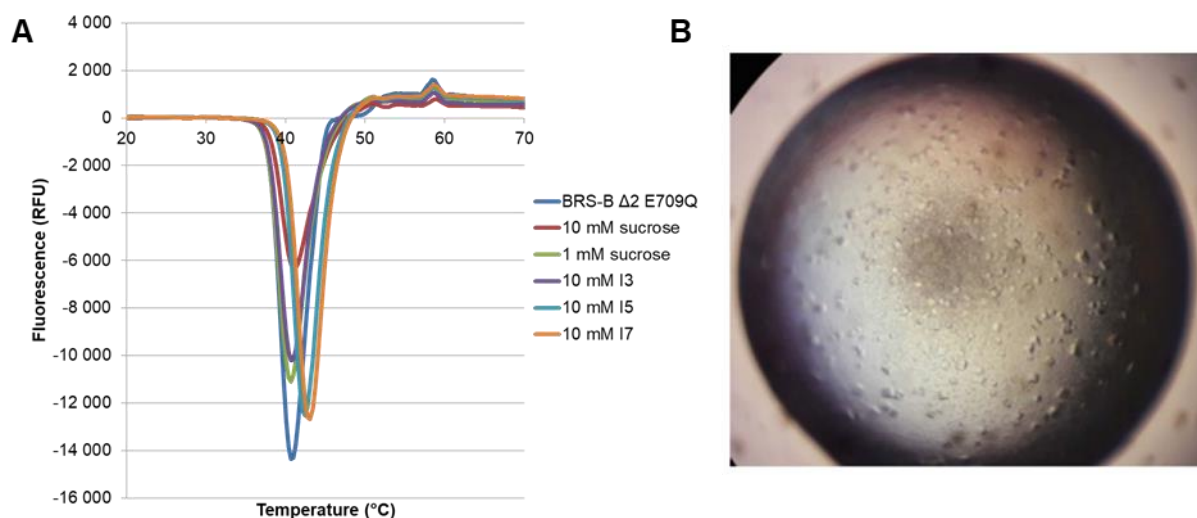


Figure 3: **A/** Melting curves obtained with Differential Scanning Fluorimetry experiments with the inactive mutant E709Q of the parental enzyme BRS-B Δ 2 only, or in presence of either sucrose or isomaltooligosaccharides (I3, I5, I7). **B/** First crystals (10 μ m) obtained with BRS-B Δ 2 in presence of sucrose and an isomaltooligosaccharide ligand.

Of prime interest for the project, we also aim in the near future at determining crystallographic structures of the mutants of interest, potentially in complex with their pentasaccharide products or **ABC'D'** acceptor. This will validate the study presented herein, and allow identifying the mutations that were mostly involved in the glycosylation selectivity.

Improved knowledge could then help to further engineer these enzymes to improve yields and hopefully access the last two remaining glycosylation patterns representative of *S. flexneri* prevalent serotypes.

Validation of the highly convergent synthesis strategy by formulating the first proof of concept glycoconjugate vaccine candidates

The next steps of this project will be to produce glyco bricks at the gram-scale. After the chemical synthesis is ready for transfer to external suppliers, gram-scale synthesis of the tetrasaccharide will allow to consider large scale enzymatic glycosylation of **ABC'D'** to produce enough pentasaccharides for the subsequent chemical elongation phases by the group of L. Mulard at the Pasteur Institute. First, chemical elongation of the glyco bricks will be attempted. The O-allyl protecting group introduced at 1_{D'} in combination with 2_{D'}-trichloroacetamide should allow assembling the glyco bricks to at least a 15mer. Moreover, thanks to the chloroacetyl group, derivatization towards acetyl or hydroxyl at position 2_{C'} will be possible at this point as required depending on the subserotype targeted (e.g. 4a or 4b). Assembly can occur either in a linear or modular way and artificial haptens could be produced to display, at the surface of carrier proteins, oligosaccharides representative of one or several *S. flexneri* serotype(s). The corresponding glycoconjugates will then be evaluated in immunology assays. Variation in the structures will also be considered by varying combinations of glyco bricks, oligosaccharide size, number of linked haptens, or also the nature of the carrier protein or linker.

In conclusion, this preliminary work paves the way for future projects which will hopefully allow:

- 1) To improve the reaction conditions for **ABC'D'** glycosylation
- 2) To diversify accessible glycosylated products using the enzymatic glycosylation toolbox of great interest to identify starting points in future projects
- 3) To accelerate selection of sucrose-active mutants from large libraries thanks to the engineered strain
- 4) To decipher the molecular determinants responsible for α 1,3 linkage specificity of BRS-B Δ 2 and improve our knowledge on this class of enzymes, in order to better guide future enzyme engineering projects
- 5) To produce at the gram scale glyco bricks described in this study for their chemical elongation and biological evaluation.

REFERENCES

A

- Abo H, Matsumura T, Kodama T, Ohta H, Fukui K, Kato K, Kagawa H. 1991. Peptide sequences for sucrose splitting and glucan binding within *Streptococcus sobrinus* glucosyltransferase (water-insoluble glucan synthetase). *J Bacteriol* **173**: 989–96. <https://doi.org/10.1128/jb.173.3.989-996.1991>.
- Abronina PI, Sedinkin SL, Podvalnyy NM, Fedina KG, Zinin AI, Torgov VI, Kononov LO. 2011. Formation of orthoester-linked d-arabinofuranose oligosaccharides and their isomerization into the corresponding glycosides. *Tetrahedron Lett* **52**: 1794–1796. <https://doi.org/10.1016/j.tetlet.2011.02.019>.
- Adamo R. 2017. Advancing homogeneous antimicrobial glycoconjugate vaccines. *Acc Chem Res* **50**: 1001–1006. <https://doi.org/10.1021/acs.accounts.7b00106>.
- Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, Tawfik DS. 2005. The “evolvability” of promiscuous protein functions. *Nat Genet* **37**: 73–76. <https://doi.org/10.1038/ng1482>.
- Aharoni A, Thieme K, Chiu CPC, Buchini S, Lairson LL, Chen H, Strynadka NCJ, Wakarchuk WW, Withers SG. 2006. High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nat Methods* **3**: 609–614. <https://doi.org/10.1038/nmeth899>.
- Alford RF, Leaver-Fay A, Jeliaskov JR, O’Meara MJ, DiMaio FP, Park H, Shapovalov M V., Renfrew PD, Mulligan VK, Kappel K, et al. 2017. The Rosetta all-atom energy function for macromolecular modeling and design. *bioRxiv* 106054. <https://doi.org/10.1101/106054>.
- Allison GE, Verma NK. 2000. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol* **8**: 17–23. [https://doi.org/10.1016/S0966-842X\(99\)01646-7](https://doi.org/10.1016/S0966-842X(99)01646-7).
- Amaya MF, Watts AG, Damager I, Wehenkel A, Nguyen T, Buschiazzi A, Paris G, Frasch AC, Withers SG, Alzari PM. 2004. Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. *Structure* **12**: 775–784. <https://doi.org/10.1016/j.str.2004.02.036>.
- André I, Potocki-Véronèse G, Barbe S, Moulis C, Remaud-Siméon M. 2014. CAZyme discovery and design for sweet dreams. *Curr Opin Chem Biol* **19**: 17–24. <https://doi.org/10.1016/j.cbpa.2013.11.014>.
- André I, Potocki-Véronèse G, Morel S, Monsan P, Remaud-Siméon M. 2010. Sucrose-utilizing transglucosidases for biocatalysis. *Top Curr Chem* **294**: 25–48. https://doi.org/10.1007/128_2010_52.
- Antranikian G, Egorova K. 2003. Extremophiles, a unique resource of biocatalysts for industrial biotechnology. In *Physiology and Biochemistry of Extremophiles* (eds. C. Gerday and N. Glandsdorff), Vol. 12 of, pp. 361–406, American Society of Microbiology, N.W., Washington, DC, U.S.A. <https://doi.org/10.1128/9781555815813.ch27>.
- Åqvist J. 1990. Ion-water interaction potentials derived from free energy perturbation simulations. *J Phys Chem* **94**: 8021–8024. <https://doi.org/10.1021/j100384a009>.
- Argüello-Morales MA, Remaud-Simeon M, Pizzut S, Sarçabal P, Willemot R, Monsan P. 2000. Sequence analysis of the gene encoding alternansucrase, a sucrose glucosyltransferase from *Leuconostoc mesenteroides* NRRL B-1355. *FEMS Microbiol Lett* **182**: 81–5. <https://doi.org/10.1111/j.1574-6968.2000.tb08878.x>.
- Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat RJ, Stoddard BL, Baker D. 2006. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**: 656–659. <https://doi.org/10.1038/nature04818>.
- Awad FN, Laborda P, Wang M, Lu AM, Li Q, Cai ZP, Liu L, Voglmeir J. 2017. Discovery and biochemical characterization of a mannose phosphorylase catalyzing the synthesis of novel β -1,3-mannosides. *Biochim Biophys Acta - Gen Subj* **1861**: 3231–3237. <https://doi.org/10.1016/j.bbagen.2017.09.013>.
- Bai Y, Gangoiti J, Dijkstra BW, Dijkhuizen L, Pijning T. 2017. Crystal Structure of 4,6- α -Glucanotransferase Supports Diet-Driven Evolution of GH70 Enzymes from α -Amylases in Oral Bacteria. *Structure* **25**: 231–242. <https://doi.org/10.1016/j.str.2016.11.023>.

- Barel L-A, Mulard LA. 2019. Classical and novel strategies to develop a *Shigella* glycoconjugate vaccine: from concept to efficacy in human. *Hum Vaccin Immunother* **0**: 21645515.2019.1606972. <https://doi.org/10.1080/21645515.2019.1606972>.
- Barry EM, Pasetti MF, Sztein MB, Fasano A, Kotloff KL, Levine MM. 2013. Progress and pitfalls in *Shigella* vaccine research. *Nat Rev Gastroenterol Hepatol* **10**: 245–55. <https://doi.org/10.1038/nrgastro.2013.12>.
- Bayly CI, Cieplak P, Cornell W, Kollman PA. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* **97**: 10269–10280. <https://doi.org/10.1021/j100142a004>.
- Becker S, Schmoldt HU, Adams TM, Wilhelm S, Kolmar H. 2004. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Curr Opin Biotechnol* **15**: 323–329. <https://doi.org/10.1016/j.copbio.2004.06.001>.
- Bélot F, Guerreiro C, Baleux F, Mulard LA. 2005. Synthesis of two linear PADRE conjugates bearing a deca- or pentadecasaccharide B epitope as potential synthetic vaccines against *Shigella flexneri* serotype 2a infection. *Chem - A Eur J* **11**: 1625–1635. <https://doi.org/10.1002/chem.200400903>.
- Bélot F, Wright K, Costachel C, Phalipon A, Mulard LA. 2004. Blockwise Approach to Fragments of the O-Specific Polysaccharide of *Shigella flexneri* Serotype 2a: Convergent Synthesis of a Decasaccharide Representative of a Dimer of the Branched Repeating Unit 1. *J Org Chem* **69**: 1060–1074. <https://doi.org/10.1021/jo035125b>.
- Ben Imeddougene A, Esque J, André I. 2018. Combining multi-scale modelling methods to decipher molecular motions of a branching sucrose from glycoside-hydrolase family 70. *PLoS One* **13**: e0201323. <https://doi.org/10.1371/journal.pone.0201323>.
- Benkoulouche M, Fauré R, Remaud-Siméon M, Moulis C, André I. 2019. Harnessing glycoenzyme engineering for synthesis of bioactive oligosaccharides. *Interface Focus* **9**. <https://doi.org/10.1098/rsfs.2018.0069>.
- 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. <https://doi.org/10.1063/1.448118>.
- Berti F, Adamo R. 2018. Antimicrobial glycoconjugate vaccines: an overview of classic and modern approaches for protein modification. *Chem Soc Rev* **47**: 9015–9025. <https://doi.org/10.1039/c8cs00495a>.
- Bissaro B, Monsan P, Fauré R, O'Donohue MJ. 2015. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem J* **467**: 17–35. <https://doi.org/10.1042/BJ20141412>.
- Bode L. 2012. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**: 1147–1162. <https://doi.org/10.1093/glycob/cws074>.
- Bopp CA, Fields PI, Nataro JP, Strockbine NA, Kaper JB. 2015. *Escherichia*, *Shigella*, and *Salmonella*. In *Manual of Clinical Microbiology, 11th Edition*, pp. 685–713, American Society of Microbiology <https://doi.org/10.1128/9781555817381.ch37>.
- Boutet J, Blasco P, Guerreiro C, Thouron F, Darteville S, Nato F, Cañada FJ, Ardá A, Phalipon A, Jiménez-Barbero J, et al. 2016. Detailed Investigation of the Immunodominant Role of O-Antigen Stoichiometric O-Acetylation as Revealed by Chemical Synthesis, Immunochemistry, Solution Conformation and STD-NMR Spectroscopy for *Shigella flexneri* 3a. *Chemistry* **22**: 10892–911. <https://doi.org/10.1002/chem.201600567>.
- Boutet J, Guerreiro C, Mulard LA. 2008. Synthesis of branched tri- to pentasaccharides representative of fragments of *Shigella flexneri* serotypes 3a and/or X O-antigens. *Tetrahedron* **64**: 10558–10572. <https://doi.org/10.1016/j.tet.2008.08.080>.
- Boutet J, Mulard LA. 2008. Synthesis of Two Tetra- and Four Pentasaccharide Fragments of *Shigella flexneri* Serotypes 3a and X O-Antigens from a Common Tetrasaccharide Intermediate. *European J Org Chem* **2008**: 5526–5542. <https://doi.org/10.1002/ejoc.200800693>.
- Bozonnet S, Dols-Laffargue M, Fabre E, Pizzut S, Remaud-Simeon M, Monsan P, Willemot RM. 2002. Molecular characterization of DSR-E, an α -1,2 linkage-synthesizing dextranucrase with two

- catalytic domains. *J Bacteriol* **184**: 5753–5761. <https://doi.org/10.1128/JB.184.20.5753-5761.2002>.
- Brison Y, Fabre E, Moulis C, Portais JC, Monsan P, Remaud-Siméon M. 2010. Synthesis of dextrans with controlled amounts of α -1,2 linkages using the transglucosidase GBD-CD2. *Appl Microbiol Biotechnol* **86**: 545–554. <https://doi.org/10.1007/s00253-009-2241-z>.
- Brison Y, Malbert Y, Czaplicki G, Mourey L, Remaud-Simeon M, Tranier S. 2016. Structural insights into the carbohydrate-binding ability of an α -(1→2) branching sucrose from glycoside-hydrolase family 70. *J Biol Chem* **291**: 7527–7540. <https://doi.org/10.1074/jbc.M115.688796>.
- 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 sucrose derived from DSR-E glucansucrase. *J Biol Chem* **287**: 7915–7924. <https://doi.org/10.1074/jbc.M111.305078>.
- Brooker BE. 1976. Surface coat transformation and capsule formation by *Leuconostoc mesenteroides* NCDO 523 in the presence of sucrose. *Arch Microbiol* **111**: 99–104. <https://doi.org/10.1007/BF00446555>.
- Brown SD, Babbitt PC. 2014. New insights about enzyme evolution from large scale studies of sequence and structure relationships. *J Biol Chem* **289**: 30221–30228. <https://doi.org/10.1074/jbc.R114.569350>.
- Buschiazzo A. 2000. Structural basis of sialyltransferase activity in trypanosomal sialidases. *EMBO J* **19**: 16–24. <https://doi.org/10.1093/emboj/19.1.16>.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**: D233–D238. <https://doi.org/10.1093/nar/gkn663>.
- Champion E, André I, Moulis C, Boutet J, Descroix K, Morel S, Monsan P, Mulard LA, Remaud-Siméon M. 2009a. Design of α -transglucosidases of controlled specificity for programmed chemoenzymatic synthesis of antigenic oligosaccharides. *J Am Chem Soc* **131**: 7379–7389. <https://doi.org/10.1021/ja900183h>.
- Champion E, André I, Mulard LA, Monsan P, Remaud-Siméon M, Morel S. 2009b. Synthesis of L-Rhamnose and N-Acetyl-D-Glucosamine Derivatives Entering in the Composition of Bacterial Polysaccharides by Use of Glucansucrases. *J Carbohydr Chem* **28**: 142–160. <https://doi.org/10.1080/07328300902755796>.
- Champion E, Guérin F, Moulis C, Barbe S, Tran TH, Morel S, Descroix K, Monsan P, Mourey L, Mulard LA, et al. 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**: 18677–88. <https://doi.org/10.1021/ja306845b>.
- Champion E, Moulis C, Morel S, Mulard LA, Monsan P, Remaud-Siméon M, André I. 2010. A pH-Based High-Throughput Screening of Sucrose-Utilizing Transglucosidases for the Development of Enzymatic Glucosylation Tools. *ChemCatChem* **2**: 969–975. <https://doi.org/10.1002/cctc.201000111>.
- Chang A, Singh S, Phillips GN, Thorson JS. 2011. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr Opin Biotechnol* **22**: 800–808. <https://doi.org/10.1016/j.copbio.2011.04.013>.
- Chassagne P, Fontana C, Guerreiro C, Gauthier C, Phalipon A, Widmalm G, Mulard LA. 2013. Structural Studies of the O -Acetyl-Containing O-Antigen from a *Shigella flexneri* Serotype 6 Strain and Synthesis of Oligo-saccharide Fragments Thereof. *European J Org Chem* **2013**: 4085–4106. <https://doi.org/10.1002/ejoc.201300180>.
- Chavaroche AAE, van den Broek LAM, Springer J, Boeriu C, Eggink G. 2011. Analysis of the polymerization initiation and activity of *Pasteurella multocida* heparosan synthase PmHS2, an enzyme with glycosyltransferase and UDP-sugar hydrolase activity. *J Biol Chem* **286**: 1777–1785. <https://doi.org/10.1074/jbc.M110.136754>.
- Chen C-Y, Georgiev I, Anderson AC, Donald BR. 2009. Computational structure-based redesign of enzyme activity. *Proc Natl Acad Sci U S A* **106**: 3764–3769.

<https://doi.org/10.1073/pnas.0900266106>.

- Chen C, Van Der Borght J, De Vreese R, D'Hooghe M, Soetaert W, Desmet T. 2014. Engineering the specificity of trehalose phosphorylase as a general strategy for the production of glycosyl phosphates. *Chem Commun* **50**: 7834–7836. <https://doi.org/10.1039/c4cc02202e>.
- Chen R. 2018. Enzyme and microbial technology for synthesis of bioactive oligosaccharides: an update. *Appl Microbiol Biotechnol* **102**: 3017–3026. <https://doi.org/10.1007/s00253-018-8839-2>.
- Chen WH, Kotloff KL. 2016. Shigella Vaccine Development: Finding the Path of Least Resistance. *Clin Vaccine Immunol* **23**: 904–907. <https://doi.org/10.1128/cvi.00444-16>.
- Cherniak R, Sundstrom JB. 1994. Polysaccharide antigens of the capsule of *Cryptococcus neoformans*. *Infect Immun* **62**: 1507–12.
- Choi YH, Kim JH, Park BS, Kim BG. 2016. Solubilization and iterative saturation mutagenesis of α 1,3-fucosyltransferase from *Helicobacter pylori* to enhance its catalytic efficiency. *Biotechnol Bioeng* **113**: 1666–1675. <https://doi.org/10.1002/bit.25944>.
- Choi YH, Kim JH, Park JH, Lee N, Kim DH, Jang KS, Park IH, Kim BG. 2014. Protein engineering of α 2,3/2,6-sialyltransferase to improve the yield and productivity of *in vitro* sialyllactose synthesis. *Glycobiology* **24**: 159–169. <https://doi.org/10.1093/glycob/cwt092>.
- Chusacultanachai S, Yuthavong Y. 2004. Random mutagenesis strategies for construction of large and diverse clone libraries of mutated DNA fragments. In *Parasite Genomics Protocols*, Vol. 270 of, pp. 319–334, Humana Press, New Jersey <https://doi.org/10.1385/1-59259-793-9:319>.
- Čihák R. 2009. Reach - an overview. *Interdiscip Toxicol* **2**: 42–44. <https://doi.org/10.2478/v10102-009-0007-1>.
- Claverie M, Cioci G, Guionnet M, Schörghuber J, Lichtenecker R, Moulis C, Remaud-Simeon M, Lippens G. 2019. Futile Encounter Engineering of the DSR-M Dextranucrase Modifies the Resulting Polymer Length. *Biochemistry* [acs.biochem.9b00373](https://doi.org/10.1021/acs.biochem.9b00373). <https://doi.org/10.1021/acs.biochem.9b00373>.
- Claverie M, Cioci G, Vuillemin M, Monties N, Roblin P, Lippens G, Remaud-Simeon M, Moulis C. 2017. Investigations on the determinants responsible for low molar mass dextran formation by DSR-M dextranucrase. *ACS Catal* **7**: 7106–7119. <https://doi.org/10.1021/acscatal.7b02182>.
- Cobucci-Ponzano B, Conte F, Bedini E, Corsaro MM, Parrilli M, Sulzenbacher G, Lipski A, Dal Piaz F, Lepore L, Rossi M, et al. 2009. β -glycosyl azides as substrates for α -glycosynthases: preparation of efficient α -L-fucosynthases. *Chem Biol* **16**: 1097–1108. <https://doi.org/10.1016/j.chembiol.2009.09.013>.
- Cobucci-Ponzano B, Strazzulli A, Rossi M, Moracci M. 2011a. Glycosynthases in biocatalysis. *Adv Synth Catal* **353**: 2284–2300. <https://doi.org/10.1002/adsc.201100461>.
- Cobucci-Ponzano B, Zorzetti C, Strazzulli A, Carillo S, Bedini E, Corsaro MM, Comfort DA, Kelly RM, Rossi M, Moracci M. 2011b. A novel α -D-galactosynthase from *Thermotoga maritima* converts β -D-galactopyranosyl azide to α -galacto-oligosaccharides. *Glycobiology* **21**: 448–456. <https://doi.org/10.1093/glycob/cwq177>.
- Cohen D, Ashkenazi S, Green M, Lerman Y, Slepon R, Robin G, Orr N, Taylor DN, Sadoff JC, Chu C, et al. 1996. Safety and immunogenicity of investigational Shigella conjugate vaccines in Israeli volunteers. *Infect Immun* **64**: 4074–4077.
- Cohen D, Atsmon J, Artaud C, Meron-Sudai S, Gougeon M, Bialik A, Goren S, Asato V, Ariel-Cohen O, Reizis A, et al. 2017. A phase I dose escalation study to assess the safety and immunogenicity of the SF2a-TT15 conjugate vaccine against *S. flexneri* 2a in healthy adult volunteers (Preliminary Results). In *Vaccines for Enteric Diseases*, Albufera.
- Cohen D, Green MS, Block C, Rouach T, Ofek I. 1988. Serum antibodies to lipopolysaccharide and natural immunity to shigellosis in an Israeli military population. *J Infect Dis* **157**: 1068–71. <https://doi.org/10.1093/infdis/157.5.1068>.
- Cohen D, Green MS, Block C, Slepon R, Ofek I. 1991. Prospective study of the association between serum antibodies to lipopolysaccharide O antigen and the attack rate of shigellosis. *J Clin Microbiol* **29**: 386–9.

- Cohen D, Meron-Sudai S, Bialik A, Asato V, Goren S, Ariel-Cohen O, Reizis A, Hochberg A, Ashkenazi S. 2019. Serum IgG antibodies to Shigella lipopolysaccharide antigens - a correlate of protection against shigellosis. *Hum Vaccin Immunother* **15**: 1401–1408. <https://doi.org/10.1080/21645515.2019.1606971>.
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **117**: 5179–5197. <https://doi.org/10.1021/ja00124a002>.
- Côté GL, Robyt JF. 1982. Isolation and partial characterization of an extracellular glucansucrase from *Leuconostoc mesenteroides* NRRL B-1355 that synthesizes an alternating (1→6), (1→3)- α -D-glucan. *Carbohydr Res* **101**: 57–74. [https://doi.org/10.1016/S0008-6215\(00\)80795-8](https://doi.org/10.1016/S0008-6215(00)80795-8).
- Côté GL, Skory CD. 2012. Cloning, expression, and characterization of an insoluble glucan-producing glucansucrase from *Leuconostoc mesenteroides* NRRL B-1118. *Appl Microbiol Biotechnol* **93**: 2387–2394. <https://doi.org/10.1007/s00253-011-3562-2>.
- Cozens C, Pinheiro VB. 2018. Darwin Assembly: fast, efficient, multi-site bespoke mutagenesis. *Nucleic Acids Res* **46**. <https://doi.org/10.1093/nar/gky067>.
- Crout DH, Vic G. 1998. Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis. *Curr Opin Chem Biol* **2**: 98–111. [https://doi.org/10.1016/s1367-5931\(98\)80041-0](https://doi.org/10.1016/s1367-5931(98)80041-0).
- D’Almeida A, Ionata M, Tran V, Tellier C, Dion M, Rabiller C. 2009. An expeditious and efficient synthesis of β -D-galactopyranosyl-(1→3)-D-N-acetylglucosamine (lacto-N-biose) using a glycosynthase from *Thermus thermophilus* as a catalyst. *Tetrahedron Asymmetry* **20**: 1243–1246. <https://doi.org/10.1016/j.tetasy.2009.05.007>.
- Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* **98**: 10089–10092. <https://doi.org/10.1063/1.464397>.
- Daudé D, André I, Monsan P. 2014. Successes in engineering glucansucrases to enhance glycodiversification. *Carbohydr Chem* **40**: 624–645. <https://doi.org/10.1039/9781849739986>.
- Daudé D, Champion E, Morel S, Guieysse D, Remaud-Siméon M, André I. 2013a. Probing substrate promiscuity of amylosucrase from *Neisseria polysaccharea*. *ChemCatChem* **5**: 2288–2295. <https://doi.org/10.1002/cctc.201300012>.
- Daudé D, Topham CM, Remaud-Siméon M, André I. 2013b. Probing impact of active site residue mutations on stability and activity of *Neisseria polysaccharea* amylosucrase. *Protein Sci* **22**: 1754–1765. <https://doi.org/10.1002/pro.2375>.
- Davies GJ, Wilson KS, Henrissat B. 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* **321** (Pt 2): 557–559. <https://doi.org/10.1007/s007920050009>.
- De Groeve MRM, De Baere M, Hoflack L, Desmet T, Vandamme EJ, Soetaert W. 2009. Creating lactose phosphorylase enzymes by directed evolution of cellobiose phosphorylase. *Protein Eng Des Sel* **22**: 393–399. <https://doi.org/10.1093/protein/gzp017>.
- De Groeve MRM, Desmet T, Soetaert W. 2011. Engineering of cellobiose phosphorylase for glycoside synthesis. *J Biotechnol* **156**: 253–60. <https://doi.org/10.1016/j.jbiotec.2011.07.006>.
- De Groeve MRM, Remmery L, Van Hoorebeke A, Stout J, Desmet T, Savvides SN, Soetaert W. 2010. Construction of cellobiose phosphorylase variants with broadened acceptor specificity towards anomeric substituted glucosides. *Biotechnol Bioeng* **107**: 413–420. <https://doi.org/10.1002/bit.22818>.
- Deangelis PL, Liu J, Linhardt RJ. 2013. Chemoenzymatic synthesis of glycosaminoglycans: re-creating, re-modeling and re-designing nature’s longest or most complex carbohydrate chains. *Glycobiology* **23**: 764–777. <https://doi.org/10.1093/glycob/cwt016>.
- DeLano WL. 2009. PyMOL: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter On Protein Crystallography*.
- Dennig A, Shivange A V., Marienhagen J, Schwaneberg U. 2011. Omnichange: The sequence independent method for simultaneous site-saturation of five codons. *PLoS One* **6**. <https://doi.org/10.1371/journal.pone.0026222>.

- Desmet T, Soetaert W. 2012. Broadening the synthetic potential of disaccharide phosphorylases through enzyme engineering. *Process Biochem* **47**: 11–17. <https://doi.org/10.1016/j.procbio.2011.10.039>.
- 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**: 10786–10801. <https://doi.org/10.1002/chem.201103069>.
- Devlamynck T, te Poele EM, Meng X, van Leeuwen SS, Dijkhuizen L. 2016. Glucansucrase Gtf180- Δ N of *Lactobacillus reuteri* 180: enzyme and reaction engineering for improved glycosylation of non-carbohydrate molecules. *Appl Microbiol Biotechnol* **100**: 7529–7539. <https://doi.org/10.1007/s00253-016-7476-x>.
- Devlamynck T, te Poele EM, Quataert K, Gerwig GJ, Van de Walle D, Dewettinck K, Kamerling JP, Soetaert W, Dijkhuizen L. 2019. Trans- α -glucosylation of stevioside by the mutant glucansucrase enzyme Gtf180- Δ N-Q1140E improves its taste profile. *Food Chem* **272**: 653–662. <https://doi.org/10.1016/j.foodchem.2018.08.025>.
- Díez-Municio M, Herrero M, Olano A, Moreno FJ. 2014. Synthesis of novel bioactive lactose-derived oligosaccharides by microbial glycoside hydrolases. *Microb Biotechnol* **7**: 315–331. <https://doi.org/10.1111/1751-7915.12124>.
- Ding L, Zhao C, Qu J, Li Y, Sugiarto G, Yu H, Wang J, Chen X. 2015. A *Photobacterium* sp. α 2-6-sialyltransferase (Psp2,6ST) mutant with an increased expression level and improved activities in sialylating Tn antigens. *Carbohydr Res* **408**: 127–133. <https://doi.org/10.1016/j.carres.2014.12.007>.
- Ding N, Li X, Chinoy ZS, Boons G-J. 2017. Synthesis of a Glycosylphosphatidylinositol Anchor Derived from *Leishmania donovani* That Can Be Functionalized by Cu-Catalyzed Azide–Alkyne Cycloadditions. *Org Lett* **19**: 3827–3830. <https://doi.org/10.1021/acs.orglett.7b01703>.
- Dupont HL. 1989. Inoculum Size in Shigellosis and Implications For Expected Mode of Transmission. *J Infect Dis* **159**: 1126–1128. <https://doi.org/10.1093/infdis/159.6.1126>.
- Epskamp S, Schmittmann VD, Borsboom D. 2012. r - correlation matrix to build networks - Stack Overflow. **48**. <https://doi.org/10.18637/jss.v048.i04>.
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. 2006. Comparative Protein Structure Modeling Using Modeller. *Curr Protoc Bioinforma* **15**: 5.6.1-5.6.30. <https://doi.org/10.1002/0471250953.bi0506s15>.
- Fabre E, Bozonnet S, Arcache A, Willemot R-M, Vignon M, Monsan P, Remaud-Simeon M. 2005. Role of the two catalytic domains of DSR-E dextranucrase and their involvement in the formation of highly α -1,2 Branched Dextran. *J Bacteriol* **187**: 296–303. <https://doi.org/10.1128/JB.187.1.296-303.2005>.
- Fang J, Li J, Chen X, Zhang Y, Wang J, Guo Z, Zhang W, Yu L, Brew K, Wang PG. 1998. Highly efficient chemoenzymatic synthesis of α -galactosyl epitopes with a recombinant α (1 \rightarrow 3)-galactosyltransferase. *J Am Chem Soc* **120**: 6635–6638. <https://doi.org/10.1021/ja9808898>.
- Flemming H-C, Wingender J. 2010. The biofilm matrix. *Nat Rev Microbiol* **8**: 623–33. <https://doi.org/10.1038/nrmicro2415>.
- Fraga Vidal R, Moulis C, Escalier P, Remaud-Siméon M, Monsan P. 2011. Isolation of a gene from *Leuconostoc citreum* B/110-1-2 encoding a novel dextranucrase enzyme. *Curr Microbiol* **62**: 1260–6. <https://doi.org/10.1007/s00284-010-9851-7>.
- France SP, Aleku GA, Sharma M, Mangas-Sanchez J, Howard RM, Steflík J, Kumar R, Adams RW, Slabu I, Crook R, et al. 2017. Biocatalytic Routes to Enantiomerically Enriched Dibenz[c,e]azepines. *Angew Chemie - Int Ed* **56**: 15589–15593. <https://doi.org/10.1002/anie.201708453>.
- Fruchterman TJ, Reingold EM. 1991. Graph drawing by force-directed placement. *Softw Pract Exp* **21**: 1129–1164. <https://doi.org/doi:10.1002/spe.4380211102>.
- Fuchs M, Farnberger JE, Kroutil W. 2015. The Industrial Age of Biocatalytic Transamination. *European J Org Chem* **2015**: 6965–6982. <https://doi.org/10.1002/ejoc.201500852>.

- Fujiwara T, Terao Y, Hoshino T, Kawabata S, Ooshima T, Sobue S, Kimura S, Hamada S. 1998. Molecular analyses of glucosyltransferase genes among strains of *Streptococcus mutans*. *FEMS Microbiol Lett* **161**: 331–336. [https://doi.org/10.1016/S0378-1097\(98\)00091-3](https://doi.org/10.1016/S0378-1097(98)00091-3).
- Fürstner A, Jeanjean F, Razon P. 2002. Total Synthesis of Woodrosin. *Angew Chemie Int Ed* **41**: 2097. [https://doi.org/10.1002/1521-3773\(20020617\)41:12<2097::AID-ANIE2097>3.0.CO;2-T](https://doi.org/10.1002/1521-3773(20020617)41:12<2097::AID-ANIE2097>3.0.CO;2-T).
- Gagarinov IA, Li T, Toraño JS, Caval T, Srivastava AD, Kruijtz JAW, Heck AJR, Boons GJ. 2017. Chemoenzymatic approach for the preparation of asymmetric bi-, tri-, and tetra-antennary N-glycans from a common precursor. *J Am Chem Soc* **139**: 1011–1018. <https://doi.org/10.1021/jacs.6b12080>.
- Gagarinov IA, Li T, Wei N, Sastre Toraño J, de Vries RP, Wolfert MA, Boons G-J. 2019. Protecting-Group-Controlled Enzymatic Glycosylation of Oligo- N -Acetylactosamine Derivatives. *Angew Chemie Int Ed* **58**: 10547–10552. <https://doi.org/10.1002/anie.201903140>.
- Gangoiti J, Corwin SF, Lamothe LM, Vafiadi C, Hamaker BR, Dijkhuizen L. 2018a. Synthesis of novel α -glucans with potential health benefits through controlled glucose release in the human gastrointestinal tract. *Crit Rev Food Sci Nutr* **0**: 1–24. <https://doi.org/10.1080/10408398.2018.1516621>.
- Gangoiti J, Pijning T, Dijkhuizen L. 2018b. Biotechnological potential of novel glycoside hydrolase family 70 enzymes synthesizing α -glucans from starch and sucrose. *Biotechnol Adv* **36**: 196–207. <https://doi.org/10.1016/j.biotechadv.2017.11.001>.
- Gangoiti J, Pijning T, Dijkhuizen L. 2016. The *Exiguobacterium sibiricum* 255-15 GtFC Enzyme Represents a Novel Glycoside Hydrolase 70 Subfamily of 4,6- α -Glucanotransferase Enzymes. *Appl Environ Microbiol* **82**: 756–66. <https://doi.org/10.1128/AEM.03420-15>.
- Gantt RW, Peltier-Pain P, Thorson JS. 2011. Enzymatic methods for glyco(diversification/randomization) of drugs and small molecules. *Nat Prod Rep* **28**: 1811–1853. <https://doi.org/10.1039/c1np00045d>.
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**: 3784–3788. <https://doi.org/10.1093/nar/gkg563>.
- Gauthier C, Chassagne P, Theillet F-X, Guerreiro C, Thouron F, Nato F, Delepierre M, Sansonetti PJ, Phalipon A, Mulard LA. 2014. Non-stoichiometric O-acetylation of *Shigella flexneri* 2a O-specific polysaccharide: synthesis and antigenicity. *Org Biomol Chem* **12**: 4218–32. <https://doi.org/10.1039/c3ob42586j>.
- Geremia RA, Samain E. 2000. Production of heterologous oligosaccharides by recombinant bacteria (recombinant oligosaccharides). *Carbohydrates Chem Biol* **2–4**: 845–860. <https://doi.org/10.1002/9783527618255.ch30>.
- Gibson GR, Hutkins R, Sanders ME, Prescott SL, Reimer RA, Salminen SJ, Scott K, Stanton C, Swanson KS, Cani PD, et al. 2017. Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nat Rev Gastroenterol Hepatol* **14**: 491–502. <https://doi.org/10.1038/nrgastro.2017.75>.
- Gibson GR, Roberfroid MB. 1995. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J Nutr* **125**: 1401–12. <https://doi.org/10.1093/jn/125.6.1401>.
- Gigg J, Gigg R, Payne S, Conant R. 1985. The allyl group for protection in carbohydrate chemistry. 17. Synthesis of propyl O-(3,6-di-O-methyl-beta-D-glucopyranosyl)-(1----4)-O-(2,3-di-O-methyl-alpha-L-rhamnopyranosyl)-(1----2)-3-O-methyl-alpha-L-rhamnopyranoside: the oligosaccharide portion of. *Chem Phys Lipids* **38**: 299–307.
- Green AP, Turner NJ. 2016. Biocatalytic retrosynthesis: Redesigning synthetic routes to high-value chemicals. *Perspect Sci* **9**: 42–48. <https://doi.org/10.1016/j.pisc.2016.04.106>.
- Grönwall A, Ingelman B. 1945. Dextran as a Substitute for Plasma. *Nature* **155**: 45–45. <https://doi.org/10.1038/155045a0>.
- Guen Y Le, Chassagne P, Heiget G Le, Urban D, Mulard LA. 2017. Allyl 4,6-O-benzylidene-2-deoxy-2-trichloroacetamido- β -d-glucopyranoside. In *Carbohydrate Chemistry*, pp. 333–342, CRC Press

<https://doi.org/10.1201/9781315120300-39>.

- Guo Y, Jers C, Meyer AS, Li H, Kirpekar F, Mikkelsen JD. 2015. Modulating the regioselectivity of a *Pasteurella multocida* sialyltransferase for biocatalytic production of 3'- and 6'-sialyllactose. *Enzyme Microb Technol* **78**: 54–62. <https://doi.org/10.1016/j.enzmictec.2015.06.012>.
- Hamura K, Saburi W, Matsui H, Mori H. 2013. Modulation of acceptor specificity of *Ruminococcus albus* cellobiose phosphorylase through site-directed mutagenesis. *Carbohydr Res* **379**: 21–25. <https://doi.org/10.1016/j.carres.2013.06.010>.
- Han NS, Kim TJ, Park YC, Kim J, Seo JH. 2012. Biotechnological production of human milk oligosaccharides. *Biotechnol Adv* **30**: 1268–1278. <https://doi.org/10.1016/j.biotechadv.2011.11.003>.
- Hanada N, Kuramitsu HK. 1989. Isolation and characterization of the *Streptococcus mutans* gtfD gene, coding for primer-dependent soluble glucan synthesis. *Infect Immun* **57**: 2079–85.
- Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* **4**: 17. <https://doi.org/10.1186/1758-2946-4-17>.
- Hargreaves JM, Le Guen Y, Guerreiro C, Descroix K, Mulard LA. 2014. Linear synthesis of the branched pentasaccharide repeats of O-antigens from *Shigella flexneri* 1a and 1b demonstrating the major steric hindrance associated with type-specific glucosylation. *Org Biomol Chem* **12**: 7728–49. <https://doi.org/10.1039/c4ob01200c>.
- Henze M, You D-J, Kamerke C, Hoffmann N, Angkawidjaja C, Ernst S, Pietruszka J, Kanaya S, Elling L. 2014. Rational design of a glycosynthase by the crystal structure of β -galactosidase from *Bacillus circulans* (BgaC) and its use for the synthesis of N-acetyllactosamine type 1 glycan structures. *J Biotechnol* **191**: 78–85. <https://doi.org/10.1016/j.jbiotec.2014.07.003>.
- Ho T-L, Wong CM. 1975. Hydroxyl Protection by Levulinylation. *Synth Commun* **5**: 91–93. <https://doi.org/10.1080/00397917508061437>.
- Hogrefe HH, Cline J, Youngblood GL, Allen RM. 2002. Creating randomized amino acid libraries with the QuikChange® multi site-directed mutagenesis kit. *Biotechniques* **33**: 1158–1165. <https://doi.org/10.2144/02335pf01>.
- Hong S, Kyung M, Jo I, Kim YR, Ha NC. 2018. Structure-based protein engineering of bacterial β -xylosidase to increase the production yield of xylobiose from xylose. *Biochem Biophys Res Commun* **501**: 703–710. <https://doi.org/10.1016/j.bbrc.2018.05.051>.
- Hsieh P-C, Vaisvila R. 2013. Protein engineering: single or multiple site-directed mutagenesis. In *Methods in molecular biology (Clifton, N.J.)*, Vol. 965 of, pp. 173–186 https://doi.org/10.1007/978-1-62703-293-3_13.
- Hsu CH, Hung SC, Wu CY, Wong CH. 2011. Toward automated oligosaccharide synthesis. *Angew Chemie - Int Ed* **50**: 11872–11923. <https://doi.org/10.1002/anie.201100125>.
- Hu M, Lan Y, Lu A, Ma X, Zhang L. 2019. Glycan-based biomarkers for diagnosis of cancers and other diseases: Past, present, and future. pp. 1–24 <https://doi.org/10.1016/bs.pmbts.2018.12.002>.
- Hu Z, Bongat White AF, Mulard LA. 2017. Efficient Iterative Synthesis of O-Acetylated Tri- to Pentadecasaccharides Related to the Lipopolysaccharide of *Shigella flexneri* Type 3 a through Di- and Trisaccharide Glycosyl Donors. *Chem Asian J* **12**: 419–439. <https://doi.org/10.1002/asia.201600819>.
- Huang PS, Boyken SE, Baker D. 2016. The coming of age of de novo protein design. *Nature* **537**: 320–327. <https://doi.org/10.1038/nature19946>.
- Hunter CD, Guo T, Daskhan G, Richards MR, Cairo CW. 2018. Synthetic strategies for modified glycosphingolipids and their design as probes. *Chem Rev* **118**: 8188–8241. <https://doi.org/10.1021/acs.chemrev.8b00070>.
- Ihssen J, Haas J, Kowarik M, Wiesli L, Wacker M, Schwede T, Thony-Meyer L. 2015. Increased efficiency of *Campylobacter jejuni* N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biol* **5**: 140227–140227. <https://doi.org/10.1098/rsob.140227>.

- Ipatieff VN, Pines H. 1936. Propylene Polymerization: Under High Pressure and Temperature with and without Phosphoric Acid. *Ind Eng Chem* **28**: 684–686. <https://doi.org/10.1021/ie50318a018>.
- Ito K, Ito S, Shimamura T, Weyand S, Kawarasaki Y, Misaka T, Abe K, Kobayashi T, Cameron AD, Iwata S. 2011. Crystal structure of glucansucrase from the dental caries pathogen *Streptococcus mutans*. *J Mol Biol* **408**: 177–86. <https://doi.org/10.1016/j.jmb.2011.02.028>.
- Jahreis K, Bentler L, Bockmann J, Meyer A, Siepelmeyer J, Joseph W, Ec S, Hans S, Lengeler JW. 2002. Adaptation of sucrose metabolism in the *Escherichia coli* wild-type strain EC3123. *J Bacteriol* **184**: 5307–5316. <https://doi.org/10.1128/JB.184.19.5307>.
- Jers C, Michalak M, Larsen DM, Kepp KP, Li H, Guo Y, Kirpekar F, Meyer AS, Mikkelsen JD. 2014. Rational design of a new *Trypanosoma rangeli* trans-sialidase for efficient sialylation of glycans. *PLoS One* **9**. <https://doi.org/10.1371/journal.pone.0083902>.
- Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, et al. 2008. De Novo Computational Design of Retro-Aldol Enzymes. *Science (80-)* **319**: 1387–1391. <https://doi.org/10.1126/science.1152692>.
- Jiao G, Yu G, Zhang J, Ewart HS. 2011. Chemical structures and bioactivities of sulfated polysaccharides from marine algae. *Mar Drugs* **9**: 196–223. <https://doi.org/10.3390/md9020196>.
- Jimenez-Rosales A, Flores-Merino M V. 2018. Tailoring proteins to re-evolve Nature: a short review. *Mol Biotechnol* **60**: 946–974. <https://doi.org/10.1007/s12033-018-0122-3>.
- Johnson KA, Goody RS. 2011. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **50**: 8264–8269. <https://doi.org/10.1021/bi201284u>.
- Joo JC, Pohkrel S, Pack SP, Yoo YJ. 2010. Thermostabilization of *Bacillus circulans* xylanase via computational design of a flexible surface cavity. *J Biotechnol* **146**: 31–39. <https://doi.org/10.1016/j.jbiotec.2009.12.021>.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**: 926–935. <https://doi.org/10.1063/1.445869>.
- Joucla G, Pizzut S, Monsan P, Remaud-Simeon M. 2006. Construction of a fully active truncated alternansucrase partially deleted of its carboxy-terminal domain. *FEBS Lett* **580**: 763–768. <https://doi.org/10.1016/j.febslet.2006.01.001>.
- Kajala I, Shi Q, Nyssölä A, Maina NH, Hou Y, Katina K, Tenkanen M, Juvonen R. 2015. Cloning and characterization of a *Weissella confusa* dextranase and its application in high fibre baking. *PLoS One* **10**: e0116418. <https://doi.org/10.1371/journal.pone.0116418>.
- Kalynych S, Morona R, Cygler M. 2014. Progress in understanding the assembly process of bacterial O-antigen. *FEMS Microbiol Rev* **38**: 1048–1065. <https://doi.org/10.1111/1574-6976.12070>.
- Kang H, Oh J, Kim D. 2009. Molecular characterization and expression analysis of the glucansucrase DSRWC from *Weissella cibaria* synthesizing a (1!6) glucan. <https://doi.org/10.1111/j.1574-6968.2008.01460.x>.
- Kärnell A, Li A, Zhao CR, Karlsson K, Nguyen BM, Lindberg AA. 1995. Safety and immunogenicity study of the auxotrophic *Shigella flexneri* 2a vaccine SFL1070 with a deleted *aroD* gene in adult Swedish volunteers. *Vaccine* **13**: 88–99.
- Kawai S, Murata K. 2016. Biofuel Production Based on Carbohydrates from Both Brown and Red Macroalgae: Recent Developments in Key Biotechnologies. *Int J Mol Sci* **17**: 145. <https://doi.org/10.3390/ijms17020145>.
- Keys TG, Fuchs HLS, Ehrit J, Alves J, Freiburger F, Gerardy-Schahn R. 2014. Engineering the product profile of a polysialyltransferase. *Nat Chem Biol* **10**: 437–42. <https://doi.org/10.1038/nchembio.1501>.
- Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, Brewer TG, Engmann CM, Houpt ER, Kang G, et al. 2018. Morbidity and mortality due to shigella and enterotoxigenic *Escherichia coli* diarrhoea: the Global Burden of Disease Study 1990-2016. *Lancet Infect Dis* **18**: 1229–1240. [https://doi.org/10.1016/S1473-3099\(18\)30475-4](https://doi.org/10.1016/S1473-3099(18)30475-4).

- Khan FI, Wei DQ, Gu KR, Hassan MI, Tabrez S. 2016. Current updates on computer aided protein modeling and designing. *Int J Biol Macromol* **85**: 48–62. <https://doi.org/10.1016/j.ijbiomac.2015.12.072>.
- Kirschner KN, Woods RJ. 2001. Solvent interactions determine carbohydrate conformation. *Proc Natl Acad Sci* **98**: 10541–10545. <https://doi.org/10.1073/pnas.191362798>.
- Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **29**: 622–55. <https://doi.org/10.1002/jcc.20820>.
- Kitaoka M. 2015. Diversity of phosphorylases in glycoside hydrolase families. *Appl Microbiol Biotechnol* **99**: 8377–8390. <https://doi.org/10.1007/s00253-015-6927-0>.
- Klemm D, Kramer F, Moritz S, Lindström T, Ankerfors M, Gray D, Dorris A. 2011. Nanocelluloses: A New Family of Nature-Based Materials. *Angew Chemie Int Ed* **50**: 5438–5466. <https://doi.org/10.1002/anie.201001273>.
- Knirel YA, Sun Q, Senchenkova SN, Perepelov A V, Shashkov AS, Xu J. 2015. O-antigen modifications providing antigenic diversity of *Shigella flexneri* and underlying genetic mechanisms. *Biochemistry (Mosc)* **80**: 901–14. <https://doi.org/10.1134/S0006297915070093>.
- Ko JA, Ryu YB, Park TS, Jeong HJ, Kim JH, Park SJ, Kim JS, Kim D, Kim YM, Lee WS. 2012. Enzymatic synthesis of puerarin glucosides using *Leuconostoc Dextran* sucrose. *J Microbiol Biotechnol* **22**: 1224–1229. <https://doi.org/10.4014/jmb.1202.02007>.
- Kobata A. 2013. Exo- and endoglycosidases revisited. *Proc Jpn Acad Ser B Phys Biol Sci* **89**: 97–117. <https://doi.org/10.2183/pjab.89.97>.
- Kobata A. 2010. Structures and application of oligosaccharides in human milk. *Proc Japan Acad Ser B* **86**: 731–747. <https://doi.org/10.2183/pjab.86.731>.
- Koeller KM, Wong C. 2000. Synthesis of complex carbohydrates and glycoconjugates: enzyme-based and programmable one-pot strategies. *Chem Rev* **100**: 4465–4494. <https://doi.org/10.1021/cr990297n>.
- Konishi N, Torii Y, Yamamoto T, Ohta H, Fukui K, Matsuno H, Komatsu H, Kodama T, Katayama E. 1999. Structure and Enzymatic Properties of Genetically Truncated Forms of the Water-Insoluble Glucan-Synthesizing Glucosyltransferase from *Streptococcus sobrinus* 1. **295**: 287–295.
- Koshland DE. 1953. Stereochemistry and the mechanism of enzymatic reactions. *Biol Rev* **28**: 416–436. <https://doi.org/10.1111/j.1469-185X.1953.tb01386.x>.
- Kotloff KL, Noriega F, Losonsky GA, Sztein MB, Wasserman SS, Nataro JP, Levine MM. 1996. Safety, immunogenicity, and transmissibility in humans of CVD 1203, a live oral *Shigella flexneri* 2a vaccine candidate attenuated by deletions in *aroA* and *virG*. *Infect Immun* **64**: 4542–8.
- Kotloff KL, Noriega FR, Samandari T, Sztein MB, Losonsky GA, Nataro JP, Picking WD, Barry EM, Levine MM. 2000. *Shigella flexneri* 2a strain CVD 1207, with specific deletions in *virG*, *sen*, *set*, and *guaBA*, is highly attenuated in humans. *Infect Immun* **68**: 1034–9. <https://doi.org/10.1128/iai.68.3.1034-1039.2000>.
- Kotloff KL, Pasetti MF, Barry EM, Nataro JP, Wasserman SS, Sztein MB, Picking WD, Levine MM. 2004. Deletion in the *Shigella enterotoxin* genes further attenuates *Shigella flexneri* 2a bearing guanine auxotrophy in a phase 1 trial of CVD 1204 and CVD 1208. *J Infect Dis* **190**: 1745–54. <https://doi.org/10.1086/424680>.
- Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AKM. 2018. Shigellosis. *Lancet (London, England)* **391**: 801–812. [https://doi.org/10.1016/S0140-6736\(17\)33296-8](https://doi.org/10.1016/S0140-6736(17)33296-8).
- Kralj S, Grijpstra P, van Leeuwen SS, Leemhuis H, Dobruchowska JM, van der Kaaij RM, Malik A, Oetari A, Kamerling JP, 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**: 8154–63. <https://doi.org/10.1128/AEM.05735-11>.
- Kralj S, Stripling E, Sanders P, van Geel-Schutten GH, Dijkhuizen L. 2005. Highly hydrolytic reuteransucrase from probiotic *Lactobacillus reuteri* strain ATCC 55730. *Appl Environ Microbiol* **71**: 3942–50. <https://doi.org/10.1128/AEM.71.7.3942-3950.2005>.

- Kuhaulomlarp S, Patron NJ, Henrissat B, Rejzek M, Saalbach G, Field RA. 2018. Identification of *Euglena gracilis* β -1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149. *J Biol Chem* **293**: 2865–2876. <https://doi.org/10.1074/jbc.RA117.000936>.
- Kwan DH, Constantinescu I, Chapanian R, Higgins MA, Kötztler MP, Samain E, Boraston AB, Kizhakkedathu JN, Withers SG. 2015. Toward efficient enzymes for the generation of universal blood through structure-guided directed evolution. *J Am Chem Soc* **137**: 5695–5705. <https://doi.org/10.1021/ja5116088>.
- Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**: 521–55. <https://doi.org/10.1146/annurev.biochem.76.061005.092322>.
- Lampel KA, Formal SB, Maurelli AT. 2018. A Brief History of Shigella. *EcoSal Plus* **8**. <https://doi.org/10.1128/ecosalplus.ESP-0006-2017>.
- Lee B-Y, Seeberger PH, Varon Silva D. 2016. Synthesis of glycosylphosphatidylinositol (GPI)-anchor glycolipids bearing unsaturated lipids. *Chem Commun* **52**: 1586–1589. <https://doi.org/10.1039/C5CC07694C>.
- Levine MM, Kotloff KL, Barry EM, Pasetti MF, Sztein MB. 2007. Clinical trials of Shigella vaccines: two steps forward and one step back on a long, hard road. *Nat Rev Microbiol* **5**: 540–553. <https://doi.org/10.1038/nrmicro1662>.
- Li A, Kärrnell A, Huan PT, Cam PD, Minh NB, Trâm LN, Quy NP, Trach DD, Karlsson K, Lindberg G. 1993. Safety and immunogenicity of the live oral auxotrophic Shigella flexneri SFL124 in adult Vietnamese volunteers. *Vaccine* **11**: 180–9.
- Li C, Wang L-X. 2016. Endoglycosidases for the synthesis of polysaccharides and glycoconjugates. *Adv Carbohydr Chem Biochem* **73**: 73–116. <https://doi.org/10.1016/bs.accb.2016.07.001>.
- Li C, Wang LX. 2018. Chemoenzymatic methods for the synthesis of glycoproteins. *Chem Rev* **118**: 8359–8413. <https://doi.org/10.1021/acs.chemrev.8b00238>.
- Li J, Su G, Liu J. 2017. Enzymatic synthesis of homogeneous chondroitin sulfate oligosaccharides. *Angew Chemie - Int Ed* **56**: 11784–11787. <https://doi.org/10.1002/anie.201705638>.
- Li T, Huang M, Liu L, Wang S, Moremen KW, Boons G-J. 2016. Divergent Chemoenzymatic Synthesis of Asymmetrical-Core-Fucosylated and Core-Unmodified N -Glycans. *Chem - A Eur J* **22**: 18742–18746. <https://doi.org/10.1002/chem.201604999>.
- Li Y, Liu X. 2014. Tunable acid-sensitive ester protecting groups in oligosaccharide synthesis. *Chem Commun* **50**: 3155. <https://doi.org/10.1039/c3cc49205b>.
- Lin S-W, Yuan T, Li J, Lin C. 2006. Carboxyl terminus of *Helicobacter pylori* α 1,3-fucosyltransferase determines the structure and stability. *Biochemistry* **45**: 8108–8116. <https://doi.org/10.1021/bi0601297>.
- Linhardt RJ, Bazin HG. 2001. Separation and Purification of Carbohydrates. In *Glycoscience: Chemistry and Chemical Biology I–III*, pp. 63–74, Springer Berlin Heidelberg, Berlin, Heidelberg https://doi.org/10.1007/978-3-642-56874-9_3.
- Liu B, Knirel YA, Feng L, Perepelov A V., Senchenkova SN, Reeves PR, Wang L. 2014. Structural diversity in Salmonella O antigens and its genetic basis. *FEMS Microbiol Rev* **38**: 56–89. <https://doi.org/10.1111/1574-6976.12034>.
- Liu G, Neelamegham S. 2014. A computational framework for the automated construction of glycosylation reaction networks. *PLoS One* **9**. <https://doi.org/10.1371/journal.pone.0100939>.
- Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, Operario DJ, Uddin J, Ahmed S, Alonso PL, et al. 2016. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *Lancet (London, England)* **388**: 1291–301. [https://doi.org/10.1016/S0140-6736\(16\)31529-X](https://doi.org/10.1016/S0140-6736(16)31529-X).
- Livio S, Strockbine NA, Panchalingam S, Tennant SM, Barry EM, Marohn ME, Antonio M, Hossain A, Mandomando I, Ochieng JB, et al. 2014. Shigella isolates from the global enteric multicenter study inform vaccine development. *Clin Infect Dis* **59**: 933–941. <https://doi.org/10.1093/cid/ciu468>.

- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: 490–495. <https://doi.org/10.1093/nar/gkt1178>.
- Lu W, Zong C, Chopra P, Pepi LE, Xu Y, Amster IJ, Liu J, Boons GJ. 2018. Controlled chemoenzymatic synthesis of heparan sulfate oligosaccharides. *Angew Chemie - Int Ed* **57**: 5340–5344. <https://doi.org/10.1002/anie.201800387>.
- Ludwiczek ML, D'Angelo I, Yalloway GN, Brockerman JA, Okon M, Nielsen JE, Strynadka NCJ, Withers SG, McIntosh LP. 2013. Strategies for modulating the pH-dependent activity of a family 11 glycoside hydrolase. *Biochemistry* **52**: 3138–3156. <https://doi.org/10.1021/bi400034m>.
- Luley-Goedl C, Nidetzky B. 2010. Carbohydrate synthesis by disaccharide phosphorylases: reactions, catalytic mechanisms and application in the glycosciences. *Biotechnol J* **5**: 1324–1338. <https://doi.org/10.1002/biot.201000217>.
- Luo X, Sun Q, Lan R, Wang J, Li Z, Xia S, Zhang J, Wang Y, Jin D, Wang Y, et al. 2012. Emergence of a novel *Shigella flexneri* serotype 1d in China. *Diagn Microbiol Infect Dis* **74**: 316–319. <https://doi.org/10.1016/j.diagmicrobio.2012.06.022>.
- Macdonald SS, Patel A, Larmour VLC, Morgan-Lang C, Hallam SJ, Mark BL, Withers SG. 2018. Structural and mechanistic analysis of a β -glycoside phosphorylase identified by screening a metagenomic library. *J Biol Chem* **293**: 3451–3467. <https://doi.org/10.1074/jbc.RA117.000948>.
- Machius M, Declerck N, Huber R, Wiegand G. 1998. Activation of *Bacillus licheniformis* alpha-amylase through a disorder-order transition of the substrate-binding site mediated by a calcium-sodium-calcium metal triad. *Structure* **6**: 281–92.
- Mackenzie LF, Wang Q, Warren RAJ, Withers SG. 1998. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J Am Chem Soc* **120**: 5583–5584. <https://doi.org/10.1021/ja980833d>.
- Mak WS, Tran S, Marcheschi R, Bertolani S, Thompson J, Baker D, Liao JC, Siegel JB. 2015. Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat Commun* **6**: 10005. <https://doi.org/10.1038/ncomms10005>.
- Mäkeläinen H, Hasselwander O, Rautonen N, Ouwehand AC. 2009. Panose, a new prebiotic candidate. *Lett Appl Microbiol* **49**: 666–672. <https://doi.org/10.1111/j.1472-765X.2009.02698.x>.
- Malakauskas SM, Mayo SL. 1998. Hyperthermophilic Protein Variant. **5**: 1–6.
- Malbert Y, Moulis C, Brison Y, Morel S, André I, Remaud-Simeon M. 2018. Engineering a branching sucrase for flavonoid glucoside diversification. *Sci Rep* **8**: 15153. <https://doi.org/10.1038/s41598-018-33394-y>.
- 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**: 208–212. [https://doi.org/10.1016/S0014-5793\(98\)01448-3](https://doi.org/10.1016/S0014-5793(98)01448-3).
- Mann J, Morenga LT. 2015. Carbohydrates in the treatment and prevention of Type 2 diabetes. *Diabet Med* **32**: 572–575. <https://doi.org/10.1111/dme.12673>.
- Mano MCR, Neri-Numa IA, da Silva JB, Paulino BN, Pessoa MG, Pastore GM. 2018. Oligosaccharide biotechnology: an approach of prebiotic revolution on the industry. *Appl Microbiol Biotechnol* **102**: 17–37. <https://doi.org/10.1007/s00253-017-8564-2>.
- Marbouty M, Cournac A, Flot J-F, Marie-Nelly H, Mozziconacci J, Koszul R. 2014. Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *Elife* **3**. <https://doi.org/10.7554/eLife.03318>.
- Marcus SL, Polakowski R, Seto NOL, Leinala E, Borisova S, Blancher A, Roubinet F, Evans S V., Palcic MM. 2003. A single point mutation reverses the donor specificity of human blood group B-synthesizing galactosyltransferase. *J Biol Chem* **278**: 12403–12405. <https://doi.org/10.1074/jbc.M212002200>.
- McCarthy PC, Sharyan A, Sheikhi Moghaddam L. 2018. Meningococcal Vaccines: Current Status and Emerging Strategies. *Vaccines* **6**: 12. <https://doi.org/10.3390/vaccines6010012>.
- Mel D, Gangarosa EJ, Radovanovic ML, Arsic BL, Litvinjenko S. 1971. Studies on vaccination against

- bacillary dysentery. 6. Protection of children by oral immunization with streptomycin-dependent *Shigella* strains. *Bull World Health Organ* **45**: 457–64.
- Meng X, Dobruchowska JM, Pijning T, López CA, Kamerling JP, Dijkhuizen L. 2014. Residue Leu940 has a crucial role in the linkage and reaction specificity of the glucansucrase GTF180 of the probiotic bacterium *Lactobacillus reuteri* 180. *J Biol Chem* **289**: 32773–82. <https://doi.org/10.1074/jbc.M114.602524>.
- Meng X, Gangoiti J, Bai Y, Pijning T. 2016. Structure – function relationships of family GH70 glucansucrase and 4 , 6- a -glucanotransferase enzymes , and their evolutionary relationships with family GH13 enzymes. *Cell Mol Life Sci* **73**: 2681–2706. <https://doi.org/10.1007/s00018-016-2245-7>.
- Meng X, Gangoiti J, Wang X, Grijpstra P, van Leeuwen SS, Pijning T, Dijkhuizen L. 2018. Biochemical characterization of a GH70 protein from *Lactobacillus kunkeei* DSM 12361 with two catalytic domains involving branching sucrose activity. *Appl Microbiol Biotechnol* **102**: 7935–7950. <https://doi.org/10.1007/s00253-018-9236-6>.
- Meng X, Pijning T, Dobruchowska JM, Gerwig GJ, Dijkhuizen L. 2015. Characterization of the Functional Roles of Amino Acid Residues in Acceptor-binding Subsite +1 in the Active Site of the Glucansucrase GTF180 from *Lactobacillus reuteri* 180. *J Biol Chem* **290**: 30131–30141. <https://doi.org/10.1074/jbc.M115.687558>.
- Meng X, Pijning T, Tietema M, Dobruchowska JM, Yin H, Gerwig GJ, Kralj S, Dijkhuizen L. 2017. Characterization of the glucansucrase GTF180 W1065 mutant enzymes producing polysaccharides and oligosaccharides with altered linkage composition. **217**: 81–90. <https://doi.org/10.1016/j.foodchem.2016.08.087>.
- Micoli F, Adamo R, Costantino P. 2018a. Protein carriers for glycoconjugate vaccines: History, selection criteria, characterization and new trends. *Molecules* **23**: 1–18. <https://doi.org/10.3390/molecules23061451>.
- Micoli F, Costantino P, Adamo R. 2018b. Potential targets for next generation antimicrobial glycoconjugate vaccines. *FEMS Microbiol Rev* **42**: 388–423. <https://doi.org/10.1093/femsre/fuy011>.
- Miller GL. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal Chem* **31**: 426–428. <https://doi.org/10.1021/ac60147a030>.
- Molina M. 2019. Exploration of the molecular determinants involved in alternansucrase specificity and stability. Institut National des Sciences Appliquées de Toulouse.
- Molina M, Moulis C, Monties N, Pizzut-Serin S, Guieysse D, Morel S, Cioci G, Remaud-Siméon M. 2019. Deciphering an Undecided Enzyme: Investigations of the Structural Determinants Involved in the Linkage Specificity of Alternansucrase. *ACS Catal* **9**: 2222–2237. <https://doi.org/10.1021/acscatal.8b04510>.
- Monchois V, Remaud-Simeon M, Monsan P, Willemot RM. 1998a. Cloning and sequencing of a gene coding for an extracellular dextransucrase (DSRB) from *Leuconostoc mesenteroides* NRRL B-1299 synthesizing only a alpha (1-6) glucan. *FEMS Microbiol Lett* **159**: 307–15. <https://doi.org/10.1111/j.1574-6968.1998.tb12876.x>.
- Monchois V, Remaud-Simeon M, Russell RRB, Monsan P, Willemot RM. 1997. Characterization of *Leuconostoc mesenteroides* NRRL B-512F dextransucrase (DSRS) and identification of amino-acid residues playing a key role in enzyme activity. *Appl Microbiol Biotechnol* **48**: 465–472. <https://doi.org/10.1007/s002530051081>.
- Monchois V, Reverte A, Remaud-Simeon M, Monsan P, Willemot RM. 1998b. Effect of *Leuconostoc mesenteroides* NRRL B-512F dextransucrase carboxy- terminal deletions on dextran and oligosaccharide synthesis. *Appl Environ Microbiol* **64**: 1644–1649.
- Monchois V, Willemot RM, Monsan P. 1999. Glucansucrases: mechanism of action and structure-function relationships. *FEMS Microbiol Rev* **23**: 131–51. <https://doi.org/10.1111/j.1574-6976.1999.tb00394.x>.
- Monsan P, Paul F. 1995. Enzymatic synthesis of oligosaccharides. *FEMS Microbiol Rev* **16**: 187–192. <https://doi.org/10.1111/j.1574-6976.1995.tb00165.x>.

- Monsan P, Remaud-Siméon M, André I. 2010. Transglucosidases as efficient tools for oligosaccharide and glucoconjugate synthesis. *Curr Opin Microbiol* **13**: 293–300. <https://doi.org/10.1016/j.mib.2010.03.002>.
- Moon YH, Kim G, Lee JH, Jin XJ, Kim DW, Kim D. 2006. Enzymatic synthesis and characterization of novel epigallocatechin gallate glucosides. *J Mol Catal B Enzym* **40**: 1–7. <https://doi.org/10.1016/j.molcatb.2006.01.030>.
- Moon YH, Lee JH, Jhon DY, Jun WJ, Kang SS, Sim J, Choi H, Moon JH, Kim D. 2007. Synthesis and characterization of novel quercetin- α -d-glucopyranosides using glucansucrase from *Leuconostoc mesenteroides*. *Enzyme Microb Technol* **40**: 1124–1129. <https://doi.org/10.1016/j.enzmictec.2006.08.019>.
- Moracci M, Trincone A, Perugino G, Ciaramella M, Rossi M. 1998. Restoration of the activity of active-site mutants of the hyperthermophilic β -glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles. *Biochemistry* **37**: 17262–17270. <https://doi.org/10.1021/bi981855f>.
- Morona R, Bosch L. 2003. Lipopolysaccharide O antigen chains mask IcsA (VirG) in *Shigella flexneri*. *FEMS Microbiol Lett* **221**: 173–180. [https://doi.org/10.1016/S0378-1097\(03\)00210-6](https://doi.org/10.1016/S0378-1097(03)00210-6).
- Moulis C, André I, Remaud-Simeon M. 2016. GH13 amylosucrases and GH70 branching sucrases, atypical enzymes in their respective families. *Cell Mol Life Sci* **73**: 2661–2679. <https://doi.org/10.1007/s00018-016-2244-8>.
- Moulis C, Arcache A, Escalier PC, Rinaudo M, Monsan P, Remaud-Simeon M, Potocki-Veronese G. 2006a. High-level production and purification of a fully active recombinant dextranucrase from *Leuconostoc mesenteroides* NRRL B-512F. *FEMS Microbiol Lett* **261**: 203–210. <https://doi.org/10.1111/j.1574-6968.2006.00347.x>.
- Moulis C, Joucla G, Harrison D, Fabre E, Potocki-Veronese G, Monsan P, Remaud-Simeon M. 2006b. Understanding the polymerization mechanism of glycoside-hydrolase family 70 glucansucrases. *J Biol Chem* **281**: 31254–31267. <https://doi.org/10.1074/jbc.M604850200>.
- Mukherjee K, Narindoshvili T, Raushel FM. 2018. Discovery of a kojibiose phosphorylase in *Escherichia coli* K-12. *Biochemistry* **57**: 2857–2867. <https://doi.org/10.1021/acs.biochem.8b00392>.
- Mulard LA, Clément M-J, Imbert A, Delepierre M. 2002. Convergent Synthesis, NMR and Conformational Analysis of Tetra- and Pentasaccharide Haptens of the *Shigella flexneri* Serotype 5a O-Specific Polysaccharide. *European J Org Chem* **2002**: 2486. [https://doi.org/10.1002/1099-0690\(200208\)2002:15<2486::AID-EJOC2486>3.0.CO;2-E](https://doi.org/10.1002/1099-0690(200208)2002:15<2486::AID-EJOC2486>3.0.CO;2-E).
- Mulard LA, Costachel C, Sansonetti PJ. 2000. Synthesis of the Methyl Glycosides of a Di- and Two Trisaccharide Fragments Specific for the *Shigella flexneri* Serotype 2a O -Antigen. *J Carbohydr Chem* **19**: 849–877. <https://doi.org/10.1080/07328300008544123>.
- Mundhada H, Marienhagen J, Scacioc A, Schenk A, Roccatano D, Schwaneberg U. 2011. SeSaM-Tv-II generates a protein sequence space that is unobtainable by epPCR. *ChemBioChem* **12**: 1595–1601. <https://doi.org/10.1002/cbic.201100010>.
- Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL, Baker D. 2009. Alteration of enzyme specificity by computational loop remodeling and design. *Proc Natl Acad Sci U S A* **106**: 9215–9220. <https://doi.org/10.1073/pnas.0811070106>.
- Nakai H, Kitaoka M, Svensson B, Ohtsubo K. 2013. Recent development of phosphorylases possessing large potential for oligosaccharide synthesis. *Curr Opin Chem Biol* **17**: 301–309. <https://doi.org/10.1016/j.cbpa.2013.01.006>.
- Nakai H, Petersen BO, Westphal Y, Dilokpimol A, Abou Hachem M, Duus JO, Schols HA, Svensson B. 2010. Rational engineering of *Lactobacillus acidophilus* NCFM maltose phosphorylase into either trehalose or kojibiose dual specificity phosphorylase. *Protein Eng Des Sel* **23**: 781–787. <https://doi.org/10.1093/protein/gzq055>.
- Nam S-H, Park J, Jun W, Kim D, Ko J-A, Abd El-Aty AM, Choi JY, Kim D-I, Yang K-Y. 2017. Transglycosylation of gallic acid by using *Leuconostoc* glucansucrase and its characterization as a functional cosmetic agent. *AMB Express* **7**: 224. <https://doi.org/10.1186/s13568-017-0523-x>.

- Nam SH, Ko EA, Jang SS, Kim DW, Kim SY, Hwang DS, Kim D. 2008. Maximization of dextranucrase activity expressed in *E. coli* by mutation and its functional characterization. *Biotechnol Lett* **30**: 135–143. <https://doi.org/10.1007/s10529-007-9498-z>.
- Newburg DS, Ruiz-Palacios GM, Morrow AL. 2005. Human milk glycans protect infants against enteric pathogens. *Annu Rev Nutr* **25**: 37–58. <https://doi.org/10.1146/annurev.nutr.25.050304.092553>.
- Nidetzky B, Gutmann A, Zhong C. 2018. Leloir glycosyltransferases as biocatalysts for chemical production. *ACS Catal* **8**: 6283–6300. <https://doi.org/10.1021/acscatal.8b00710>.
- Nielsen MM, Pedersen CM. 2018. Catalytic glycosylations in oligosaccharide synthesis. *Chem Rev* **118**: 8285–8358. <https://doi.org/10.1021/acs.chemrev.8b00144>.
- Niggemann J, Kamerling JP, Vliegenthart JFG. 1998. β -1,4-galactosyltransferase-catalyzed synthesis of the branched tetrasaccharide repeating unit of *Streptococcus pneumoniae* type 14. *Bioorg Med Chem* **6**: 1605–1612. [https://doi.org/10.1016/S0968-0896\(98\)00095-9](https://doi.org/10.1016/S0968-0896(98)00095-9).
- Nigudkar SS, Demchenko A V. 2015. Stereocontrolled 1,2-cis glycosylation as the driving force of progress in synthetic carbohydrate chemistry. *Chem Sci* **6**: 2687–2704. <https://doi.org/10.1039/C5SC00280J>.
- Nishihara K, Kanemori M, Yanagi H, Yura T. 2000. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl Environ Microbiol* **66**: 884–9. <https://doi.org/10.1128/aem.66.3.884-889.2000>.
- Nishimoto M, Kitaoka M. 2007. Practical preparation of lacto-N-biose I, a candidate for the bifidus factor in human milk. *Biosci Biotechnol Biochem* **71**: 2101–4. <https://doi.org/10.1271/bbb.70320>.
- Noriega FR, Liao FM, Maneval DR, Ren S, Formal SB, Levine MM. 1999. Strategy for cross-protection among *Shigella flexneri* serotypes. *Infect Immun* **67**: 782–8.
- Nycholat CM, Peng W, McBride R, Antonopoulos A, de Vries RP, Polonskaya Z, Finn MG, Dell A, Haslam SM, Paulson JC. 2013. Synthesis of biologically active N- and O-linked glycans with multisialylated poly-N-acetylglucosamine extensions using *P. damsela* α -6 sialyltransferase. *J Am Chem Soc* **135**: 18280–18283. <https://doi.org/10.1021/ja409781c>.
- Nyffenegger C, Nordvang RT, Jers C, Meyer AS, Mikkelsen JD. 2017. Design of *Trypanosoma rangeli* sialidase mutants with improved trans-sialidase activity. *PLoS One* **12**: 1–11. <https://doi.org/10.1371/journal.pone.0171585>.
- O'Neill EC, Field RA. 2015. Enzymatic synthesis using glycoside phosphorylases. *Carbohydr Res* **403**: 23–37. <https://doi.org/10.1016/j.carres.2014.06.010>.
- Oldrini D, Fiebig T, Romano MR, Proietti D, Berger M, Tontini M, De Ricco R, Santini L, Morelli L, Lay L, et al. 2018. Combined chemical synthesis and tailored enzymatic elongation provide fully synthetic and conjugation-ready *Neisseria meningitidis* serogroup X vaccine antigens. *ACS Chem Biol* **13**: 984–994. <https://doi.org/10.1021/acscchembio.7b01057>.
- Ollis AA, Zhang S, Fisher AC, DeLisa MP. 2014. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat Chem Biol* **10**: 816–822. <https://doi.org/10.1038/nchembio.1609>.
- Onufriev A, Bashford D, Case DA. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**: 383–394. <https://doi.org/10.1002/prot.20033>.
- 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. <https://doi.org/10.1021/bi061444w>.
- Oscarson S, Tedebark U. 1996. Synthesis and Acidic Opening of Chlorinated Carbohydrate Orthoacetates. *J Carbohydr Chem* **15**: 507–512. <https://doi.org/10.1080/07328309608005670>.
- Osorio MI, Zúñiga MA, Mendoza F, Jaña GA, Jiménez VA. 2019. Modulation of glucan-enzyme interactions by domain V in GTF-SI from *Streptococcus mutans*. *Proteins* **87**: 74–80. <https://doi.org/10.1002/prot.25624>.
- Packer MS, Liu DR. 2015. Methods for the directed evolution of proteins. *Nat Rev Genet* **16**: 379–394.

<https://doi.org/10.1038/nrg3927>.

- Palcic MM. 1999. Biocatalytic synthesis of oligosaccharides. *Curr Opin Biotechnol* **10**: 616–624. [https://doi.org/10.1016/S0958-1669\(99\)00044-0](https://doi.org/10.1016/S0958-1669(99)00044-0).
- Palcic MM. 2011. Glycosyltransferases as biocatalysts. *Curr Opin Chem Biol* **15**: 226–233. <https://doi.org/10.1016/j.cbpa.2010.11.022>.
- Paris G, Cremona ML, Amaya MF, Buschiazzi A, Giambiagi S, Frasch ACC, Alzari PM. 2001. Probing molecular function of trypanosomal sialidases: single point mutations can change substrate specificity and increase hydrolytic activity. *Glycobiology* **11**: 305–311. <https://doi.org/10.1093/glycob/11.4.305>.
- Paris G, Ratier L, Amaya MF, Nguyen T, Alzari PM, Frasch ACC. 2005. A sialidase mutant displaying trans-sialidase activity. *J Mol Biol* **345**: 923–934. <https://doi.org/10.1016/j.jmb.2004.09.031>.
- Park BH, Karpinets T V, Syed MH, Leuze MR, Uberbacher EC. 2010. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20**: 1574–1584. <https://doi.org/10.1093/glycob/cwq106>.
- Passerini D, Vuillemin M, Ufarté L, Morel S, Loux V, Fontagné-Faucher C, Monsan P, Remaud-Siméon M, Moulis C. 2015. Inventory of the GH70 enzymes encoded by *Leuconostoc citreum* NRRL B-1299 - Identification of three novel α -transglucosylases. *FEBS J* **282**: 2115–2130. <https://doi.org/10.1111/febs.13261>.
- Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, Ferguson D, Seibel G, Kollman P. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun* **91**: 1–41. [https://doi.org/10.1016/0010-4655\(95\)00041-D](https://doi.org/10.1016/0010-4655(95)00041-D).
- Pell J, Hintze A, Canino-Koning R, Howe A, Tiedje JM, Brown CT. 2012. Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. *Proc Natl Acad Sci* **109**: 13272–13277. <https://doi.org/10.1073/pnas.1121464109>.
- Perepelov A V., L'vov VL, Liu B, Senchenkova SN, Shekht ME, Shashkov AS, Feng L, Aparin PG, Wang L, Knirel YA. 2009. A new ethanolamine phosphate-containing variant of the O-antigen of *Shigella flexneri* type 4a. *Carbohydr Res* **344**: 1588–1591. <https://doi.org/10.1016/j.carres.2009.03.022>.
- Perepelov A V., Shekht ME, Liu B, Shevelev SD, Ledov VA, Senchenkova SN, L'vov VL, Shashkov AS, Feng L, Aparin PG, et al. 2012. *Shigella flexneri* O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. *FEMS Immunol Med Microbiol* **66**: 201–210. <https://doi.org/10.1111/j.1574-695X.2012.01000.x>.
- Pergolizzi G, Kuhaudomlarp S, Kalita E, Field RA. 2017. Glycan phosphorylases in multi-enzyme synthetic processes. *Protein Pept Lett* **24**: 696–709. <https://doi.org/10.2174/0929866524666170811125109>.
- Petschacher B, Nidetzky B. 2016. Biotechnological production of fucosylated human milk oligosaccharides: prokaryotic fucosyltransferases and their use in biocatalytic cascades or whole cell conversion systems. *J Biotechnol* **235**: 61–83. <https://doi.org/10.1016/j.jbiotec.2016.03.052>.
- Phalipon A, Costachel C, Grandjean C, Thuizat A, Guerreiro C, Tanguy M, Nato F, Vulliez-Le Normand B, Bélot F, Wright K, et al. 2006. Characterization of functional oligosaccharide mimics of the *Shigella flexneri* serotype 2a O-antigen: implications for the development of a chemically defined glycoconjugate vaccine. *J Immunol* **176**: 1686–94. <https://doi.org/10.1093/infdis/jii176.3.1686> [pii].
- Phalipon A, Tanguy M, Grandjean C, Guerreiro C, Bélot F, Cohen D, Sansonetti PJ, Mulard LA. 2009. A synthetic carbohydrate-protein conjugate vaccine candidate against *Shigella flexneri* 2a infection. *J Immunol* **182**: 2241–7. <https://doi.org/10.4049/jimmunol.0803141>.
- Pham H, Boger MCL, Dijkhuizen L, van Leeuwen SS. 2019. Stimulatory effects of novel glucosylated lactose derivatives GL34 on growth of selected gut bacteria. *Appl Microbiol Biotechnol* **103**: 707–718. <https://doi.org/10.1007/s00253-018-9473-8>.
- Pham H, Dijkhuizen L, van Leeuwen SS. 2017. Structural characterization of glucosylated lactose

- derivatives synthesized by the *Lactobacillus reuteri* GtfA and Gtf180 glucansucrase enzymes. *Carbohydr Res* **449**: 59–64. <https://doi.org/10.1016/j.carres.2017.07.002>.
- Pham H, Pijning T, Dijkhuizen L, Van Leeuwen SS. 2018. Mutational Analysis of the Role of the Glucansucrase Gtf180- Δ N Active Site Residues in Product and Linkage Specificity with Lactose as Acceptor Substrate. *J Agric Food Chem* **66**: 12544–12554. <https://doi.org/10.1021/acs.jafc.8b04486>.
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. 2005. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**: 1781–802. <https://doi.org/10.1002/jcc.20289>.
- Pierdominici-Sottile G, Palma J, Roitberg AE. 2014. Free-energy computations identify the mutations required to confer trans-sialidase activity into *Trypanosoma rangeli* sialidase. *Proteins Struct Funct Bioinforma* **82**: 424–435. <https://doi.org/10.1002/prot.24408>.
- Pijning T, Vujičić-Žagar A, Kralj S, Dijkhuizen L, Dijkstra BW. 2014. Flexibility of truncated and full-length glucansucrase GTF180 enzymes from *Lactobacillus reuteri* 180. *FEBS J* **281**: 2159–2171. <https://doi.org/10.1111/febs.12769>.
- Pijning T, Vujičić-Žagar A, Kralj S, Dijkhuizen L, Dijkstra BW. 2012. Structure of the α -1,6/ α -1,4-specific glucansucrase GTFA from *Lactobacillus reuteri* 121. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**: 1448–1454. <https://doi.org/10.1107/S1744309112044168>.
- Platts-Mills JA, Babji S, Bodhidatta L, Gratz J, Haque R, Havt A, McCormick BJ, McGrath M, Olortegui MP, Samie A, et al. 2015. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet Glob Heal* **3**: e564-75. [https://doi.org/10.1016/S2214-109X\(15\)00151-5](https://doi.org/10.1016/S2214-109X(15)00151-5).
- Plotkin SA. 2005. Vaccines: past, present and future. *Nat Med* **11**: S5-11. <https://doi.org/10.1038/nm1209>.
- Polotsky VY, Robbins JB, Bryla D, Schneerson R. 1994. Comparison of conjugates composed of lipopolysaccharide from *Shigella flexneri* type 2a detoxified by two methods and bound to tetanus toxoid. *Infect Immun* **62**: 210–4.
- Porras-Domínguez JR, Rodríguez-Alegría ME, Ávila-Fernández Á, Montiel-Salgado S, López-Munguía A. 2017. Levan-type fructooligosaccharides synthesis by a levansucrase-endolevanase fusion enzyme (LevB1SacB). *Carbohydr Polym* **177**: 40–48. <https://doi.org/10.1016/j.carbpol.2017.08.040>.
- Pozsgay V, Chu C, Pannell L, Wolfe J, Robbins JB, Schneerson R. 1999. Protein conjugates of synthetic saccharides elicit higher levels of serum IgG lipopolysaccharide antibodies in mice than do those of the O-specific polysaccharide from *Shigella dysenteriae* type 1. *Proc Natl Acad Sci U S A* **96**: 5194–7. <https://doi.org/10.1073/pnas.96.9.5194>.
- Prapulla SG, Subhaprada V, Karanth NG. 2000. Microbial production of oligosaccharides: a review. *Adv Appl Microbiol* **47**: 299–343. [https://doi.org/10.1016/S0065-2164\(00\)47008-5](https://doi.org/10.1016/S0065-2164(00)47008-5).
- Priem B, Gilbert M, Wakarchuk WW, Heyraud A, Samain E. 2002. A new fermentation process allows large-scale production of human milk oligosaccharides by metabolically engineered bacteria. *Glycobiology* **12**: 235–240. <https://doi.org/10.1093/glycob/12.4.235>.
- Puchart V. 2015. Glycoside phosphorylases: structure, catalytic properties and biotechnological potential. *Biotechnol Adv* **33**: 261–276. <https://doi.org/10.1016/j.biotechadv.2015.02.002>.
- Puzari M, Sharma M, Chetia P. 2018. Emergence of antibiotic resistant *Shigella* species: A matter of concern. *J Infect Public Health* **11**: 451–454. <https://doi.org/10.1016/j.jiph.2017.09.025>.
- Qiu S, Wang Z, Chen C, Liu N, Jia L, Liu W, Wang L, Hao R, Zhang L, Wang Y, et al. 2011. Emergence of a novel *Shigella flexneri* serotype 4s strain that evolved from a serotype X variant in China. *J Clin Microbiol* **49**: 1148–1150. <https://doi.org/10.1128/JCM.01946-10>.
- Qiu S, Xu X, Wang Y, Yang G, Wang Z, Wang H, Zhang L, Liu N, Chen C, Liu W, et al. 2012. Emergence of resistance to fluoroquinolones and third-generation cephalosporins in *Shigella flexneri* subserotype 1c isolates from China. *Clin Microbiol Infect* **18**. <https://doi.org/10.1111/j.1469-0691.2012.03768.x>.

- R Development Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna.
- Rance M, Sørensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. 1983. Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun* **117**: 479–485. [https://doi.org/10.1016/0006-291X\(83\)91225-1](https://doi.org/10.1016/0006-291X(83)91225-1).
- Reetz MT, Bocola M, Carballeira JD, Zha D, Vogel A. 2005. Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew Chemie - Int Ed* **44**: 4192–4196. <https://doi.org/10.1002/anie.200500767>.
- Reetz MT, Carballeira JD. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc* **2**: 891–903. <https://doi.org/10.1038/nprot.2007.72>.
- Reeve SM, Gainza P, Frey KM, Georgiev I, Donald BR, Anderson AC. 2015. Protein design algorithms predict viable resistance to an experimental antifolate. *Proc Natl Acad Sci* **112**: 749–754. <https://doi.org/10.1073/pnas.1411548112>.
- Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D. 2011. De novo enzyme design using Rosetta3. *PLoS One* **6**: e19230. <https://doi.org/10.1371/journal.pone.0019230>.
- Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**: W320–W324. <https://doi.org/10.1093/nar/gku316>.
- Roe DR, Cheatham TE. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9**: 3084–95. <https://doi.org/10.1021/ct400341p>.
- Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, Dechancie J, Betker J, Gallaher JL, Althoff E, Zanghellini A, Dym O, et al. 2008. Kemp elimination catalysts by computational enzyme design. *Nature* **453**: 190–5. <https://doi.org/10.1038/nature06879>.
- Ruffing A, Chen RR. 2006. Metabolic engineering of microbes for oligosaccharide and polysaccharide synthesis. *Microb Cell Fact* **5**: 1–9. <https://doi.org/10.1186/1475-2859-5-25>.
- Sabri S, Nielsen LK, Vickers CE. 2013a. Molecular control of sucrose utilization in Escherichia coli W, an efficient sucrose-utilizing strain. *Appl Environ Microbiol* **79**: 478–87. <https://doi.org/10.1128/AEM.02544-12>.
- Sabri S, Steen JA, Bongers M, Nielsen LK, Vickers CE. 2013b. Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the Escherichia coli chromosome at well-characterised loci. *Microb Cell Fact* **12**: 60. <https://doi.org/10.1186/1475-2859-12-60>.
- Said Hassane F, Phalipon A, Tanguy M, Guerreiro C, Bélot F, Frisch B, Mulard LA, Schuber F. 2009. Rational design and immunogenicity of liposome-based diepitope constructs: application to synthetic oligosaccharides mimicking the Shigella flexneri 2a O-antigen. *Vaccine* **27**: 5419–26. <https://doi.org/10.1016/j.vaccine.2009.06.031>.
- Sakaguchi K, Katoh T, Yamamoto K. 2016. Transglycosidase-like activity of *Mucor hiemalis* endoglycosidase mutants enabling the synthesis of glycoconjugates using a natural glycan donor. *Biotechnol Appl Biochem* **63**: 812–819. <https://doi.org/10.1002/bab.1433>.
- Sakurama H, Fushinobu S, Hidaka M, Yoshida E, Honda Y, Ashida H, Kitaoka M, Kumagai H, Yamamoto K, Katayama T. 2012. 1,3-1,4- α -L-Fucosyltransferase that specifically introduces Lewis a/x antigens into type-1/2 chains. *J Biol Chem* **287**: 16709–16719. <https://doi.org/10.1074/jbc.M111.333781>.
- Salamone S, Guerreiro C, Cambon E, André I, Remaud-Siméon M, Mulard LA. 2015a. Programmed chemo-enzymatic synthesis of the oligosaccharide component of a carbohydrate-based antibacterial vaccine candidate. *Chem Commun* **51**: 2581–2584. <https://doi.org/10.1039/c4cc08805k>.
- Salamone S, Guerreiro C, Cambon E, Hargreaves JM, Tarrat N, Remaud-Siméon M, André I, Mulard LA. 2015b. Investigation on the Synthesis of Shigella flexneri Specific Oligosaccharides Using Disaccharides as Potential Transglucosylase Acceptor Substrates. *J Org Chem* **80**: 11237–11257. <https://doi.org/10.1021/acs.joc.5b01407>.

- Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. 2013. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **9**: 3878–88. <https://doi.org/10.1021/ct400314y>.
- Sánchez-Rodríguez A, Tytgat HLP, Winderickx J, Vanderleyden J, Lebeer S, Marchal K. 2014. A network-based approach to identify substrate classes of bacterial glycosyltransferases. *BMC Genomics* **15**. <https://doi.org/10.1186/1471-2164-15-349>.
- Sanz ML, Polemis N, Morales V, Corzo N, Drakoularakou A, Gibson GR, Rastall RA. 2005. In vitro investigation into the potential prebiotic activity of honey oligosaccharides. *J Agric Food Chem* **53**: 2914–2921. <https://doi.org/10.1021/jf0500684>.
- Saumonneau A, Champion E, Peltier-Pain P, Molnar-Gabor D, Hendrickx J, Tran V, Hederos M, Dekany G, Tellier C. 2015. Design of an α -L-transfucosidase for the synthesis of fucosylated HMOs. *Glycobiology* **26**: 261–269. <https://doi.org/10.1093/glycob/cwv099>.
- Schmaltz RM, Hanson SR, Wong CH. 2011. Enzymes in the synthesis of glycoconjugates. *Chem Rev* **111**: 4259–4307. <https://doi.org/10.1021/cr200113w>.
- Schmid J. 2018. Recent insights in microbial exopolysaccharide biosynthesis and engineering strategies. *Curr Opin Biotechnol* **53**: 130–136. <https://doi.org/10.1016/j.copbio.2018.01.005>.
- Schmolzer K, Czabany T, Luley-Goedl C, Pavkov-Keller T, Ribitsch D, Schwab H, Gruber K, Weber H, Nidetzky B. 2015. Complete switch from α -2,3- to α -2,6-regioselectivity in *Pasteurella dagmatis* β -D-galactoside sialyltransferase by active-site redesign. *Chem Commun* **51**: 3083–3086. <https://doi.org/10.1039/c4cc09772f>.
- Sethi A, Eargle J, Black AA, Luthey-Schulten Z. 2009. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci* **106**: 6620–6625. <https://doi.org/10.1073/pnas.0810961106>.
- Seto NO., Compston CA, Szpacenko A, Palcic MM. 2000. Enzymatic synthesis of blood group A and B trisaccharide analogues. *Carbohydr Res* **324**: 161–169. [https://doi.org/10.1016/S0008-6215\(99\)00297-9](https://doi.org/10.1016/S0008-6215(99)00297-9).
- Seto NOL, Palcic MM, Compston CA, Li H, Bundle DR, Narang SA. 1997. Sequential interchange of four amino acids from blood group B to blood group A glycosyltransferase boosts catalytic activity and progressively modifies substrate recognition in human recombinant enzymes. *J Biol Chem* **272**: 14133–14138. <https://doi.org/10.1074/jbc.272.22.14133>.
- Shahnaij M, Latif HA, Azmi IJ, Amin MB, Luna SJ, Islam MA, Talukder KA. 2018. Characterization of a serologically atypical *Shigella flexneri* Z isolated from diarrheal patients in Bangladesh and a proposed serological scheme for *Shigella flexneri*. *PLoS One* **13**: 1–16. <https://doi.org/10.1371/journal.pone.0202704>.
- Shashkov AS, Senchenkova SN, Sun Q, Lan R, Wang J, Perepelov A V, Knirel YA, Xu J. 2013. Structure of the O-antigen of a novel *Shigella flexneri* serotype, 1d (I: 7,8). *Carbohydr Res* **373**: 93–96. <https://doi.org/10.1016/j.carres.2013.03.015>.
- Siegel JB, Smith AL, Poust S, Wargacki AJ, Bar-Even A, Louw C, Shen BW, Eiben CB, Tran HM, Noor E, et al. 2015. Computational protein design enables a novel one-carbon assimilation pathway. *Proc Natl Acad Sci U S A* **112**: 3704–3709. <https://doi.org/10.1073/pnas.1500545112>.
- Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, et al. 2010. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**: 309–13. <https://doi.org/10.1126/science.1190239>.
- Smith MB. 1999. Protection for the Amino Group. In *Protective Groups in Organic Synthesis*, pp. 494–653, John Wiley & Sons, Inc., New York, USA <https://doi.org/10.1002/0471220574.ch7>.
- Smith MR, Khara E, Wen F. 2015. Engineering novel and improved biocatalysts by cell surface display. *Ind Eng Chem Res* **54**: 4021–4032. <https://doi.org/10.1021/ie504071f>.
- Sobhany M, Kakuta Y, Sugiura N, Kimata K, Negishi M. 2012. The structural basis for a coordinated reaction catalyzed by a bifunctional glycosyltransferase in chondroitin biosynthesis. *J Biol Chem* **287**: 36022–36028. <https://doi.org/10.1074/jbc.M112.375873>.
- Sprenger GA, Baumgärtner F, Albermann C. 2017. Production of human milk oligosaccharides by enzymatic and whole-cell microbial biotransformations. *J Biotechnol* **258**: 79–91.

<https://doi.org/10.1016/j.jbiotec.2017.07.030>.

- Stanley P, D. Cummings R. 2017. Structures common to different glycans. In *Essentials of Glycobiology [Internet] 3rd edition*. (eds. A. Varki, R. D. Cummings, J. Esko, and E. Al.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press
<https://doi.org/10.1101/glycobiology.3e.014>.
- Stemmer WP. 1994. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc Natl Acad Sci* **91**: 10747–51. <https://doi.org/10.1111/j.1399-6576.2007.01374.x>.
- Stenutz R, Weintraub A, Widmalm G. 2006. The structures of Escherichia coli O-polysaccharide antigens. *FEMS Microbiol Rev* **30**: 382–403. <https://doi.org/10.1111/j.1574-6976.2006.00016.x>.
- Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures. **41**: 207–234. <https://doi.org/10.1016/j.pep.2005.01.016>.
- Suflita M, Fu L, He W, Koffas M, Linhardt RJ. 2015. Heparin and related polysaccharides: synthesis using recombinant enzymes and metabolic engineering. *Appl Microbiol Biotechnol* **99**: 7465–7479. <https://doi.org/10.1007/s00253-015-6821-9>.
- Swarts BM, Guo Z. 2010. Synthesis of a Glycosylphosphatidylinositol Anchor Bearing Unsaturated Lipid Chains. *J Am Chem Soc* **132**: 6648–6650. <https://doi.org/10.1021/ja1009037>.
- Tagliabue A, Rappuoli R. 2018. Changing Priorities in Vaccinology: Antibiotic Resistance Moving to the Top. *Front Immunol* **9**: 1068. <https://doi.org/10.3389/fimmu.2018.01068>.
- Tamesada M, Kawabata S, Fujiwara T, Hamada S. 2004. Synergistic effects of streptococcal glucosyltransferases on adhesive biofilm formation. *J Dent Res* **83**: 874–9. <https://doi.org/10.1177/154405910408301110>.
- Tanaka SI, Takahashi T, Koide A, Ishihara S, Koikeda S, Koide S. 2015. Monobody-mediated alteration of enzyme specificity. *Nat Chem Biol* **11**: 762–764. <https://doi.org/10.1038/nchembio.1896>.
- 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 beta-glycosidase into a beta-transglycosidase. *Protein Eng Des Sel* **27**: 13–19. <https://doi.org/10.1093/protein/gzt057>.
- Thomas JG, Ayling A, Baneyx F. 1997. Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli. To fold or to refold. *Appl Biochem Biotechnol* **66**: 197–238.
- Trouilleux P, Potier P, Driguez P. 2019. Kilogram-scale Production of Synthetic Heparin Analogs: Some Chemical Considerations. In *Protecting Groups*, pp. 473–491, Wiley
<https://doi.org/10.1002/9783527697014.ch17>.
- Uchiyama T, Miyazaki K. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20**: 616–622. <https://doi.org/10.1016/j.copbio.2009.09.010>.
- Umekawa M, Huang W, Li B, Fujita K, Ashida H, Wang LX, Yamamoto K. 2008. Mutants of *Mucor hiemalis* endo- β -N-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities. *J Biol Chem* **283**: 4469–4479. <https://doi.org/10.1074/jbc.M707137200>.
- Umekawa M, Li C, Higashiyama T, Huang W, Ashida H, Yamamoto K, Wang LX. 2010. Efficient glycosynthase mutant derived from *Mucor hiemalis* endo- β -N-acetylglucosaminidase capable of transferring oligosaccharide from both sugar oxazoline and natural N-glycan. *J Biol Chem* **285**: 511–521. <https://doi.org/10.1074/jbc.M109.059832>.
- Urashima T, Hirabayashi J, Sato S, Kobata A. 2018. Human Milk Oligosaccharides as Essential Tools for Basic and Application Studies on Galectins. *Trends Glycosci Glycotecnol* **30**: SE51–SE65. <https://doi.org/10.4052/tigg.1734.1SJ>.
- Urbach C, Halila S, Guerreiro C, Driguez H, Mulard LA, Armand S. 2014. CGTase-catalysed cis-glycosylation of L-rhamnosides for the preparation of *Shigella flexneri* 2a and 3a haptens. *ChemBiochem* **15**: 293–300. <https://doi.org/10.1002/cbic.201300597>.

- Van der Borgh J, Soetaert W, Desmet T. 2012. Engineering the acceptor specificity of trehalose phosphorylase for the production of trehalose analogs. *Biotechnol Prog* **28**: 1257–1262. <https://doi.org/10.1002/btpr.1609>.
- van der Put RMF, Kim TH, Guerreiro C, Thouron F, Hoogerhout P, Sansonetti PJ, Westdijk J, Stork M, Phalipon A, Mulard LA. 2016. A Synthetic Carbohydrate Conjugate Vaccine Candidate against Shigellosis: Improved Bioconjugation and Impact of Alum on Immunogenicity. *Bioconjug Chem* **27**: 883–892. <https://doi.org/10.1021/acs.bioconjchem.5b00617>.
- van Gunsteren WF, Berendsen HJC. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys* **34**: 1311–1327. <https://doi.org/10.1080/00268977700102571>.
- van Leeuwen SS, Kralj S, van Geel-Schutten IH, Gerwig GJ, Dijkhuizen L, Kamerling JP. 2008a. Structural analysis of the α -d-glucan (EPS180) produced by the *Lactobacillus reuteri* strain 180 glucansucrase GTF180 enzyme. *Carbohydr Res* **343**: 1237–1250. <https://doi.org/10.1016/j.carres.2008.01.042>.
- van Leeuwen SS, Kralj S, van Geel-Schutten IH, Gerwig GJ, Dijkhuizen L, Kamerling JP. 2008b. Structural analysis of the α -d-glucan (EPS35-5) produced by the *Lactobacillus reuteri* strain 35-5 glucansucrase GTFA enzyme. *Carbohydr Res* **343**: 1251–1265. <https://doi.org/10.1016/j.carres.2008.01.044>.
- Varki A. 2017. Biological roles of glycans. *Glycobiology* **27**: 3–49. <https://doi.org/10.1093/glycob/cww086>.
- Vergès A, Cambon E, Barbe S, Moulis C, Remaud-Siméon M, André I. 2017. Novel product specificity toward erlose and panose exhibited by multisite engineered mutants of amylosucrase. *Protein Sci* **26**: 566–577. <https://doi.org/10.1002/pro.3106>.
- Vergès A, Cambon E, Barbe S, Salamone S, Le Guen Y, Moulis C, Mulard LA, Remaud-Siméon M, André I. 2015. Computer-aided engineering of a transglycosylase for the glucosylation of an unnatural disaccharide of relevance for bacterial antigen synthesis. *ACS Catal* **5**: 1186–1198. <https://doi.org/10.1021/cs501288r>.
- Vuillemin M, Claverie M, Brison Y, Séverac E, Bondy P, Morel S, Monsan P, Moulis C, Remaud-Siméon M. 2016. Characterization of the first α -(1→3) branching sucraes of the GH70 family. *J Biol Chem* **291**: 7687–702. <https://doi.org/10.1074/jbc.M115.688044>.
- Vuillemin M, Grimaud F, Claverie M, Rolland-Sabaté A, Garnier C, Lucas P, Monsan P, Dols-Lafargue M, Remaud-Siméon M, Moulis C. 2018. A dextran with unique rheological properties produced by the dextransucrase from *Oenococcus kitaharae* DSM 17330. *Carbohydr Polym* **179**: 10–18. <https://doi.org/10.1016/j.carbpol.2017.09.056>.
- Vuillemin M, Malbert Y, Laguerre S, Remaud-Siméon M, Moulis C. 2014. Optimizing the production of an α -(1→2) branching sucrose in *Escherichia coli* using statistical design. *Appl Microbiol Biotechnol* **98**: 5173–84. <https://doi.org/10.1007/s00253-014-5627-5>.
- Vujicic-Zagar A, Pijning T, Kralj S, López C a, Eeuwema W, Dijkhuizen L, Dijkstra BW. 2010. Crystal structure of a 117 kDa glucansucrase fragment provides insight into evolution and product specificity of GH70 enzymes. *Proc Natl Acad Sci U S A* **107**: 21406–11. <https://doi.org/10.1073/pnas.1007531107>.
- Wang G, Zhang W, Lu Z, Wang P, Zhang X, Li Y. 2009. Convenient synthesis of an N-glycan octasaccharide of the bisecting type. *J Org Chem* **74**: 2508–15. <https://doi.org/10.1021/jo900016j>.
- Wang W, Malcolm BA. 1999. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *Biotechniques* **26**: 680–2. <https://doi.org/10.2144/99264st03>.
- Wang Z, Chinoy ZS, Ambre SG, Peng W, McBride R, De Vries RP, Glushka J, Paulson JC, Boons GJ. 2013. A general strategy for the chemoenzymatic synthesis of asymmetrically branched N-glycans. *Science (80-)* **341**: 379–383. <https://doi.org/10.1126/science.1236231>.
- Warmerdam A, Paudel E, Jia W, Boom RM, Janssen AEM. 2013. Characterization of β -Galactosidase isoforms from *Bacillus circulans* and their contribution to GOS production. *Appl Biochem Biotechnol* **170**: 340–358. <https://doi.org/10.1007/s12010-013-0181-7>.

- Weijers CAGM, Franssen MCR, Visser GM. 2008. Glycosyltransferase-catalyzed synthesis of bioactive oligosaccharides. *Biotechnol Adv* **26**: 436–456. <https://doi.org/10.1016/j.biotechadv.2008.05.001>.
- Weiser J, Shenkin PS, Still WC. 1999. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem* **20**: 217–230. [https://doi.org/10.1002/\(SICI\)1096-987X\(19990130\)20:2<217::AID-JCC4>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1096-987X(19990130)20:2<217::AID-JCC4>3.0.CO;2-A).
- Wen L, Edmunds G, Gibbons C, Zhang J, Gadi MR, Zhu H, Fang J, Liu X, Kong Y, Wang PG. 2018. Toward automated enzymatic synthesis of oligosaccharides. *Chem Rev* **118**: 8151–8187. <https://doi.org/10.1021/acs.chemrev.8b00066>.
- Wessel H-P, Bundle DR. 1985. Strategies for the synthesis of branched oligosaccharides of the *Shigella flexneri* 5a, 5b, and variant X serogroups employing a multifunctional rhamnose precursor. *J Chem Soc Perkin Trans 1* 2251. <https://doi.org/10.1039/p19850002251>.
- West NP, Sansonetti P, Mounier J, Exley RM, Parsot C, Guadagnini S, Prévost M-C, Prochnicka-Chalufour A, Delepierre M, Tanguy M, et al. 2005. Optimization of virulence functions through glucosylation of *Shigella* LPS. *Science* **307**: 1313–7. <https://doi.org/10.1126/science.1108472>.
- Wickson EJ, Grossman RF. 2008. Formulation Development. In *Handbook of Vinyl Formulating* (ed. R.F. Grossman), *Wiley Series on Plastics Engineering and Technology*, pp. 1–12, John Wiley & Sons, Inc., Hoboken, NJ, USA <https://doi.org/10.1002/9780470253595.ch1>.
- Williams S. 2015. Transglycosylases.
- Woo HJ, Kang HK, Nguyen TTH, Kim GE, Kim YM, Park JS, Kim D, Cha J, Moon YH, Nam SH, et al. 2012. Synthesis and characterization of ampelopsin glucosides using dextransucrase from *Leuconostoc mesenteroides* B-1299CB4: Glucosylation enhancing physicochemical properties. *Enzyme Microb Technol* **51**: 311–318. <https://doi.org/10.1016/j.enzmictec.2012.07.014>.
- Woods RJ, Chappelle R. 2000. Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates. *Theochem* **527**: 149–156. [https://doi.org/10.1016/S0166-1280\(00\)00487-5](https://doi.org/10.1016/S0166-1280(00)00487-5).
- World Health Organization. 2005. Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* type 1. p. 64.
- Wren BW, Russell RR, Tabaqchali S. 1991. Antigenic cross-reactivity and functional inhibition by antibodies to *Clostridium difficile* toxin A, *Streptococcus mutans* glucan-binding protein, and a synthetic peptide. *Infect Immun* **59**: 3151–5.
- Xiao Z, Guo Y, Liu Y, Li L, Zhang Q, Wen L, Wang X, Kondengaden SM, Wu Z, Zhou J, et al. 2016. Chemoenzymatic Synthesis of a Library of Human Milk Oligosaccharides. *J Org Chem* **81**: 5851–5865. <https://doi.org/10.1021/acs.joc.6b00478>.
- Xu Y, Fan Y, Ye J, Wang F, Nie Q, Wang L, Wang PG, Cao H, Cheng J. 2018. Successfully engineering a bacterial sialyltransferase for regioselective α 2,6-sialylation. *ACS Catal* **8**: 7222–7227. <https://doi.org/10.1021/acscatal.8b01993>.
- Yamamoto T, Yamashita H, Mukai K, Watanabe H, Kubota M, Chaen H, Fukuda S. 2006. Construction and characterization of chimeric enzymes of kojibiose phosphorylase and trehalose phosphorylase from *Thermoanaerobacter brockii*. *Carbohydr Res* **341**: 2350–2359. <https://doi.org/10.1016/j.carres.2006.06.024>.
- Yan M, Wang B, Xu X, Chang P, Hang F, Wu Z, You C, Liu Z. 2018. Molecular and Functional Study of a Branching Sucrase-Like Glucansucrase Reveals an Evolutionary Intermediate between Two Subfamilies of the GH70 Enzymes. *Appl Environ Microbiol* **84**: 1–13. <https://doi.org/10.1128/aem.02810-17>.
- Yang C, Li P, Zhang X, Ma Q, Cui X, Li H, Liu H, Wang J, Xie J, Wu F, et al. 2016. Molecular characterization and analysis of high-level multidrug-resistance of *Shigella flexneri* serotype 4s strains from China. *Sci Rep* **6**: 1–11. <https://doi.org/10.1038/srep29124>.
- Ye X, Zhang C, Zhang YHP. 2012. Engineering a large protein by combined rational and random approaches: stabilizing the *Clostridium thermocellum* cellobiose phosphorylase. *Mol Biosyst* **8**: 1815–1823. <https://doi.org/10.1039/c2mb05492b>.

- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. DbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**: 445–451. <https://doi.org/10.1093/nar/gks479>.
- Zeuner B, Jers C, Mikkelsen JD, Meyer AS. 2014. Methods for improving enzymatic trans-glycosylation for synthesis of human milk oligosaccharide biomimetics. *J Agric Food Chem* **62**: 9615–9631. <https://doi.org/10.1021/jf502619p>.
- Zeuner B, Vuillemin M, Holck J, Muschiol J, Meyer AS. 2018. Loop engineering of an α -1,3/4-L-fucosidase for improved synthesis of human milk oligosaccharides. *Enzyme Microb Technol* **115**: 37–44. <https://doi.org/10.1016/j.enzmictec.2018.04.008>.
- Zhang F, Zhang Z, Linhardt RJ. 2010. Glycosaminoglycans. In *Handbook of Glycomics* (eds. R. Cummings and J. Pierce), pp. 59–80, London NW1 7BY, UK <https://doi.org/10.1016/B978-0-12-373600-0.00003-2>.
- Zhang X, Pagadala V, Jester HM, Lim AM, Pham TQ, Goulas AMP, Liu J, Linhardt RJ. 2017. Chemoenzymatic synthesis of heparan sulfate and heparin oligosaccharides and NMR analysis: paving the way to a diverse library for glycobiologists. *Chem Sci* **8**: 7932–7940. <https://doi.org/10.1039/c7sc03541a>.
- Zhu Y, Romain C, Williams CK. 2016. Sustainable polymers from renewable resources. *Nature* **540**: 354–362. <https://doi.org/10.1038/nature21001>.
- Zimmermann S, Lepenies B. 2015. Glycans as Vaccine Antigens and Adjuvants: Immunological Considerations. pp. 11–26 https://doi.org/10.1007/978-1-4939-2874-3_2.
- Zivkovic AM, German JB, Lebrilla CB, Mills DA. 2011. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *Proc Natl Acad Sci* **108**: 4653–4658. <https://doi.org/10.1073/pnas.1000083107>.