Computer-aided design and engineering of sucrose-utilizing transglucosylases for oligosaccharide synthesis

Alizee Verges

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

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

Présentée et soutenue par :
Alizée Vergès
le mercredi 8 avril 2015

Titre :

Computer-aided Design and Engineering of Sucrose-Utilizing Transglucosylases for Oliogsaccharide Synthesis

---

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

Unité de recherche :
LISBP, UMR CNRS 5504, UMR INRA 792, INSA de Toulouse

Directeur/trice(s) de Thèse :

Isabelle ANDRE, Directrice de Recherche, CNRS

Jury :
Birte SVENSSON, Professeur
Charles TELLIER, Professeur
Laurence MULARD, Directrice de Recherche, Institut Pasteur
Sylvain COTTAZ, Professeur
Magali REMAUD-SIMEON, Professeur

THÈSE

Présentée devant

L'Institut National Des Sciences Appliquées de Toulouse

Spécialité : Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries

Filière : Microbiologie et Biocatalyse Industrielle

Par **Alizée VERGÈS**

Ingénieur de l'Institut National des Sciences Appliquées de Toulouse

Computer-aided Design and Engineering of Sucrose-Utilizing Transglucosylases for Oligosaccharide Synthesis

Directrice de thèse :

Isabelle André, Directrice de Recherche CNRS

Co-encadrante :

Magali Remaud-Siméon, Professeur des Universités INSA

Remerciements

Je pense que dans la vie, l'une des choses les plus importantes est de savoir remercier et d'exprimer sa gratitude envers les personnes qui nous entourent, qui nous soutiennent ou que l'on croise simplement et qui nous marquent. Par les quelques mots qui vont suivre, je souhaiterais tout simplement dire MERCI à ces personnes.

Tout d'abord je tiens à remercier ma directrice de thèse Isabelle André et ma co-directrice de thèse, Magali Remaud-Siméon qui m'ont permis d'entrer dans le monde de la recherche depuis mon stage de fin d'études. Elles ont su me faire partager leur goût pour les défis et surtout le dépassement de soi. Elles m'ont également épaulée dans les moments difficiles auxquels on doit savoir faire face pendant tout projet de recherche, en étant patiente et à l'écoute. Merci à vous !

Ensuite, cette thèse n'aurait pas eu lieu sans une collaboration fructueuse et dans la bonne humeur avec Laurence Mulard de l'Institut Pasteur qui depuis des années déjà montre combien la chimie et la biocatalyse sont complémentaires ! Merci Laurence, mais aussi Stéphane et Yan qui ont su répondre à nos attentes (de substrat) en urgence !!! Je remercie également Emmanuelle Cambon qui m'a formée dès mon stage de Master. Elle a su être patiente et pédagogue afin de me faire découvrir ce qu'est l'ingénierie d'enzyme et surtout me donner l'envie de continuer en thèse. Merci Manue pour tous ces bons moments dans et en dehors du labo.

Tout au long des ces 4 années, j'ai également rencontré des personnes qui sont devenues des amis ou qui ont une importance pour moi et à qui je souhaite témoigner mon affection. Nelly, Sandra, Laurence, Florence, Grimaud la Granule, Delphine, Bastien, Marlène, Maéva, Betty, Alvaro, Etienne, Marina, Marion, Yoann, Yannick, Seydou, Christopher et Marc. Ils ont participé tous à leur façon à l'aboutissement de ce travail. Merciiiii

Une petite parenthèse pour Anne-Marie, David et Soraya avec qui j'ai travaillé directement et qui ont une place toute particulière dans mon cœur….

Merci à Samuel Tranier qui a su me faire découvrir la cristallographie des protéines et surtout me transmettre sa passion pour la recherche ! J'ai passé de très beaux moments à l'IPBS ! Merci aussi à Gianlucas qui a pris le relais avec la passion à l'italienne !

Enfin une pensée particulière pour Claude Maranges et Pierre Monsan sans qui mon parcours aurait été très différent et avec qui on peut discuter sans limites et sans peur d'être jugée. Merci pour votre soutien.

Je tiens également à remercier les membres du jury et les rapporteurs, Birte Svensson, Charles Tellier, Laurence Mulard, Sylvain Cottaz et Magali qui m'ont fait l'honneur de lire, d'analyser et de discuter de mon travail de façon agréable et détendue.

Merci à toutes les personnes de l'équipe CIMES qui chaque jour apportent joie et bonne humeur !

Merci à ma mère et à ma sœur d'être là encore et encore malgré toutes les épreuves, et à mes amis du labo et extérieurs pour leur soutien, leur sourire et leur folie !

SURNAME: VERGÈS **Forename:** Alizée

Title: Computer-aided Design and Engineering of Sucrose-utilizing Transglucosylases for Oligosaccharide Synthesis.

Speciality: Ecological, Veterinary, Agronomic Sciences and Bioengineering. **Field:** Enzymatic and Microbial Engineering.

Year: 2015 **Place:** INSA Toulouse

Summary:

Chemical synthesis of complex oligosaccharides still remains critical. Enzymes have emerged as powerful tools to circumvent chemical boundaries of glycochemistry. However, natural enzymes do not necessarily display the required properties and need to be optimized by molecular engineering. Combined use of chemistry and tailored biocatalysts may thus be attractive for exploring novel synthetic routes, especially for glyco-based vaccines development. The objective of this thesis was thus to apply semi-rational engineering strategies to *Neisseria polysaccharea* amylosucrase (*Np*AS), a sucrose-utilizing α-transglucosylase, in order to conceive novel substrate specificities and extend the potential of this enzyme to catalyze novel reactions, going beyond what nature has to offer.

In a first study, a computer aided-approach was followed to reshape the active site of the enzyme (subsites +1, +2 and +3) for the recognition and α-1,4 glucosylation of a non-natural disaccharide acceptor molecule (allyl 2-deoxy-2-*N*-trichloroacetyl-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranose). The trisaccharide product is a building block for the chemo-enzymatic synthesis of oligosaccharides mimicking the repetitive units of the *Shigella flexneri* lipopolysaccharides, and ultimately, for the production of a vaccine against Shigellosis disease. Using computational tools dedicated to the automated protein design, combined with sequence analysis, a library of about $2.7x10^4$ sequences was designed and experimentally constructed and screened. Altogether, 55 mutants were identified to be active on sucrose (the donor substrate), and one, called mutant F3, was subsequently found able to catalyze the α-1,4 glucosylation of the target disaccharide. Impressively, this mutant contained seven mutations in the first shell of the active site leading to a drastic reshaping of the catalytic pocket without significantly perturbing the original specificity for sucrose donor substrate.

In a second study, three variants were identified from the screening of the semi-rational library on sole sucrose as displaying totally novel product specificities. They were further characterized, as well as their products, at both biochemical and structural level. These mutants, called 37G4, 39A8 and 47A10, contained between 7 and 11 mutations into their active site. They were found able to use sucrose and maltose (a reaction product from sucrose) as both donor and acceptor substrates to produce in varying amounts erlose (α-D-Glucopyranosyl-(1→4)-α-D-Glucopyranosyl-(1→2)-β-D-Fructose) and panose (α-D-Glucopyranosyl-(1→6)-α-D-Glucopyranosyl-(1→4)-α-D-glucose) trisaccharides, which are not produced at all by parental wild-type enzyme. Relatively high yields were obtained for the production of these molecules, which are known to have acariogenic and sweetening properties and could be of interest for food applications.

In a last part, another mutant 30H3 was isolated due to its high activity on sucrose (6.5-fold improvement compared to wild-type activity) from primary screening of the library. When characterized, the mutant revealed a singular product profile compared to that of wild-type *Np*AS. It appeared highly efficient for the synthesis of soluble maltooligosaccharides of controlled size chains, from DP 3 to 21, and with a low polydispersity. No formation of insoluble polymer was found. The X-ray structure of the mutant was determined and revealed the opening of the catalytic pocket due to the presence of 9 mutations in the first sphere. Molecular dynamics simulations suggested a role of mutations onto flexibility of domain B' that might interfere with oligosaccharide binding and explain product specificity of the mutant.

KEYWORDS: Amylosucrase, Transglycosylase, Glucosylation, Computer-Aided Library Design, Enzyme design and engineering, Chemo-enzymatic synthesis, *Shigella flexneri*, Antigenic oligosaccharides, Erlose, Panose, Maltooligosaccharide.

Doctorate school: SEVAB (Ecological, Veterinary, Agronomic Sciences and Bioengineering).
Laboratory: Laboratory of Biosystems and Chemical Engineering (UMR CNRS 5504, UMR INRA 792) of INSA Toulouse.

NOM : VERGÈS **Prénom :** Alizée

Titre : Design Computationnel et Ingénierie de Transglucosylases pour la Synthèse d'oligosaccharides.

Specialité : Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénierie. **Domaine :** Microbiologie et Biocatalyse Industrielle.

Année : 2015 **Lieu :** INSA Toulouse

Résumé :

La synthèse d'oligosides complexes reste difficilement réalisable par voie chimique. Le recours aux catalyseurs enzymatiques permettrait de pallier aux contraintes de la chimie mais les enzymes naturelles ne présentent pas toujours les propriétés adéquates et nécessitent d'être optimisées par ingénierie moléculaire. Le couplage de la chimie et de biocatalyseurs conçus « sur mesure », peut offrir une alternative prometteuse pour explorer de nouvelles voies de synthèse des sucres, notamment pour la mise au point de glycovaccins. L'objectif de cette thèse a ainsi visé à mettre en œuvre des stratégies d'ingénierie semi-rationnelles de l'amylosaccharase de *Neisseria polysaccharea* (AS*Np*), une α-transglucosylase utilisant le saccharose comme substrat, afin de concevoir de nouvelles spécificités de substrats et d'étendre le potentiel de cette enzyme à catalyser de nouvelles réactions, permettant ainsi d'aller bien au-delà de ce que la Nature peut offrir.

Dans une première étude, une approche assistée par ordinateur a été suivie afin de remodeler le site actif de l'enzyme (sous-sites +1, +2 et +3) pour la reconnaissance et la glucosylation en α-1,4 d'un accepteur disaccharidique non-naturel (l'allyl 2-deoxy-2-*N*-trichloroacetyl-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranose). Le produit attendu, un trisaccharide, est un précurseur dans la synthèse chimio-enzymatique des oligosaccharides mimant les unités répétitives des lipopolysaccharides de *Shigella flexneri*, dont l'utilisation ultime est le développement de vaccins contre la Shigellose. Une approche computationnelle faisant appel à des outils dédiés au design automatisé de protéines et à une analyse des séquences a conduit au design d'une librairie d'environ $2.7x10^4$ séquences, qui a ensuite été construite expérimentalement puis criblée. Au final, 55 variants actifs sur saccharose (le substrat donneur) ont été identifiés, et un mutant, appelé F3, a révélé sa capacité à glucosyler en α-1,4 le disaccharide cible. De manière étonnante, ce mutant possède 7 mutations au sein de son site actif, nécessaires au déploiement de sa nouvelle spécificité tout en maintenant son aptitude à utiliser le saccharose comme donneur d'unité glucosyle.

Dans une deuxième étude, trois variants ont été identifiés lors du criblage de la librairie semi-rationnelle sur saccharose comme présentant de nouvelles spécificités de produits. Ces mutants ont été caractérisés plus en détails, ainsi que leurs produits, sur un plan biochimique et structural. Ces mutants, appelés 37G4, 39A8 et 47A10, contiennent entre 7 et 11 mutations dans leur site actif. Il a été montré qu'ils étaient capables de reconnaitre le saccharose et le maltose (un produit de la réaction avec le saccharose) comme donneur et accepteur pour synthétiser en quantités variables de l'erlose (α-D-Glucopyranosyl-(1→4)-α-D-Glucopyranosyl-(1→2)-β-D-Fructose) et du panose (α-D-Glucopyranosyl-(1→6)-α-D-Glucopyranosyl-(1→4)-α-D-glucose), des molécules non produites par l'enzyme sauvage. Des taux de production relativement élevés ont été obtenus pour ces molécules, dont les propriétés acariogènes et le pouvoir sucrant pourraient présenter un intérêt applicatif pour l'industrie alimentaire.

Dans une dernière partie, un autre mutant, appelé 30H3, a été isolé lors du criblage primaire de la librairie de par son activité élevée sur saccharose (une amélioration d'un facteur 6.5 comparé à l'enzyme sauvage). Après caractérisation, le mutant s'est avéré synthétiser un profil unique de produits en comparaison de l'enzyme sauvage AS*Np*. Il s'est ainsi montré très efficace pour la synthèse de maltooligosaccharides solubles, de taille de chaînes contrôlée allant d'un DP 3 à 21, et de faible polydispersité. Aucun polymère insoluble n'a été identifié. La structure 3D du mutant résolue par cristallographie des rayons X a révélé un agrandissement de la poche catalytique en raison de la présence de 9 mutations introduites dans la première sphère. Des simulations de dynamique moléculaire ont suggéré un rôle de ces mutations sur la flexibilité du domaine B' qui pourrait vraisemblablement interférer avec l'arrimage des oligosaccharides à la surface de l'enzyme et être à l'origine de la spécificité de produits du mutant

MOTS CLES : Amylosaccharase, Transglycosylase, Glucosylation, Design computationnel de librairies, Design et Ingénierie d'Enzymes, Synthèse Chimio-enzymatique, *Shigella flexneri*, Oligosaccharides antigéniques, Erlose, Panose, Maltooligosaccharide.

Ecole Doctorale : SEVAB (Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénierie).
Laboratoire: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (UMR CNRS 5504, UMR INRA 792) – INSA de Toulouse.

Publications

Computer-aided engineering of a transglycosylase for the glucosylation of an unnatural disaccharide of relevance for bacterial antigen synthesis. **Vergès A.,** Cambon E., Barbe S., Salamone S., Le Guen Y., Moulis C., Mulard L.A., Remaud-Siméon M. and Andre I. ; ACS Catal., 2015, *5* (2), pp 1186–1198

Novel product specificity toward erlose and panose exhibited by multi-site engineered mutants of amylosucrase. **Vergès A.** *et al*., In preparation

Isolation and characterization of an efficient mutant of *Neisseria polysaccharea* amylosucrase for the synthesis of controlled size maltooligosaccharides. **Vergès A.** *et al.,* In preparation

Oral communications

Computer-aided design and engineering of *Neisseria polysaccharea* amylosucrase toward the synthesis of novel oligosaccharides. **Vergès A.,** Cambon E., Barbe S., Tranier S., Remaud-Siméon M. & André I., *International Carbohydrate Symposium*, January 12-17, 2014, Bangalore, India

Molecular engineering perspectives of GH family 13 amylosucrases toward novel products: from natural promiscuity to tailored catalysis. Daudé D., **Vergès A.,** Cambon E., Barbe S., Monsan P., Remaud-Siméon M. & André I., *International Carbohydrate Symposium*, January 12-17, 2014, Bangalore, India

A Neutral-Drift based approach for building polymorphic variant library of an α - transglucosidase. Daudé D., Cambon E., **Vergès A**., Emond S., André I. & Remaud-Siméon M., *Biotrans*, July 21-25, 2013, Manchester, England

Diversification of *Neisseria polysaccharea* amylosucrase activity through a semi-rational engineering approach. **Vergès A.,** Cambon E., Barbe S., Tranier S., Remaud-Siméon M. & André I., *Congrès annuel de la Société Française de Biochimie et Biologie Moléculaire*, GT Enzymes, September 4-6, 2013, Paris, France

One enzyme – Many activities: Use of multiple enzyme engineering approaches to create novel functions of a carbohydrate-active enzyme. Cambon E., **Vergès A**., Daudé D., Barbe S., Moulis C., Remaud-Siméon M. & André I., *Congrès annuel de la Société Française de Biochimie et Biologie Moléculaire,* GT Enzymes, September 4-6, 2013, Paris, *France*

Diversification of *Neisseria Polysaccharea* amylosucrase activity by semi-rational engineering, **(award for best student oral communication). Vergès A.,** Cambon E., Barbe S., Tranier S., Remaud-Siméon M. & André I., *Glucidoc meeting,* April 8-11, 2013, Landéda, France

Structure-based engineering of glycoenzymes for the development of chemo-enzymatic pathways to access bacterial carbohydrates. Cambon E., Guérin F., Champion E., **Vergès A**., Barbe S., Moulis C., Morel S., Salamone S., Mulard L.A., Remaud-Siméon M. & André I*., International Carbohydrate Symposium*, July 22-27, 2012, Madrid, Spain

"A Neutral-drift based approach for the generation of polymorphic variants library of *Neisseria polysaccharea* amylosucrase". Daudé D., Cambon E., **Vergès A.,** Emond S., André I. & Remaud-Siméon M., *International Carbohydrate Symposium*, July 22 – 27, 2012, Madrid, Spain

Neutral-Drift based approach for building polymorphic variants library of *Neisseria polysaccharea* amylosucrase.
Daudé D., Cambon E., **Vergès A.**, Tarrat N., Morel S., Guieysse D., André I., Remaud-Siméon M., *Congrès annuel de la Société Française de Biochimie et Biologie Moléculaire*, October, 2011, Ax-les-Thermes, France

Posters

Semi-rational engineering of *Neisseria polysaccharea* amylosucrase towards the synthesis of novel oligosaccharides for glycobased vaccines **(bursary from SFBBM).** **Vergès A.**, Cambon E., Barbe S., Salamone S., Le Guen Y., Moulis C., Mulard L.A., Remaud-Siméon M. and André I., FEBS-EMBO, August 30-september 4, 2014, Paris, France

Diversification of *Neisseria Polysaccharea* amylosucrase activity by semi-rational engineering. **Vergès A.,** Cambon E., Barbe S., Tranier S., Remaud-Siméon M. & André I., *Cosm'ing meeting,* June 26-28, 2013, Saint-Malo, France

Diversification of *Neisseria Polysaccharea* amylosucrase activity by semi-rational engineering. **Vergès A.,** Cambon E., Barbe S., Tranier S., Remaud-Siméon M. & André I., *Journée de l'Ecole Doctorale SEVAB,* November 6, 2012, Toulouse, France

Stratégie originale de la construction d'une banque de *Neisseria polysaccharea* basée sur la théorie neutraliste de l'évolution. Daudé D., Cambon E., **Vergès A.**, Tarrat N., Morel S., Guieysse D., André I. & Remaud-Siméon M., *Congrès annuel de la Société Française de Biochimie et Biologie Moléculaire*, October 12-14, 2011, Ax-Les-Thermes, France

Table of contents

Abbreviation list

(A), (B), (C): L- Rhamnopyranoside
(A'): α-L- Methyl Rhamnopyranoside
*Ac*AS: *Arthrobacter chlorophenolicus* amylosucrase
*Am*AS: *Alteromonas macleodii* amylosucrase
ANR: National Agency of the Research
ASR: Alternansucrase

BTB: BromoThymol Blue

CAZy: Carbohydrate-Active enZymes
CCP4: Collaborative Computational Project No.4

(D): *N*-acetyl-D-glucosamine
(D'): Allyl 2-*N*-trichloroacetyl-2-deoxy-β-D-glucopyranoside
(D'A'): Allyl 2-deoxy-2-*N*-trichloroacetyl-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside
*Dg*AS: *Deinococcus geothermalis* amylosucrase
DNS: 3,5-dinitrosalicylic acid
DP: Degree of Polymerization
*Dr*AS: *Deinococcus radiodurans* amylosucrase
DSF: Differential Scanning Fluorimetry
DSR: Dextransucase
DTT: DiThioThreitol

(E): D-glucopyranose
EDTA: (EthyleneDinitrilo)Tetraacetic Acid
epPCR: Error prone PCR
ESRF: European Synchrotron Radiation Facility

FACS: Fluorescence Activated Cell Sorting
Fru: Fructose

GH: Glycoside-Hydrolase
Glc: Glucose
GS: GlucanSucrase
GST: Glutathion -S-Transferase
GT: GlycosylTransferase

HEPES: 4-(2-HydroxyEthyl)-1-PiperazineEthaneSulfonic acid
HPAEC-PAD: High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection
HPLC: High Performance Liquid Chromatography
HPSEC: High-Performance Size Exclusion Chromatography

IPTG: IsoPropyl β-D-1-ThioGalactopyranoside

Km: Michaelis dissociation constant

LB: Luria-Bertani medium
LC/MS: Liquid Chromatography/Mass Spectrometry
LISBP: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés
LOS: Lipo-OligoSide
LPS: LipoPolySaccharide

MD: Molecular Dynamics
*Mf*AS: *Methylobacillus flagellatus* amylosucrase

*Nm*AS: *Neisseria meningitidis* amylosucrase
NMR: Nuclear Magnetic Resonance
*Np*AS : *Neisseria polysaccharea* amylosucrase

O-Ag: O-Antigen
OB: Oligosaccharide Binding site
O-Sp: O-Specific antigen

PBS: Phosphate Buffer Saline
PCR: Polymerase Chain Reaction
PDB: Protein Data Bank
PEG: PolyEthyleneGlycol

SB: Sucrose Binding site
SDS-PAGE: Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
S. flexneri: *Shigella flexneri*
Suc: Sucrose

Tm: Melting Temperature
Treha:Trehalulose
TRIS: 2-amino-2-hydroxyméthyl-1,3-propanediol
Tur: Turanose

U: one unit (U) corresponds to 1µmol of fructose liberated per minute
UV: Ultraviolet light

Introduction

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 prevented extensive exploration of chemo-enzymatic routes to complex carbohydrates. To circumvent this limitation, protein engineering has proven to be very efficient to tailor enzymes with novel substrate specificities. However, the outcome of protein engineering strategies strongly depends on our understanding of enzyme catalysis and our comprehension of the inter-relationships between protein structure, activity and dynamics.

Computational biology has also led to major advances in the prediction of protein structures from their amino acid sequences but also, in recent years, in the development of methods dedicated to the remodelling of known protein scaffolds in order to modify the function/property of existing enzymes, or to design from scratch (*de novo* design) new enzymes to fulfill a particular function. These methods enable to guide evolution of proteins on structural regions of relevance to achieve the desired function, and thus reduce the size of sequence libraries to build and screen experimentally. Enzyme catalysts with unprecedented properties, even catalyzing reactions not found in nature, have been designed using such computational approaches. However, these methods have been scarcely applied to the engineering of carbohydrate-active enzymes.

In this work, we have explored the potential of computer-aided design strategies to engineer carbohydrate-active enzymes endowed with a novel substrate specificity and able to catalyze new transglucosylation reactions which will be integrated into programmed chemo-enzymatic routes of synthesis to access carbohydrate components of microbial cell-surface oligosaccharides. This study fits into a collaborative ANR project called "GlucoDesign" (2009-2012) between the Catalysis and Enzyme Molecular Engineering group of the Laboratory of Biosystems and Chemical Engineering (LISBP) from INSA Toulouse and the GlycoChemistry group led by Laurence Mulard at the Institut Pasteur in Paris. The objective *in fine* is to develop chemo-enzymatic routes that include enzymatic steps at different stages of the process using engineered enzymes able to act on lightly chemically protected intermediates and enabling to overcome the lack of stereoselectivity of chemical 1,2-*cis* glucosylation with the purpose of accessing easily to carbohydrates that enter in the composition of vaccines.

Focusing on the synthesis of *Shigella flexneri* oligosaccharides, which are responsible for endemic shigellosis (a diarrheic disease), semi-rational engineering strategies have already been successfully applied to conceive novel transglycosylases able to catalyze the regio-specific glucosylation of two partially protected monosaccharides, namely the α-L-rhamnopyranoside and the allyl 2-acetamido-2-deoxy-α-D-glucopyranoside. Building up on this prior work, and aspiring at exploring further the potential offered by state-of-the-art computational design approaches, we have undertaken a larger challenge within this thesis as the objective was to deeply re-engineer the active site of a sucrose-utilizing enzyme, amylosucrase from *Neisseria polysaccharea*, to recognize a partially protected disaccharide, the allyl 2-deoxy-2-*N*-trichloroacetyl-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside, and glucosylate it with the wanted regiospecificity without activity loss for sucrose donor substrate.

Using a computer-aided approach, a library of ~2.7×10^4 mutants targeting 23 positions of the active site first shell was constructed and screened on sucrose and targeted disaccharide as acceptor. As a result, a variety of original activities were isolated in comparison to parental wild-type enzyme. Some of these activities were those desired, such as the novel acceptor

specificity for the glucosylation of the targeted disaccharide. Extended screening of this library also allowed the discovery of other activities, such as the ability to produce molecules not naturally produced by the wild – type enzyme or to synthesize a controlled distribution of soluble short chain oligosaccharides from sole sucrose. These newly discovered activities were further characterized and molecular determinants involved in these changes of specificity were investigated by a combination of biochemical and biophysical techniques.

The first chapter of this manuscript will give an overview on the characteristics of amylosucrases, the family of enzymes investigated here, followed by a survey of enzyme engineering strategies: random, (semi)-rational and computer-aided design, which can be used for *in vitro* evolution. Screening methods developed are also presented followed by a presentation of main achievements in amylosucrase engineering.

Results obtained within the frame of this thesis are reported as a series of three articles, with one already published and two that will be shortly submitted for publication. In chapter three, the computer-aided approach that was followed to construct the semi-rational library is described and the main results obtained from the screening of the library are presented, in particular the isolation of a mutant able to glucosylate a partially protected β-linked disaccharide allyl (2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside, a potential intermediate in the synthesis of *Shigella flexneri* cell-surface oligosaccharides. The fourth chapter reports the characterization of three mutants isolated from this library and that produce trisaccharides, erlose and panose, not synthesized by wild-type enzyme, from sole sucrose. To finish, the fifth chapter focuses on another mutant isolated from the same library that turned out to be very efficient for the production of soluble short maltooligosaccharide chains, of very low polydispersity, thus revealing the potentiality and diversity of this structurally guided library of medium size that targeted two acceptor binding subsites.

Chapter I:

Bibliographic section

Over the past years, sucrose-utilizing transglucosylases have emerged as valued tools in chemistry to generate glycodiversification. Not only these enzymes use as substrate an abundant agroresource, sucrose, but they also share a remarkable versatility regarding the acceptor substrate, allowing the structurally-controlled synthesis of diverse glucosylated products. With the help of enzyme engineering techniques, significant progresses have been made in recent years to optimize the properties of natural enzymes and to enlarge the repertoire of accessible reactions. In particular, advanced computational methods used in combination with evolution techniques have contributed to the tailoring of novel biocatalysts answering specific needs and having the targeted activity, often with no reported equivalent in Nature. The first bibliographic chapter gives an overview on this class of enzymes, focusing particularly on our model enzyme, amylosucrase from *Neisseria polysaccharea*. This enzyme is very singular as it has been classified in the Glycoside-Hydrolase family 13, which contains mostly starch degrading enzymes, although it is a polymerase that presents some unique structural characteristics which will be detailed in the following. It is one of the sucrose-utilizing enzymes that has been the most tickled by enzyme engineering, either by directed evolution (random approach) or by structure-guided mutagenesis, what has revealed its remarkable malleability and adaptability to novel substrates and reaction conditions. Main achievements in amylosucrase engineering will be presented in this chapter, after a survey of current enzyme engineering approaches and screening methods which can be used for *in vitro* evolution and that will be illustrated with some relevant examples.

I. Amylosucrase: an unique glucansucrase from GH13 family

I.1. Overview of glucansucrases

Glucansucrases (GS) are α-transglucosylases that use sucrose as glucosyl donor to synthesize α-glucans with a concomitant release of fructose **(Figure 1)**. The α-glucans synthesized by these enzymes differ by the type of glucosidic linkages as well as by the degree and spatial arrangements of their branches, their structure being dependent on enzyme specificity. Small amounts of glucose and sucrose isomers are also formed. They result from sucrose hydrolysis or glucosyl transfer onto fructose molecule released during the reaction course, respectively. Sucrose isomers that may be synthesized by glucansucrases include turanose (α-D-glucosylpyranosyl-1,3-D-fructofuranose), trehalulose (α-D-glucosylpyranosyl-1,1-D-fructofuranose), leucrose (α-D-glucosylpyranosyl-1,5-D-fructofuranose) or isomaltulose (α-D-glucosylpyranosyl-1,6-D-fructofuranose).

Glucansucrases have been classified by the Enzyme Commission in the EC 2.4.1.4, EC 2.4.1.5 or EC 2.4.1.140 subclassifications. In the sequence-based CAZy classification, they have been placed in families 13 and 70 of Glycoside-Hydrolases (GH)[1].

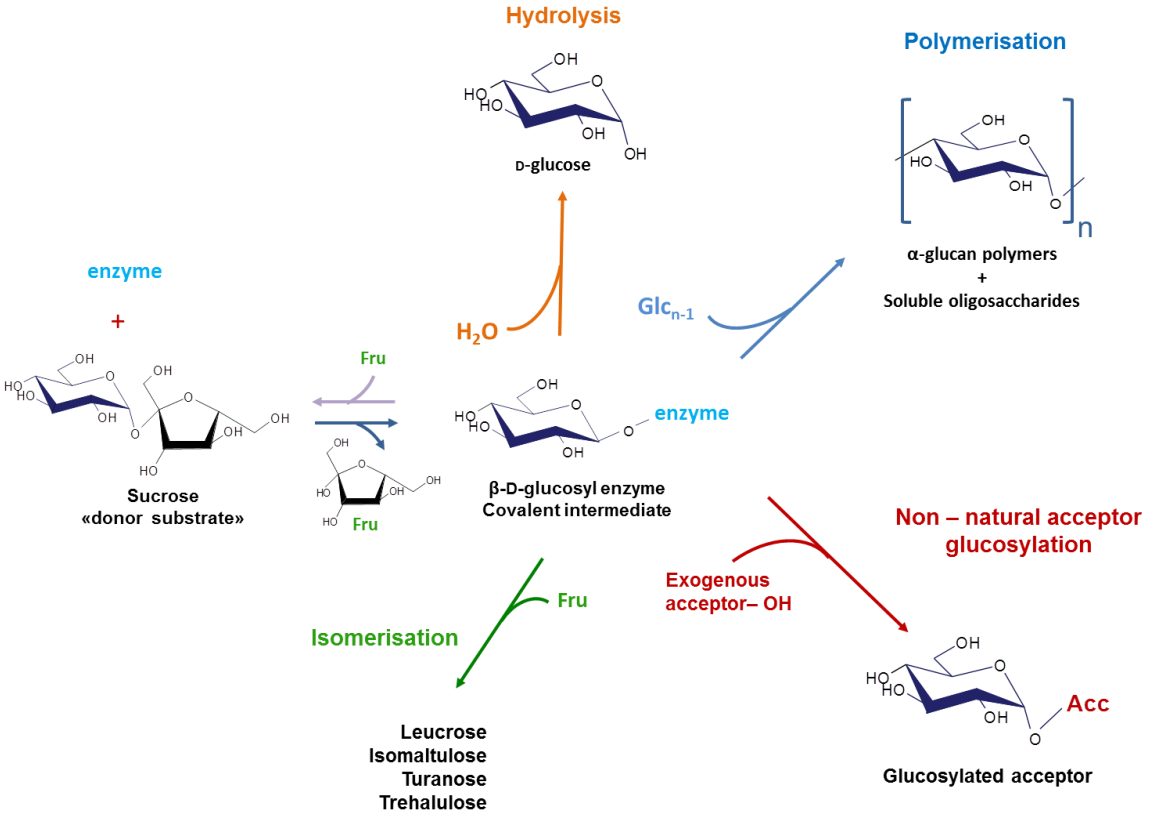


Figure 1: Reactions catalyzed by glucansucrases. *In orange is illustrated the hydrolysis reaction; in blue the polymerization reaction (synthesis of α-glucans); in green the isomerization reaction; in red the non-natural acceptor reaction; Acc: acceptor molecule; Glc: glucose; Fru: fructose; n: degree of polymerization.*

GS from GH70 family are all produced by Gram^{+} lactic acid bacteria (*Leuconostoc sp.*, *Streptococcus sp.*, *Lactococcus sp.*, *Lactobacillus sp.*, and *Weissella sp.*)[2,3,4,5,6,7,8,9] and are classified into 4 groups depending on the structure of the polymer they synthesize **(Figure 2).**

We can distinguish:

- alternansucrases (ASR) (GH70, E.C.2.4.1.140) that produce alternan, a glucan containing alternating α-(1,6) and α-(1,3) linkages in the main chain and some ramifications also grafted with either α-(1,3) or α-(1,6) osidic bonds.

- dextransucrases (DSR) (GH70, E.C.2.4.1.5) that produce dextran, a glucan showing more than 50% α-(1,6) linkages in its main chain and containing branches linked by α-(1,2), α-(1,3) or α-(1,4) glucosidic bonds.

- mutansucrases (GH70, E.C.2.4.1.5) that produce mutan, a glucan with more than 50% of contiguous α-(1,3) linkages.

- reuteransucrases (GH70, E.C.2.4.1.5), that produce reuteran, a glucan containing more than 50% of α-(1,4) linkages.

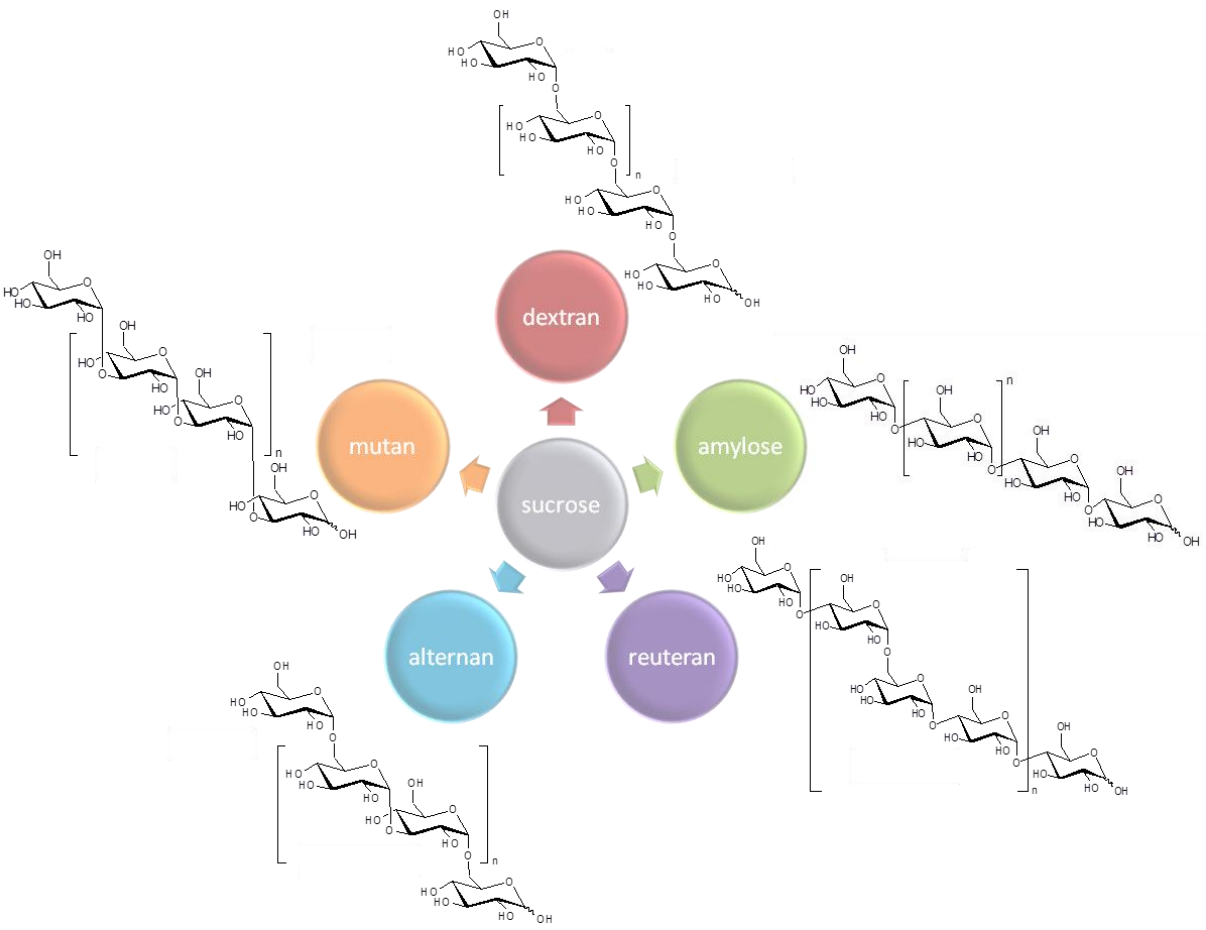


Figure 2: Schematic representation of basic α-glucan polymers synthesized by glucansucrases from GH70 and GH13 families[10].

Amylosucrases constitute another class of glucansucrases. They mainly produce an amylose-like polymer from sucrose **(Figure 2)** and are classified in the GH13 family. This family is also named the α-amylase family, as it mainly contains enzymes involved in starch biotransformation and belongs to the same GH-H clan together with GH70 and GH77 families. It is the largest family of the CAZy classification with more than 20, 900 members identified in October 2014. Noteworthy, different enzymatic activities have been reported in this family including hydrolase, transferase or isomerase activities differing by their substrate or product specificities[1]. This family also contains non-catalytic members such as amino acid transporters[11]. On the basis of their sequences and phylogenetic analyses, GH13 family was subdivided into 40 subfamilies[12] **(Table 1).** Amylosucrases are the sole polymerases; they belong to GH13_4 subfamily.

Table 1: Subfamilies of GH13 family[12].

Subfamily	EC number	Enzyme activities	Subfamily	EC number	Enzyme activities
1	3.2.1.1	α-Amylase	21	3.2.1.20	α-Glucosidase
2	3.2.1.1 2.4.1..19 3.2.1.133	α-Amylase Cyclodextrin glucanotransferase Maltogenic α-Amylase	22	2.4.1.183	α-1,3-Glucan synthase
3	ND	Unknown activity	23	ND	Unknown activity
4	2.4.1.4 3.2.2.-	Amylosucrase Sucrose hydrolase	24	3.2.1.1	α-Amylase
5	3.2.1.1	α-Amylase	25	3.2.1.33	Amylo-α-1,6-glucosidase
6	3.2.1	α-Amylase	26	5.4.99.15	(1,4)-α-Glucan 1-α-glucosylmutase
7	3.2.1.1	α-Amylase	27	3.2.1.1	α-Amylase
8	2.4.1.8	1,4-α-Glucan branching enzyme	28	3.2.1.1	α-Amylase
9	2.4.1.8	1,4-α-Glucan branching enzyme	29	3.2.1.93	α-Phosphotrehalase
10	3.2.1.141	4-α-(1,4-α-Glucano)trehalose-trehalohydrolase	30	3.2.1.20	α-Glucosidase
11	3.2.1.68	Isoamylase	31	3.2.1.70 3.2.1.10	Glucan 1,6-α-glucosidase Oligo-1,6-glucosidase
12	3.2.1.41	Pullulanase	32	3.2.1.1	α-Amylase
13	3.2.1.41	Pullulanase	33	5.4.99.16	Trehalose synthase
14	3.2.1.41	Pullulanase	34	NA	Amino acid transporter
15	3.2.1.1	α-Amylase	35	NA	Amino acid transporter
16	5.4.99.16	Maltose α-glucosyltransferase	36	ND	Unknown activity
17	3.2.1.20	α-Glucosidase	37	ND	α-Amylase
18	2.4.1.7	Sucrose phosphorylase	38	ND	Unknown activity
19	3.2.1.1 3.2.1.98 3.2.1.-	α-Amylase Maltohexaose-forming α-amylase Maltopentaose – forming α-amylase	39	ND	Unknown activity
20	3.2.1.54 3.2.1.133 3.2.1.135	Cyclomaltodextrinase Maltogenic α-amylase Neopullulanse	40	3.2.1.20	α-Glucosidase

ND: not determined
NA: not applicable

Glucansucrases from GH13 and GH70 families also catalyze glucosylation of non-natural hydroxylated acceptors (**Figure 1**). A large panel of acceptors has been reported including:

- mono- and disaccharides (D-glucose, maltose, isomaltose, nigerose, gentiobiose, cellobiose,…..) that lead to short gluco-oligosaccharides[13, 56, 15, 16]

- oligosaccharides (panose, acarbose..)[13, 17]

- functionalized sugars (anhydrofructose, dianhydre fructose, alkylglucosides…)[13,18, 19, 20, 21, 22]

- alditols (D-glucitol, D-mannitol, maltitol) [13, 20]

- alcohols (methanol, ethanol, propanol….)[23]

- aromatic compounds (catechins, polyphenols)[24, 25, 26, 27]

- amino acid derivatives[23, 27, 25.]

Of note, acceptor recognition and glucosylation are dependent on the enzyme specificity. Acceptor reaction has been broadly exploited to generate transglucosylation products such as prebiotic molecules[28, 29], antigenic oligosaccharides[17, 21] or antioxydants[24, 25, 27, 30, 18], which are of interest for a wide range of applications.

Glucansucrases from GH families 13 and 70 share the same α-retaining mechanism that will be described in more details on **page 40**. Their three-dimensional structures are related but reveal major differences as shown in **Figure 3**. Amylosucrase catalytic domains adopt the common fold with the A, B and C domains usually found in family GH13 and two additional domains (N and B' domain) that are specific to amylosucrases. The **A catalytic domain** is characterized by a $(\beta/\alpha)_8$ TIM barrel consisting in eight β-strands alternating with eight α-helix (**Figure 3A**). GH70 enzymes are larger in size than amylosucrases (≥120kDa) and adopt a structure in which the $(\beta/\alpha)_8$ catalytic domain (domain A) is formed by non-contiguous fragments of sequences that are homologous to those of GH13 family but differently organized, thus confirming the predicted circular permutation in which the secondary elements of the barrel are found in the following order: α3, β4/α4, β5/α5, β6/α6, β7/α7, β8/α8, β1/α1, β2/α2, β3 (**Figure 3B**).

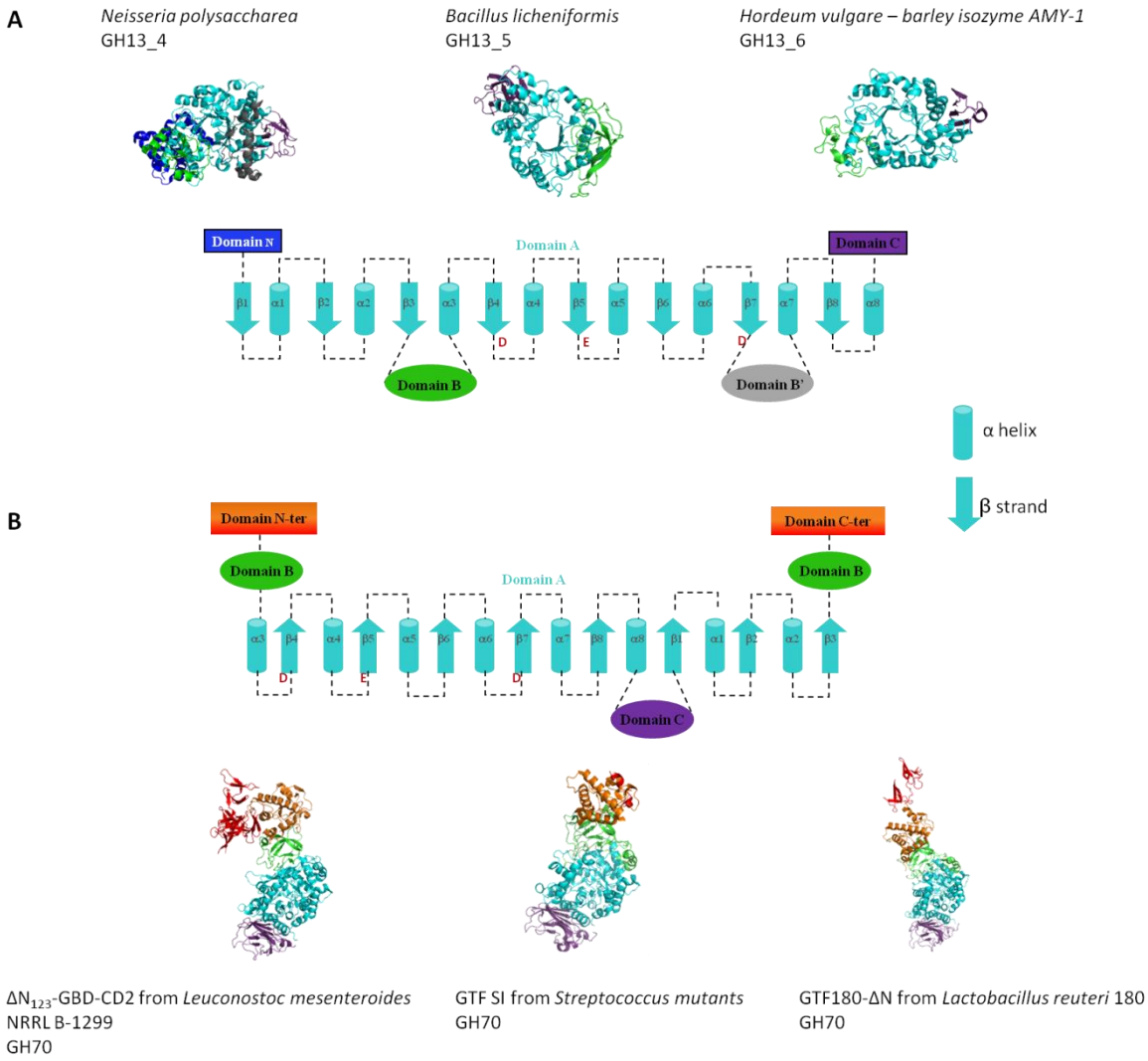


Figure 3: Topology diagrams of members of GH13 (A) and GH70 families (B)[31]. ***A:*** *View of selected enzymes from family GH13. From left to right:* Neisseria polysaccharea *GH13_4 (PDB code: 1G5A)[32];* Bacillus licheniformis *GH13_5 (PDB code: 1BLI)[33] and* Hordeum vulgare *– barley isozyme AMY-1 GH13_6 (PDB code: 1P6W)[34].* ***B:*** *View of selected enzymes from family GH70. From left to right: ΔN$_{123}$-GBD-CD2 from* Leuconostoc mesenteroides *NRRL B-1299 (PDB code: 3TTQ)[35] ; GTF SI from* Streptococcus mutans *(PDB code: 3AIE)[36] and GTF180-ΔN from* Lactobacillus reuteri *180 (PDB code: 3KLK)[37]. Individual domains are colored as follows: catalytic domain A in cyan; domain B in green; domain B' in grey; domain C in purple; Domain N in blue; Domain C-ter and Domain N-ter (which are composed of domains IV and V for GH70) in red and orange. The approximate sites of the three catalytic residues (D, E, D) are indicated in red.*

As amylosucrases are the enzymes at the center of our work, the following sections were essentially devoted to these enzymes.

I.2. Mechanistic and structural properties of amylosucrases

a- Producing organisms

The first intracellular amylosucrase produced by *Neisseria perflava* was discovered by Hehre et Hamilton in 1946[38, 39, 40]. Then, other intracellular enzymes were produced by various species of *Neisseria* (*N. canis, N. cinerea, N. dentrificans, N. sicca* and *N. subflava*[41]). These bacteria belong to the microbiote of the buccal cavity and may be potentially involved in formation of dental carries[42, 43]. In 1983, another strain of *Neisseria*, isolated from the throat of healthy children[44], *Neisseria polysaccharea*, was identified as a producer of extracellular amylosucrase. The gene encoding this enzyme was cloned and sequenced[45] and the recombinant enzyme was biochemically and structurally characterized[46,70]. Since 2005, several recombinant amylosucrases from *Deinococcus radiodurans*[47], *Deinococcus geothermalis*[48], *Arthrobacter chlorophenolicus*[49], *Alteromonas macleodii*[50] *Methylobacillus flagellatus*[51] and *Synechococcus sp*.[52] have been described. Two three-dimensional structures of amylosucrases from *Deinococcus radiodurans*[53] and *Deinococcus geothermalis*[54] have also been recently reported in 2012 and 2013, respectively. **Table 2** summarizes the main biochemical properties of these 7 characterized amylosucrases. Their percentage of identity is presented in **Table 3.**

Table 2: Biochemical and structural data of characterized recombinant amylosucrases at this date (October 2014).

Organism	Living conditions	Enzyme size	Reaction conditions	Specific activity (U/mg^{-1})			Conformation in solution	3-D structures (PDB code)
				30°C	45°C	50°C		
N. polysaccharea Potocki de Montalk G. *et al.*, FEBS Lett., 2000	mesophile	1.911 nucleotides 70.000Da	$T°_{opt}$=37°C pH_{opt}=6.8-7.3°C $t_{1/2}$=21h (30°C) $t_{1/2}$=3min (50°C) Tm=49.6°C	9.6	n.d.	n.d.	monomeric	Active form: apo (1G5A)[32] D-glucose (1JG9)[55] Suc (1MW1-1MW2-1MW3)[56] Inactive form: Suc (1JGI)[55] G7 (1MW0-1MVY)[56] α-D-glucosyl unit (1S46)[57] Suc+G7 (1ZS2)[58]
D. geothermalis Emond S. *et al.*, FEMS Microbiol. Lett., 2008 Guérin F. *et al.*, JBC, 2012	moderate thermophile	1.950 nucleotides 73.300Da	$T°_{opt}$=50°C pH_{opt}=7.0 $t_{1/2}$=22h (50°C) $t_{1/2}$=65h (30°C) Tm=58.9°C	19	n.d.	44	dimeric	Active form: apo (3UCQ)[54] turanose (3UER)[54]
D. radiodurans Pizzut-Serin S. *et al.*, FEBS Lett., 2005	polyextremophile	1.935 nucleotides 71.155Da	$T°_{opt}$=n.d. pH_{opt}=7.0 $t_{1/2}$=15h (30°C) Tm=n.d.	4	n.d.	n.d.	dimeric	Active form: apo (4AYS)[53]
Synechococcus sp. **PCC 7002** Perez-Cenci M. *et al.*, Plant science, 2012	mesophile	1.986 nucleotides 74.499Da	$T°_{opt}$=30°C pH_{opt}=6.5-7.0 $t_{1/2}$=n.d Tm=n.d	n.d.	n.d.	n.d.	n.d.	n.d.
A.Chlorophenolicus A6 Dong-Ho S. *et al.*, J. Microbiol. Biotechnol., 2012	mesophile (can survive under psychrophilic conditions)	1.932 nucleotides 71.705Da	$T°_{opt}$=45°C pH_{opt}=8.0 $t_{1/2}$=33.3min (40°C) $t_{1/2}$=14.3min (50°C) Tm=n.d.	n.d.	n.d.	n.d.	monomeric	n.d.
A.Macleodii Ha S. *et al.*, Biosci.Biotechnol.Biochem., 2009	mesophile	1.950 nucleotides 73.7863 Da	$T°_{opt}$=45°C pH_{opt}=8.0°C $t_{1/2}$=30min at 45°C Tm= n.d.	~150	432	~400	n.d.	n.d.
M. flagellatus **KT ATCC51484** Jeong J.-W. *et al.*, Appl.Biochem. Biotechnol., 2014	mesophile	1.953 nucleotides 74.556Da	$T°_{opt}$= 45°C pH_{opt}=8.0-8.5 $t_{1/2}$=19.3min (50°C) $t_{1/2}$=4.3min (55°C) Tm=50.6°C	n.d.	n.d.	n.d.	dimeric	n.d.

Table 3: Percentage of amino acid identity between characterized amylosucrases. *Multiple sequence alignment performed using ClustalW (http://www.genome.jp/tools/clustalw).*

	Neisseria polysaccharea	*Deinococcus geothermalis*	*Deinococcus radiodurans*	*Arthrobacter chlorophenolicus*	*Methylobacillus flagellatus*	*Synechococcus sp. PCC 7002*	*Alteromonas macleodii*
Neisseria polysaccharea		38.2	37.4	53.8	34.7	36.1	43.9
Deinococcus geothermalis			73.4	39.5	37.8	36.5	37.8
Deinococcus radiodurans				38.4	36.3	34.1	39.8
Arthrobacter chlorophenolicus					36.5	36.1	42.5
Methylobacillus flagellatus						57.4	37.4
Synechococcus sp. PCC 7002							37.4
Alteromonas macleodii							

% of identity	< 20	20-30	30-40	40-50	>50

The physiological role of amylosucrases has not been studied in detail. Buttcher *et al*. proposed that the polymer produced by *Np*AS might protect the cell[59]. It was also proposed that the enzyme could be involved in energy storage or could be used to elongate glycogen-like polymers[32]. However, these hypotheses have never been demonstrated. More recently, Perez-Cenci *et al*. proposed that amylosucrase from *Synechococcus sp*. PCC 7002 cyanobacterium could play a role in sucrose catabolism[52]. In this species, the amylosucrase encoding gene was co-expressed with 3 other genes forming the Suc cluster. Two of them, sppA and spsA encode a sucrose-phosphate phosphatase and a sucrose-phosphate synthase, respectively, and are responsible for sucrose synthesis whereas amsA and frkA encoding an amylosucrase and a fructokinase would be involved in sucrose catabolism to yield maltooligosaccharides and fructose-6P. These results indicate that in cyanobacteria, amylosucrase participate in energy storage from sucrose.

b- Biochemical properties

Profile of reaction products

From sucrose, amylosucrases synthesize glucose, fructose, maltose, maltotriose, sucrose isomers (turanose and trehalulose) and soluble maltooligosaccharides of degree of polymerization (DP) comprised between 3 and 25. Insoluble maltoligosaccharides of DP 26 to 90 are also formed and precipitate during the reaction course to yield insoluble and highly crystalline amylose[46]. The distribution of the reaction products is influenced by sucrose concentrations **(Figure 4)**. The production of insoluble glucan is maximal at 300mM sucrose (72%). In standard conditions, at 100mM sucrose, amylosucrase converts 57% of sucrose into insoluble glucan, 4% in glucose, 14% in turanose, 3% in trehalulose, and 20% in soluble maltooligosides[47, 61].

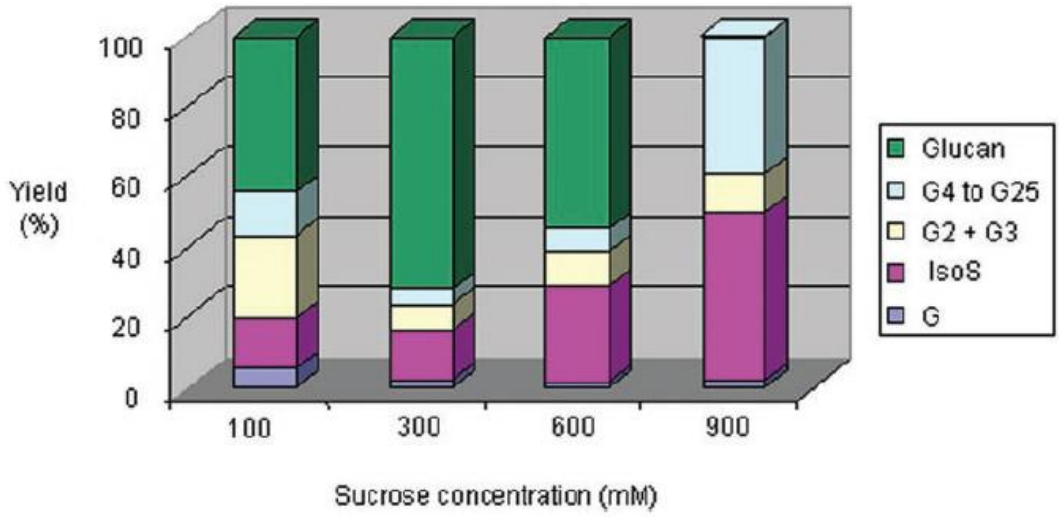


Figure 4: Distribution of the products synthesized by wild-type *N. polysaccharea* amylosucrase using various sucrose concentrations. *G: glucose; IsoS: sucrose isomers; Gn: maltooligosaccharides of DP n; glucan: insoluble fraction* (taken from Albenne C. *et al.*, 2004).

Of note, *Np*AS also catalyzes disproportionation of maltooligosaccharides by cleaving the α-1,4 glucosidic linkage at the non-reducing end of one maltooligosaccharide molecule and transferring the glucosyl moiety onto another molecule. In this reaction, MOS play the role of either donor or acceptor molecule. For example, from maltohexaose (G_6), the enzyme synthesizes maltopentaose (G_5) and maltoheptaose (G_7) in equimolar ratios. Then, maltotetraose (G_4) and maltooctaose (G_8) are also produced equimolarly. After 4h reaction, the equilibrium is reached. Disproportionation reaction starts to be efficient with MOS substrate composed of at least 5 glucosyl units [61].

Kinetic parameters

Two different Michaelis-Menten equations were used to model initial velocity values of sucrose consumption versus initial sucrose concentrations. The first one correctly fitted the values obtained between 0 and 20mM sucrose and the second one, fitted the values obtained for concentrations higher than 20mM. Two sets of kinetic parameters were determined **(Table 4).** Hydrolysis is favoured at low sucrose concentrations, whereas transglucosylation is preferred at high sucrose concentrations. This singular kinetic behaviour was attributed to the presence of a second sucrose binding site[62] differing from the sucrose binding site found in the catalytic pocket. At high sucrose concentration, sucrose binding was proposed to induce conformational changes impacting k_{cat} and favouring transglucosylation. Above 300mM sucrose concentration, enzyme activity is diminished due to sucrose inhibition. *Np*AS was also shown to be inhibited by fructose[63].

Table 4: Apparent kinetic parameters for sucrose consumption, glucose released and initial polymerization rates.

	[Sucrose] < 20mM			[Sucrose] > 20mM		
	K_m	V_m	k_{cat}	K_m	V_m	k_{cat}
Sucrose consumption	1.9mM	470µmol.min^{-1}.g^{-1}	33 min^{-1}	50.2mM	1100µmol.min^{-1}.g^{-1}	77 min^{-1}
Sucrose hydrolysis	1.7mM	288µmol of glucose released.min^{-1}.g^{-1}	20 min^{-1}	38.7mM	472µmol of glucose released.min^{-1}.g^{-1}	33 min^{-1}
Tranglucosylation	1.9mM	147µmol of glucose incorporated into α-glucan.min^{-1}.g^{-1}	10 min^{-1}	387mM	1620µmol of glucose incorporated into α-glucan.min^{-1}.g^{-1}	113 min^{-1}

c- Structure

Consensus sequences of GH13_4 subfamily.

Amylosucrases are proteins composed of about 600 amino acids. For example, the *Np*AS gene is composed of 1911 bp, and corresponds to a protein of 636 amino acids. Like all the enzymes from GH13 family, amylosucrase sequences share **8 consensus sequences** shown in **Figure 5 and Figure 6**. Numerotation is based upon *Np*AS sequence: **region I** (from residue 182 to 187 on β3 strand), **region II** (282-293 on β4 strand), **region III** (328-331 on β5 strand), **region IV** (388-395 on loop 7), **region V** (250-258 domain B), **region VI** (125-134 on β3 strand), **region VII** (480-489 on β8 strand) and **region VIII**[64, 65–68] (144 - 149 in the loop 2 of the $(\beta/\alpha)_8$ barrel).

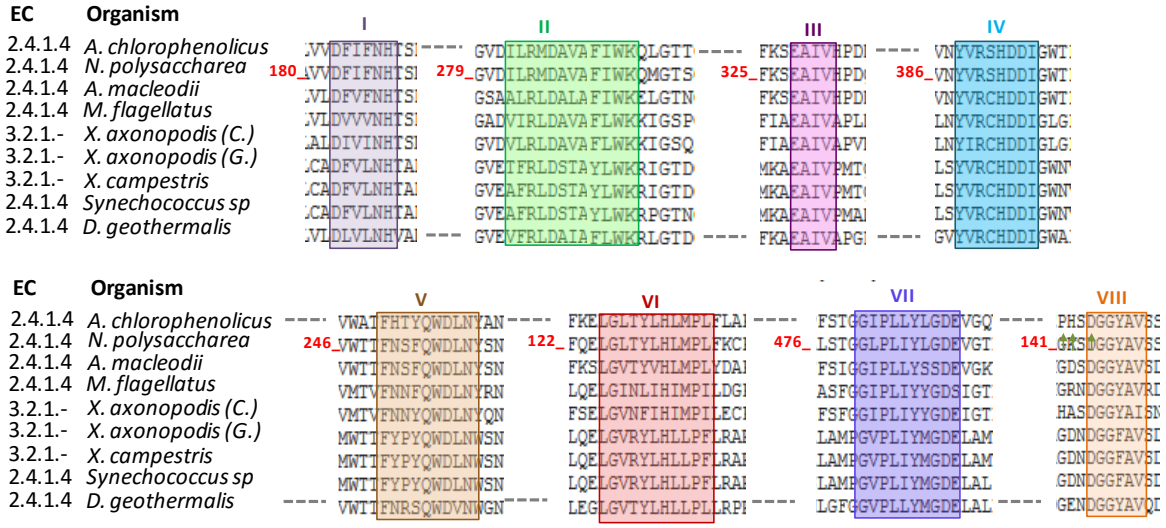


Figure 5: Conserved regions in the GH13 family. *Highlighted are the 8 conserved regions*[62]. *Enzyme sources:* A. chlorophenolicus: Arthrobacter chlorophenolicus A6; N. polysaccharea: Neisseria polysaccharea ATCC 43768; A. macleodii: Alteromonas macleodii KCTC 2957; M. flagellatus: Methylobacillus flagellatus KT; X. axonopodis (C.): Xanthomonas axonopodis pv. citri str. 306; X. axonopodis (G.): Xanthomonas axonopodis pv. glycines 8RA; X. campestris: Xanthomonas campestris pv. campestris str. ATCC 33913; Synechococcus sp: Synechococcus sp. PCC 7002 ATCC 27264; D. geothermalis: Deinococcus geothermalis DSM 11300. *Numbering is given with respect to* Np*AS sequence (indicated in red).*

In these regions, 7 residues are highly conserved **(Figure 5)** and they play a role in the mechanism of action which will be further detailed on **page 40**:

- 3 carboxylic acids corresponding respectively to the catalytic nucleophile **Asp** (region II), the acid-base catalyst **Glu** (region III) and the aspartate **Asp** (region IV)) involved in the transition state stabilization. Only this triade is **strictly** conserved. In *Np*AS, the catalytic triade is formed by nucleophile Asp286, acid-base Glu328 and stabilizer Asp393.

- 2 histidines that stabilize the covalent intermediate formed from substrate cleavage (**His** (region I), and **His** (region IV)). In *Np*AS, they correspond to His187 and His392.

- **Arg** (region II), and **Asp** (region I) that are important for structural integrity. In *Np*AS, these residues are Arg284 and Asp182.

Three additional residues (**Gly126, Pro134 and Gly480** using *Np*AS numbering) were proposed to play a role in α-amylase calcium binding[69].

Three additional consensus sequences were identified as specific of the sub-family GH13_4 which gathers amylosucrases and also sucrose-hydrolases (**Figure 6**)[63]. For *Np*AS, these regions are defined as follows:

- region IX encompassed by residues 225 to 237 and corresponding to the domain B,
- region XI defined by residues 446 to 456 corresponding to the domain B',
- region X corresponding to residues 509 to 514 from loop 8 of the $(\beta/\alpha)_8$ barrel.

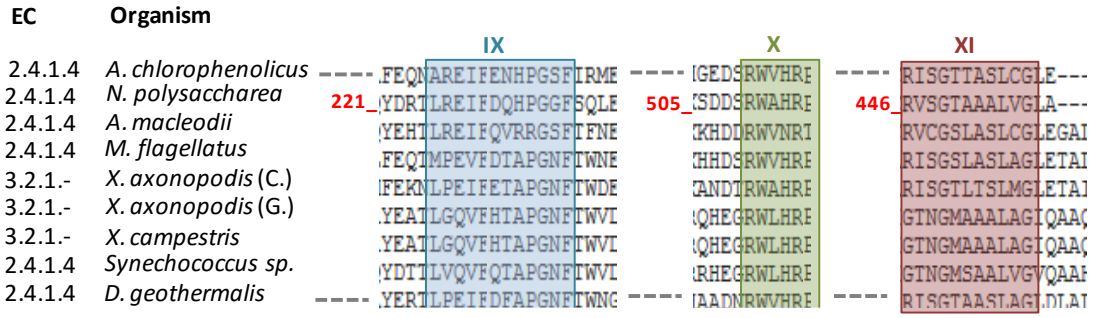


Figure 6: Additional consensus sequences found in the GH13_4 subfamily[59]. *Enzyme sources:* A. chlorophenolicus: Arthrobacter chlorophenolicus A6; N. polysaccharea: Neisseria polysaccharea ATCC 43768; A. macleodii: Alteromonas macleodii KCTC 2957; M. flagellatus: Methylobacillus flagellatus KT; X. axonopodis (C.): Xanthomonas axonopodis pv. citri str. 306 ; X. axonopodis (G.) : Xanthomonas axonopodis pv. glycines 8RA; X. campestris: Xanthomonas campestris pv. campestris str. ATCC 33913; Synechococcus sp: Synechococcus sp. PCC 7002 ATCC 27264; D. geothermalis: Deinococcus geothermalis DSM 11300. *Numbering is given with respect to* Np*AS sequence (indicated in red).*

3D structure of amylosucrases

Amylosucrase from *N. polysaccharea* is the first amylosucrase that had its crystallographic structure determined. Various 3D-structures of *Np*AS are available in the PDB (Protein Data Bank, http://www.rcsb.org)either) in apo form or in complex with various ligands (including sucrose, glucose, maltoheptaose and turanose complexes)[56, 55, 54]. The structure of a covalently bound glucosyl-enzyme intermediate has also been determined[57]. More recently, the

crystallographic structures of amylosucrase from *Deinococcus geothermalis* (*Dg*AS) and *Deinococcus radiodurans* (*Dr*AS) were released[53, 54].

In the following part, we will focus on the 3D-structure of *Np*AS, the enzyme studied in this work.

As mentioned earlier in this chapter, the three-dimensional structure of *Np*AS consists of 5 domains: the **A, B and C domains** which are common to all the GH family 13 enzymes and the **N and B' domains** that are specific to amylosucrases[32, 70] **(Figure 7).**

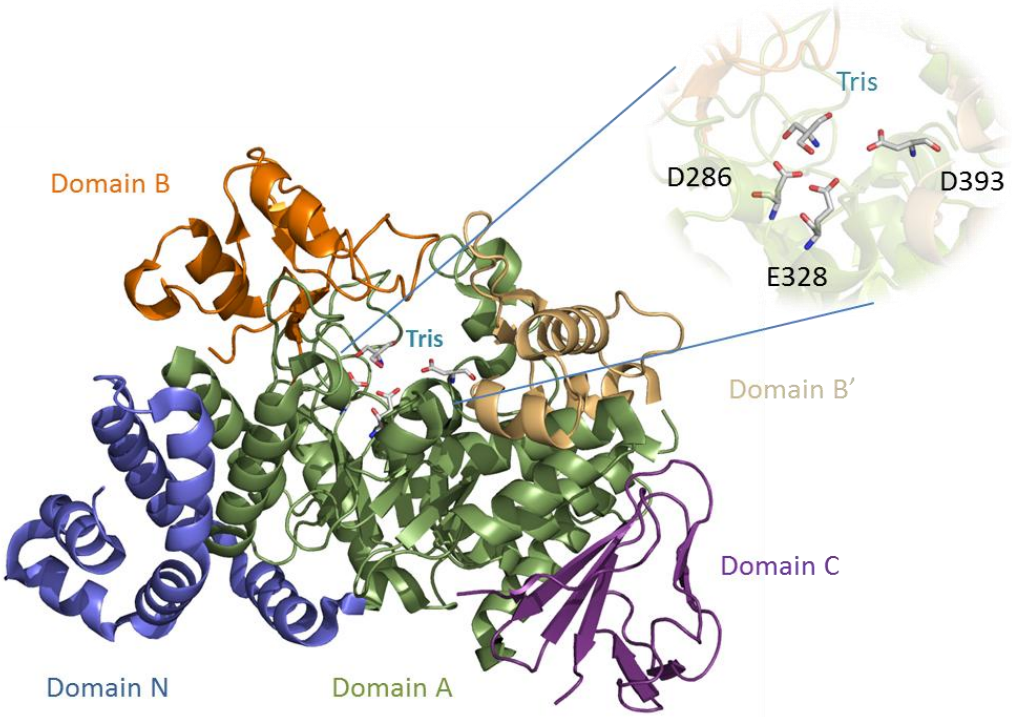


Figure 7: Three-dimensional structure of amylosucrase from *Neisseria polysaccharea* (PDB code: 1G5A) with a bound Tris molecule in the active site. *The different domains are colored: domain A (green), domain B (orange), domain C (purple), domain B' (brown), domain N (blue). The catalytic residues, Asp286 (nucleophile), Glu328 (acide/base catalyst), Asp393 (covalent intermediate stabilizer) and Tris molecule are represented in sticks.*

Domain A (residues 90-184, 261-395 and 461-553) is a $(\beta/\alpha)_8$ barrel domain. The catalytic site is formed by catalytic residues **Asp286** (nucleophile), **Glu328** (acid/base catalyst) and **Asp393** (transition state stabilizer). These residues are located at the bottom of a narrow pocket delimited by a salt bridge formed between residues Asp144 and Arg509. The active site is composed of sugar binding subsites that span from -1 at the non-reducing end to +6 at the reducing end **(Figure 8),** according to the nomenclature defined by Davies and Henrissat[71]. Sucrose (α-D-glucopyranosyl-1,2-β-D-fructofuranoside) occupies subsites -1 and +1 and substrate cleavage occurs between the two subsites.

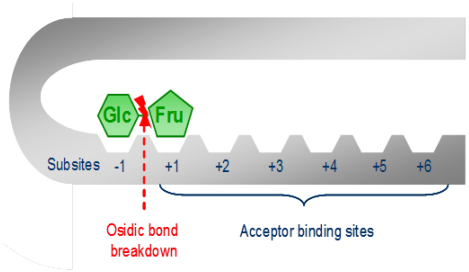


Figure 8: Schematic representation of sucrose binding into the active site of *Neisseria polysaccharea* amylosucrase. *In green is shown a sucrose molecule composed of a glucosyl moiety (Glc) linked to a fructosyl moiety (Fru).*

Domain B (residues 185-260) corresponds to the loop 3 inserted between β-strand 3 and α-helix 3. This domain was shown to be often involved in the substrate or product specificity as well as in the stability of GH13 enzymes[11].

Domain C constitutes the greek key domain. It is composed of an 8 stranded β-sandwich (residues 554-628) whose role has not been clearly elucidated yet. In Barley α-amylases, it is assumed to be involved in malto-oligosaccharide binding[34, 72, 73, 74].

N-terminal domain (residues 1-89) is exclusively formed of α-helices and is specific to amylosucrases. Its function has not been elucidated.

Domain B' (residues 396-460), is an extended loop located between the β_7 -strand and the α_7-helix of the catalytic barrel. This domain covers up the active site and it is responsible for its pocket-like topology. It is composed of two helices followed by a β-strand and is terminated by a last short α-helix[70, 32].

Sucrose binding sites

Determination of the crystallographic structure of an inactive form of *Np*AS (containing an E328Q mutation) in complex with sucrose (PDB code: 1JGI[55]) enabled the identification of two sucrose binding sites named respectively, SB1[55] and SB2[40,61] **(Figure 9A).** SB1 corresponds to the catalytic site. As shown in **Figure 9B**, the glucosyl ring is tightly bound at subsite -1 through stacking interactions with Tyr147 and Phe250 and a network of hydrogen bonds involving Asp286, His187, Arg284, Asp393, His392, Asp144 and Arg509[52, 68, 62]. The fructosyl ring of sucrose is bound in subsite +1 and it is maintained by a weaker network of hydrogen bonding interactions involving Glu328Gln (the inactivated acid/base residue in X-ray structure that is replaced by Glu328 in functional *Np*AS), Asp393, Asp394 and Arg446. Noteworthy, Asp144, Arg509, Asp394 and Arg446 are not conserved in the GH13 family and might be responsible of *Np*AS substrate specificity[76, 75, 55]. SB2 is located at the surface of B' domain and it has been hypothesized that it could serve as a guide to feed SB1 with sucrose by a movement of the B'domain[61, 62].

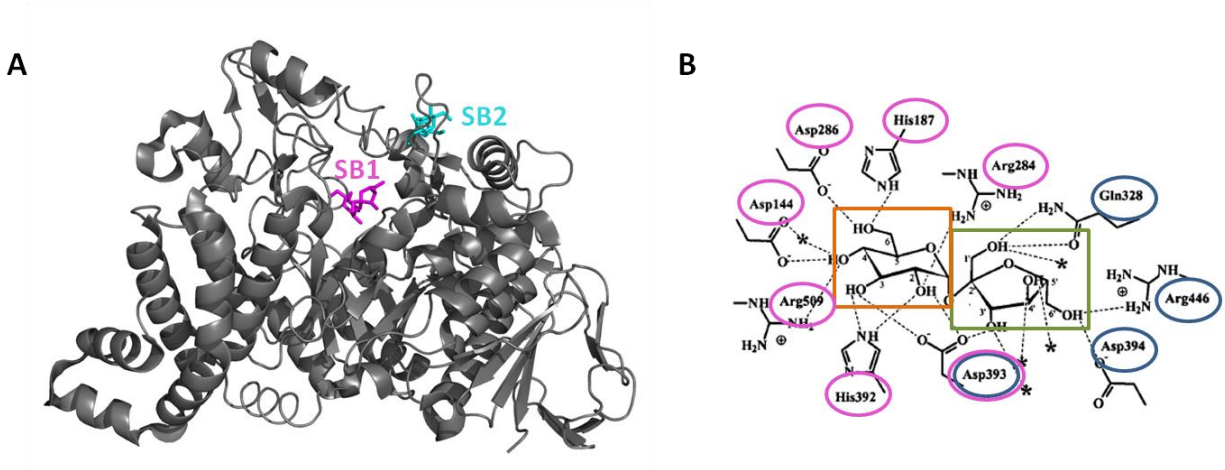


Figure 9: (A) Representation of sucrose binding sites SB1 (magenta) and SB2 (cyan) identified in the crystallographic structure of inactivated *Np*AS in complex with sucrose (PDB code: 1JGI and 1MW1). (B): Schematic view of *Np*AS/Sucrose interactions at SB1 site *(Glu328 has been inactivated by introduction of a Gln mutation). Hydrogen bonds are shown as dotted lines. Water molecules are marked as asterisks. Orange square highlights the glucopyranosyl ring; green square indicates the fructosyl moiety; amino acids encircled in blue and magenta are involved into hydrogen bonding interactions with the fructosyl and glucosyl rings, respectively* (adapted from Mirza O. *et al.*, 2001)[55].

Additional oligosaccharide binding sites:

The structure of the inactive E328Q-*Np*AS/maltoheptaose complex (PDB codes: 1MVY[56] and 1MWO[56]) further revealed the identification of three oligosaccharide binding sites named OB1, OB2 and OB3[56] **(Figure 10)**. OB1 spans over the -1 and +1 subsites and also includes five additional subsites (+2 to +6 defined by the 5 last glucosyl units of maltoheptaose). Maltoheptaose binding in OB1 is ensured by a network of interactions described in **Figure 11**. The OB2 and OB3 sites are located at the surface of B' and C domains, respectively. OB2 was suggested to play a role onto amylose and/or glycogen anchoring during elongation[56]. No function was attributed to OB3 so far.

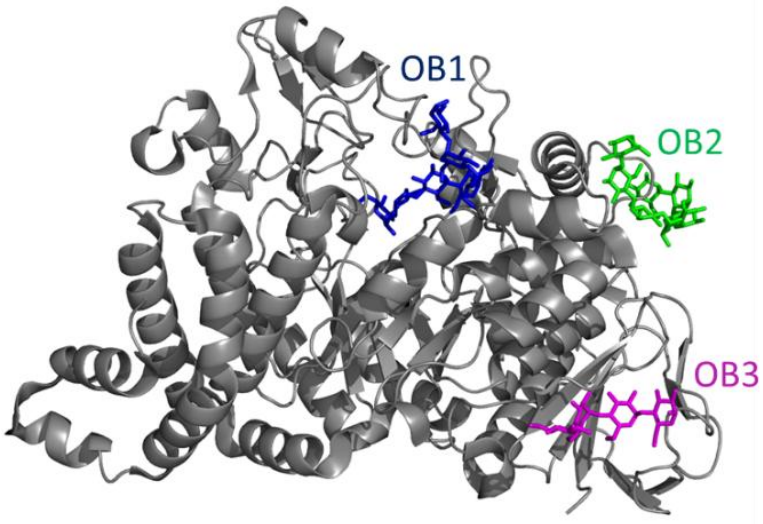


Figure 10: View of oligosaccharide binding sites of *Np*AS.
OB1 (blue) corresponding to the active site, OB2 (green) at the surface of the B'domain and OB3 (magenta) at the surface of the C domain of inactive Np*AS-E328Q/maltoheptaose complex (code PDB: 1MWO).*

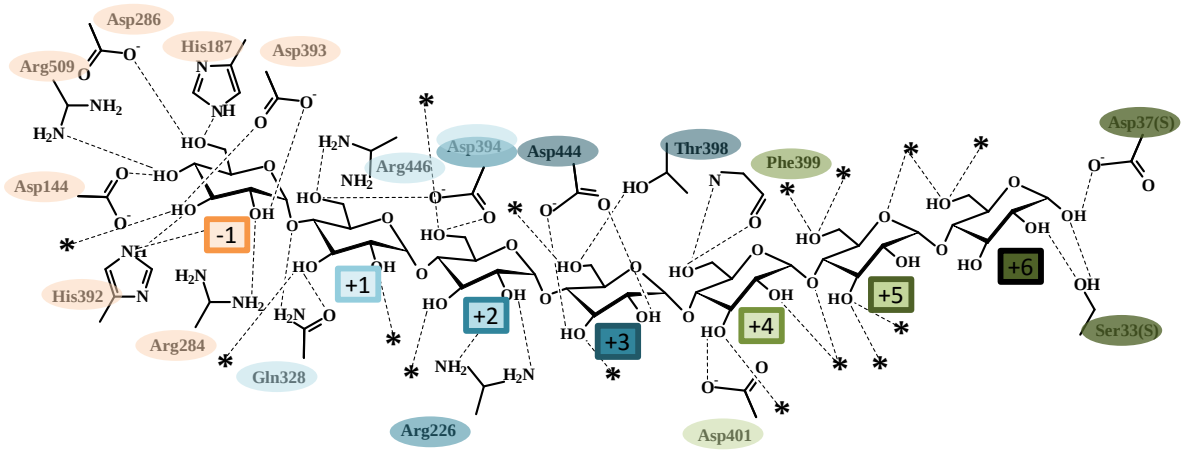


Figure 11: Mapping of the hydrogen bonding interaction network at OB1 site formed between maltoheptaose and E328Q-*Np*AS (PDB code: 1MVY). *Dotted lines represent hydrogen bonds; amino acids in interaction with subsite -1 are colored in orange; with subsite +1 in light cyan; with subsite +2 in cyan; with subsite +3 in dark cyan; with subsite +4 in light green; with subsite +5 in green and with subsite +6 in dark green. * Represents water molecules.* (adapted from Skov, L.K. *et al.*, 2002).

d- Mechanism of action

Amylosucrase follows the double-displacement mechanism common to all enzymes of the GH13 family. The α-retaining mechanism involves two steps **(Figure 12).** In the **first step**, donor binds at the bottom of the catalytic pocket with the glucosyl ring occupying the -1 subsite and the fructosyl ring the +1 subsite. Next, the Glu328 acid/base assists the nucleophilic attack exerted by Asp286 nucleophile. Glu328 gives its proton to the anomeric oxygen O_1 while Asp286 attacks the anomeric carbon of the glucosyl ring at -1 position. The reaction goes through the formation of an oxocarbonium leading to the formation of the **covalent β-glucosyl-enzyme intermediate** with an **inversion of the anomeric carbon configuration**. In a **second step**, the deprotonated Glu328 serves as base to activate the hydroxyl group of an acceptor molecule, which can be either water (hydrolysis reaction) or the hydroxyl group of an acceptor (transglucosylation)[77]. Additional residues play also an important role in catalysis, such as Arg284 which enables the proper catalytic positioning of Asp286, His187 and His392 that contribute to the stabilization of the transition state. Tyr147 and Phe250 are also key residues that maintain the glucosyl ring at -1 subsite through stacking interactions[75].

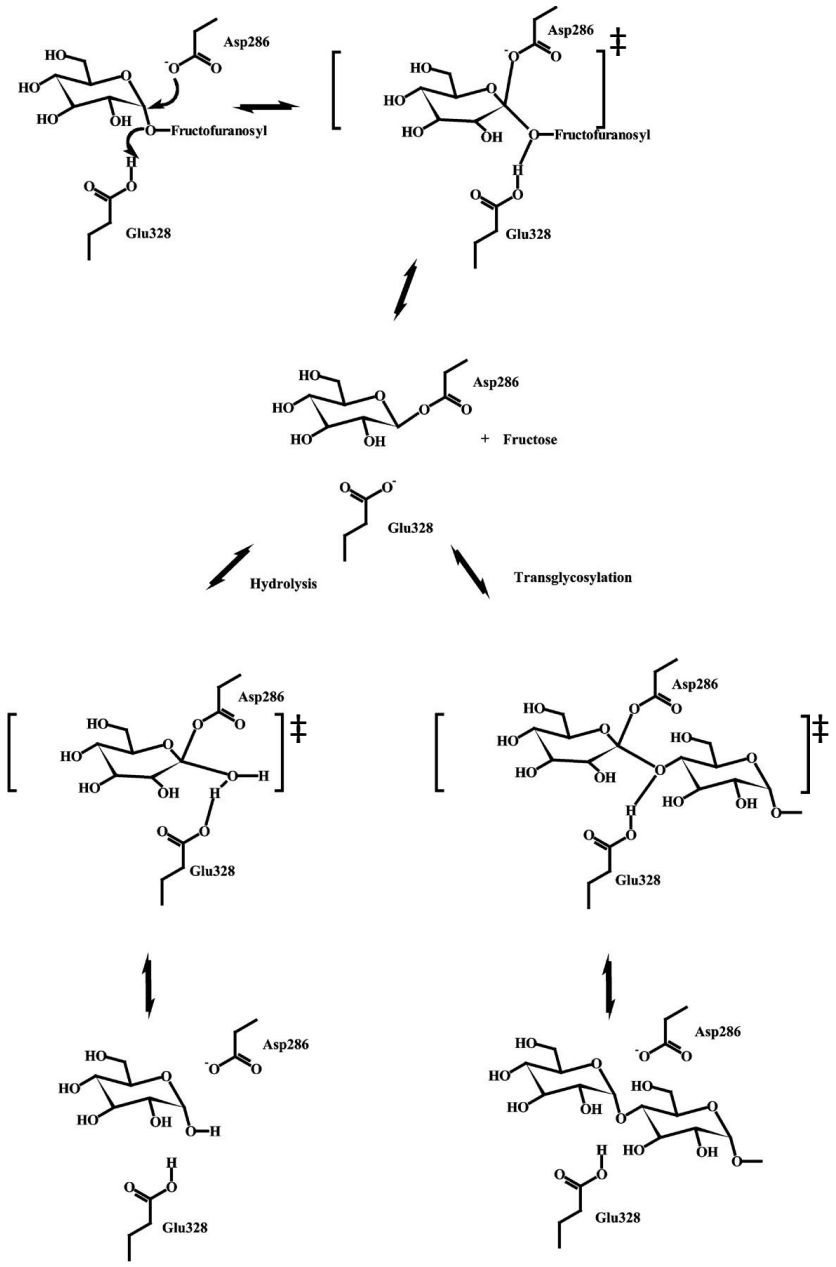


Figure 12: Retaining mechanism of sucrose-utilizing α-glycoside-hydrolases and α-transglycosylases. *Numbering of catalytic residues is that of* Np*AS.* (Skov L. *et al.* 2001)[70].

Polymer formation is the result of successive transglucosylation reactions occurring at the non-reducing end of maltooligosaccharides through a multi-chain process of elongation **(non-processive elongation)**. Glucose is first released from sucrose hydrolysis and then acts as acceptor to yield maltose which in turn will be glucosylated and so forth. Some residues from OB1 binding site, such as Arg415 located at +4 subsite, Asp394 and Arg446 at +1 subsite are described to be important for the oligosaccharide elongation **(Figure 11)**, as their mutation was shown to affect the size of produced maltooligosaccharides. Another residue, the Arg226 at subsites +2/+3, was also shown to play a major role in the polymerization as its replacement by a smaller amino-acid is accompanied by synthesis of longer malto-oligosaccharides and increase of insoluble glucans produced during the course of the reaction[78, 60].

e- Glycosylation of non-natural substrates

Glucoconjugates and oligosaccharides play an important role in host-defense, cell-cell recognition, ligand-receptor binding and cell signaling. They find numerous applications as therapeutics, food ingredients and/or fine chemicals. However, their chemical synthesis can be tedious due to the lack of regio-specificity of glycosylation reactions, requiring thus multiple protection-deprotection steps of reactive hydroxyl groups under harsh conditions which limit yields and production rates. In this context, enzymatic synthesis offers an attractive alternative thanks to the regio- and stereo- selectivity of enzymes and their mild reaction conditions[79]. In particular, the ability of amylosucrases to transglucosylate exogenous acceptors has been widely explored to produce novel carbohydrate-based molecules[80, 81, 82, 83, 84] **(Table 5)**. Indeed, natural amylosucrases have revealed a high plasticity of their +1 subsite which is able to accommodate and transglucosylate many different types of hydroxylated molecules, emphasizing thus the tremendous acceptor promiscuity of amylosucrases. Conversely, *Np*AS has shown a very narrow donor substrate promiscuity, mainly due to a strong specificity for the glucosyl moiety at -1 subsite. Attempts to enlarge the repertoire of donor substrates recognized by amylosucrases have been quite unsuccessful so far **(Table 6)** as sucrose remains so far the preferred donor substrate[85].

Table 5: Molecules recognized as acceptors by amylosucrases.

Acceptor molecule	Enzyme / organism	Glucosylated product(s)	Applications	References
Salicin	*Dg*AS ---------- *Np*AS	α-D-glucopyranosyl-(1,4)-salicin ---------- α-D-glucopyranosyl-(1,4)-salicin α-D-glucopyranosyl-(1,4)-α-D-glucopyranosyl-(1,4)-salicin	Analgesic and antipyretic compound	81
Glycerol	*Mf*AS	(2R/S)-1-*O*-α-D-glucosyl-glycerol 2-O-α-D-glucosyl-glycerol via an α-(1,4) glycosidic linkage	Better thermostable, low heat-colorability, low Maillard reactivity, low hygroscopicity, high water holding capacity, non-cariogenicity and low digestibility (compared to sucrose)	51
Hydroquinone	*Dg*AS	α - arbutins	Skin whitening agent Cosmetic and medical uses	83
(+)-catechin	*Dg*AS	(+)-catechin-3'-O-α-D-glucopyranoside (+)-catechin-3'-O-α-D-maltoside	Skin-whitening application Antioxidant activities (protective effect against cardiovascular diseases and cancers)	82
α-allyl-*N*-acetyl-D-glucosamine **L-rhamnose**	*Np*AS	α-D-glucopyranosyl-(1,6)-*N*-acetyl-D-glucosamine α-D-glucopyranosyl-(1,4)-*N*-acetyl-D-glucosamine α-D-glucopyranosyl-(1,1)-β-L-rhamnopyranoside α-D-glucopyranosyl-(1,4)- α-D-glucopyranosyl-(1,1)-β-L-rhamnopyranoside	Carbohydrate-based vaccine against Shigella disease	86
Glycogen	*Np*AS	Polymer of α-(1,4) linkages	n.d.	87, 84
D- and L- arabinose **D- and L-galactose** **D- and L-altrose** **D-fucose** **D- and L-allose** **D- and L-xylose** **D- and L-mannose** **D-sorbitol** **Xylitol** **D-arabitol** **D-mannitol** **Myo-inositol** **Maltitol**	*Np*AS	n.d.	n.d.	85
α-D-maltopyranosyl fluoride	*Np*AS	n.d.	n.d.	88
Allose **Mannose** **Galactose**	*Np*AS	n.d.	n.d.	89
Piceid **Aesculin** **Arbutin**	*Am*AS	α-D-glycosyl-(1,4)-piceid	Antioxidant polyphenolic compound Cardiovascular protective effects	90
β-arbutin	*Dg*AS	4-hydroxyphenyl-β-maltoside	Inhibitor of tyrosinase- Whitening effect on skin	91
Phytoglycogen **Waxy maize Amylopectin** **Normal maize flour** **Hydrothermically-treated wheat flour** **Limit dextrins A and B** **Waxy maize starch** **Lintnerized potato starch**	*Np*AS	α-glucans	n.d.	28

Table 6: Molecules recognized as donor by amylosucrases.

Enzyme/organism	Donor molecule	References
Amylosucrase from *N. polysaccharea*	Maltose (G_2) Maltotriose (G_3) Maltotetraose (G_4) Maltopentaose (G_5) Maltohexaose (G_6) Maltoheptaose (G_7) pNP-G_2 to pNP-G_7 ---------- α-D-glucopyranosyl fluoride ---------- pNP-glucose	61, 88, 85
Amylosucrase from *N. perflava*	α-D-glucopyranosyl fluoride ---------- α-D-galactopyranosyl-1,2-β-D-fructofuranoside	92, 93

II. Amylosucrase Engineering

II.1. Enzyme engineering approaches

The large diversity of enzymes existing in nature offers a large panel of activities and specificities. Enzymes have demonstrated their usefulness for many industrial applications and have a good potential for the development of green enzyme-based processes. However, they also often suffer from a lack of robustness in terms of stability or efficiency. In addition, they may not possess the requested specificity, the latter being sometimes too narrow or in contrast too broad for a given reaction. Enzyme engineering approaches have been developed to circumvent those drawbacks, develop new catalysts and adapt enzymes to industrial usages in terms of pH, temperature, resistance to solvent, efficiency, stability, expression, substrate specificity and elimination of non-desired functions[94].

Enzyme engineering has first started in the sixties with rational site-directed mutagenesis to explore structure-function relationships[95,96] but really gained in importance with the introduction of the "directed molecular evolution" concept proposed by William Stemmer and Frances Arnold in the nineties[97, 98, 99, 100, 101]. Nowadays, the trend is to mix both rational and random approaches to propose semi-rational strategies relying on computer-aided design and which aims at increasing the odds to identify a positive hit by focusing the evolution on sequence regions identified to be crucial for the wanted property, while reducing the size of mutant libraries and screening efforts **(Figure 53)** .

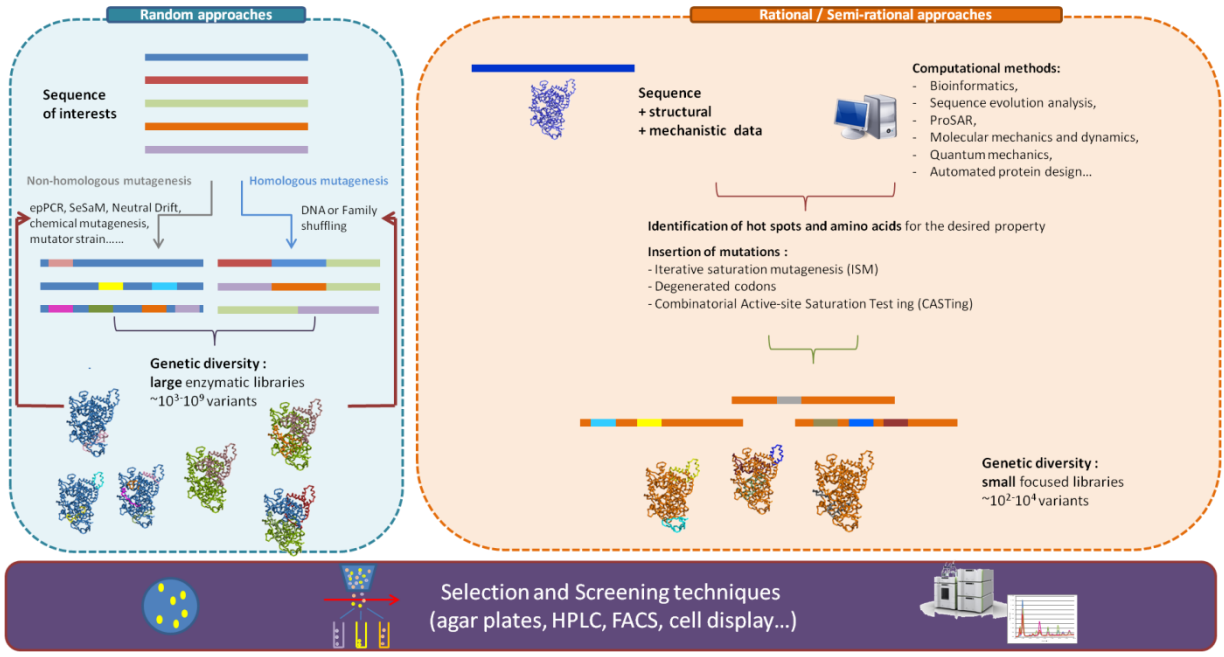


Figure 53: Different approaches of enzyme engineering. *Random approaches of engineering consist of iterative rounds of error-prone PCR or DNA-shuffling; Rational approaches and semi-rational approaches are based on structural, sequence and mechanistic data and computational methods. Sequence libraries are then screened using different types of methods (selection, solid medium screening, analytical methods….)*

a- Random approaches of engineering

These approaches, also named **"directed molecular evolution"**, were initially developed to mimic **Darwinian evolution** in laboratory and short time frame. The methods consist of iterative rounds of *in vitro* mutations and/or recombinations coupled with phenotypic selection or screening to isolate the variants showing the desired properties **(Figure 53)**. The size of the libraries is usually ranging from 10^4 to 10^9 mutants[102]. No structural or mechanistic knowledge is required, only gene functional expression is essential. They are often employed to improve catalytic efficiency, thermostability, solvent tolerance or expression.

Directed molecular evolution

To generate diversity, **error prone PCR** remains the most employed method **(Figure 53)**. Taq DNA polymerase (DNA polymerase I from *Thermus aquaticus*) can be used with excess Mg^{2+} or Mn^{2+} to perform high error rate of mutations. Some other low fidelity DNA polymerases, specifically engineered for this purpose (Mutazyme Kit (Stratagene)[103]) or natural ones (MutaGen^{TM} [104]) can also be used as well as chemical products[105], physical agents[106,107] and mutator strains[108,109, 110, 111]. Other approaches have been proposed to overcome limitations of epPCR, better control mutational bias and facilitate mutations of consecutive amino acid positions[112, 113]. For extensive review, one can consult the article from Ruff *et al.*[114] but also associated articles that describe Multicodon Scanning MuTagenesis (MCST)[115], 2'-deoxyinosine 5'-triphosphate (dIT)P-PCR[116], Prolonged Overlap Extension-PCR[117] (POE-PCR), Meganneal[118], Tandem Repeat Insertion[119] (TRINS), error-prone multiply-primed Rolling Circle Amplification[120] (epRCA), Sequence Saturation Mutagenesis (SeSaM)-Tv-II[121] and (SeSaM)-Tv-III[122].

Gene recombination may also be coupled to epPCR to further expand genetic diversity. The first method of ***in vitro* recombination** was proposed by Stemmer *et al.*, in 1994[101]. In this method, the variants of a gene are usually generated by epPCR and then digested by DNAse I. Then, fragments are ligated by PCR without primers through the formation of heteroduplexes. The entire gene is reconstituted using a final PCR with primers corresponding to the gene extremities **(Figure 53)**. Family shuffling follows the same procedure except that the shuffled genes are naturally related genes. The procedure results into chimeric libraries of the target gene. The technique is not very efficient for long gene recombination (> 3 kbp). The number of recombination points per sequence is limited to regions of high sequence identity and is usually comprised between 1 and 4[123]. The main advantage in comparison to random mutagenesis is that there is a limited risk of insertion of stop codon or destabilizing amino acids. Most of the time, the recombinants are correctly folded and functional[124, 125].

To improve the method, a lot of variant procedures were then proposed notably to increase the number of crossover points and promote recombination between variants showing low percentage of identity. These methods are reviewed in Ruff *et al*[114] and include, SISDC (sequence-independent site-directed chimera-genesis)[126], RACHIT[127] (RAndom CHImeragenesis on Transient templates), ITCHY[128, 129] (Incremental Truncation for Creation of Hybrid enzymes), SCRATCHY[130, 131], SHIPREC[132, 133] (Sequence Homology–Independent Protein RECombination), OE-PCR[134] (Overlap Extension PCR), Staggered Extension Process (StEP)[135], metagenomic gene-specific methods ((TMGS-PCR)[136], user-friendly DNA recombination ((USERec)[137], GoldenGate Shuffling (GGS)[138], PhosphorThioate Recombination ((PTRec)[139] and Integron[140].

Directed evolution generates large libraries of 10^3 to 10^9 variants and generally accumulates a high rate of deleterious mutations and a low frequency of beneficial mutations. The larger the library is, the greater is the chance to find a positive hit. One challenge is thus to reach a compromise between the size of the sequence space to cover and the chance to obtain a positive hit. For a protein with N amino acids with M targeted substitutions, the number of variants is $19^M[N!(N-M)!M!]$[141]. Isolation of positive hits rapidly becomes very challenging and necessitates efficient assays for selection or screening **(cf. p 54)**. To limit the size of the library to be screened, another evolutionary approach inspired by the Neutral Theory of Evolution (NTE), and called *Neutral Drift* [142], has been proposed and applied to the evolution of various enzymes [143, 144 , 145, 146].

The **Neutral Theory of Evolution**[147, 148, 149], first proposed by Kimura *et al*., is based on the observation that neutral mutations occurring under non-adaptative conditions were conserved as they do not impact natural protein function and were starting points to increase genetic diversity at the origin of new phenotypes. This theory was recently revisited by protein engineers to examine the polymorphism of neutral libraries of paraoxonase and cytochrome P450 and their ability to transform naturally poorly recognized substrates of different structures[150, 151, 152, 143, 153]. From small libraries generated by epPCR, only a hundred of variants showing the natural phenotypes are usually selected and re-submitted to iterative cycles (3 to 5 cycles) of mutations and selection on the natural phenotype **(Figure 3)**[153]. Using this approach, PON1, a lactonase that uses as preferred substrate (a lipophylic lactone) and shows some latent promiscuous activities (arylesterase, phosphotriesterase and aromatic lactonase) was evolved[154]. A GFP protein was fused to PON1 to serve as a fluorescent reporter of properly folded variants[144]. Three iterative rounds of random mutagenesis (performed by epPCR) and selection (using *in vitro* compartmentalization in double-emulsions and FACS) **(cf. page 56)** were applied to isolate properly folded mutants expressing lipo-lactonase activity[155]. About 300 variants were defined as neutral or close to neutral, half of them exhibiting changes into their latent promiscuous activities, specificities or inhibitions[153] **(Figure 14)**. The same strategy was applied to find mutants exhibiting improved specificity and activities towards toxic organophosphates. The 21 best mutants were then recombined by DNA shuffling to evaluate the potential of neutral drift libraries as starting point for directed evolution. 77 variants, from a library of 360 mutants, showed an increased catalytic efficiency (up to 77-fold) towards a fluorogenic organophosphate (CMP-MeCyC). The same strategy allowed to isolate 10 variants with improved activities against TBL (ϒ-thiobutyrolactone) showing that neutral drift can provide a small-size but very useful platform of genetic diversity for subsequent directed evolution experiments[153].

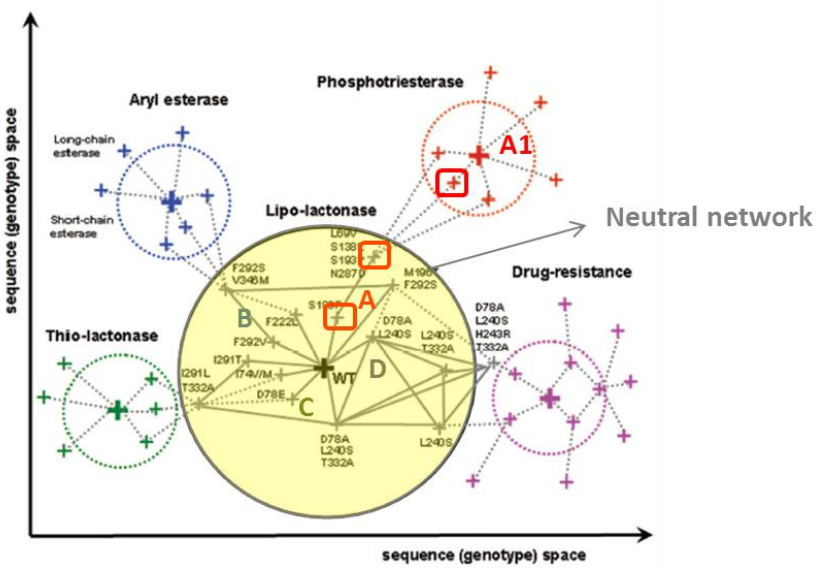


Figure 14: Neutral Drift approach applied to PON1 enzyme. *The length of edges corresponds to their neutrality state; neutral variants are scattered in a distance of 1 from the wild-type enzyme (WT); nearly neutral variants are scattered around the WT in a distance of 2; and possibly neutral in a distance of 3. The large circle denotes the boundaries of the neutral network of PON1 native phenotype; the dashed borders and edges are hypothetical. Filled edges are based on common mutations. Red squares represent transition points. A, B, C, D, A1, mutant A; B, C, D, A1.* (adapted from Amitai G. *et al.*, 2007)[153].

Over the past few years, there has been a spectacular advent of computational methods and tools to support rational and semi-rational engineering. Remarkable progresses have been achieved in this field which will be briefly overviewed in the following section and illustrated by some examples of applications.

b- Rational/Semi-rational engineering

Rational engineering is usually driven by sequence and structural knowledge available on the enzyme to evolve. The use of bioinformatic tools enabling **sequence alignments** (SA), **phylogenetic** and **co-evolutionary analyses** is relatively widespread to investigate sequence variability, ancestral relationships and for identifying signatures or consensus sequences putatively related to the desired properties. On the basis of these analyses, key amino acids can be identified and targeted by site-directed mutagenesis to verify their functional role or alter and improve properties of enzymes. Many tools are available over the web to perform SA analysis such as ClustalW[156], BLAST[157], Muscle[158], JANUS[159], Zebra[160]. Co-evolutionary and/or phylogenetic analyses (using PHYLIP (PHYLogeny Inference Package) and PhyML (Phylogenetic estimation using (Maximum) Likelihood)) can also be used for the reconstruction of ancestral biocatalysts[161, 162].

Rational engineering has also benefited from the increasing number of **high-resolution three-dimensional protein structures**, in particular when bound to ligands, to elucidate molecular determinants involved in catalytic activity and substrate specificity, which could be targeted by mutagenesis to alter properties of native enzymes[268]. Statistical analysis of sequence and structural data gathered in databases has also led to the development of computational tools and softwares to predict with high accuracy, the **secondary and tertiary structures** of novel proteins, providing insightful structural information on novel and unstudied proteins and enabling to estimate or anticipate the effect of mutations onto protein structural stability. Among others, webservers are available for comparative 3D-modelling such as I-TASSER[163],

Swiss-Model[164], Modeller[165], HHpred2[166] and others to estimate protein stability upon mutation such as FoldX, PoPMuSiC[167] and I-Mutant 3.0[168].

The use of computational tools combining 3D structural information to sequence data is another alternative to strengthen predictive models and guide enzyme design. One remarkable example is coming from the algorithm SCHEMA which identifies amino acid fragments within a set of related enzymes that can be interchanged with minimal disruption of interactions and preservation of the protein fold[169]. The Bio-Prodict 3DM system has also provided insightful information for the engineering of novel substrate specificity and selectivity. By analyzing frequency of amino acid residues found at each position within a protein superfamily and amino acid correlations, 3DM enables to identify key positions for a given function and was used to drive the construction of small enzyme libraries[170, 171].

Classical molecular modelling techniques are widely used to study the effects of mutations on the stability, folding and dynamics of proteins, using more or less CPU consuming methods. The empirical force field FoldX[172] is probably one of the most commonly used for rapid evaluation of the free energy. It is when alteration of substrate specificity of enzymes is searched that computational tools are the most insightful, in particular to study **enzyme-substrate recognition and interactions** using **molecular docking methods**. Among widely used molecular docking programs are AutoDock[173], FlexX[174] or GOLD (Genetic Optimisation for Ligand Docking)[175]. Derived information can then be used to verify by mutagenesis the role of amino acids in substrate recognition and suggest mutations that could lead to optimized properties.

Another class of computational tools has also been developed to investigate formation of **channels and tunnels**, what is of particular interest for buried active sites to describe solvent access, substrate entrance and product release. Softwares such as CAVER[176], MOLE[177] and Biomove3D[178,179] enabled to guide the redesign of substrate specificity, modulation of enantioselectivity or regioselectivity of different types of enzymes. For illustration, Biedermannova *et al.* showed how re-engineering of substrate access tunnels of an haloalkane dehalogenase, led to drastic effects on the rate of product release[180]. In similar way, Lafaquière *et al*. controlled enantioselectivity of the lipase from *B. cepacia* by introducing specific mutations that affected selectively substrate access to active site[181].

Worth mentioning, quantum Mechanics-based methods, such as **QM/MM** (**Quantum Mechanical/Molecular Mechanics**) and **CPMD** (Car-Parrinello Molecular Dynamics)[182], although very time consuming have provided outstanding information to understand in detail the molecular mechanisms of enzymes[183, 184]. However, their incorporation in enzyme design remains limited and the impact of mutations on the substrate recognition and the catalytic reactions is still poorly explored so far.

More and more, rational insight is combined to combinatorial engineering to generate focused semi-rational libraries of limited size by saturation mutagenesis and recombination[185,186]. The idea is to limit screening effort while taking advantage of knowledge to increase the chances of getting positive hits.

In practice, **site-saturation mutagenesis** is performed by PCR reactions employing **degenerated oligonucleotide primers** that target one or several sites of mutations identified by **rational means** involving either sequence or structure analysis as described above. NNN codons may be used to allow the equimolar introduction of the 4 nucleotides and the representation of the 64 codons[187]. To limit screening efforts and bias in favor of the most

represented amino acids of the genetic code, alternative codons such as NNK (32 codons encoding all amino acids) or NDT (12 codons encoding 12 amino acids) can be used[188]. The use of NDT codons allows mixing polar, nonpolar, aromatic, aliphatic negatively and positively charged residues while limiting structurally similar amino acid representation and considerably reducing screening efforts. As presented in **Table 7**, Reetz *et al.* compared the efficiency of codon degeneracy to represent 95% of desired sequences. By limiting the number of variants to screen, NDT library was better explored and led to a higher number of positive variants, along which mutants of higher catalytic efficiency. A new randomization was also investigated using MAX codon which has been shown to be a more efficient scheme compared to NNN, NNB, NNK, NDT, because it removes genetic code redundancy using one codon for one amino acid with a zero probability for a stop codon[189]. Numerous commercial kits are available to facilitate library construction (QuikChange[190], QuickChange Multi Site-Directed Kit[190] and Omnichange kit[191]).

Various computational tools have also been developed to help in the design and construction of the primers, maximize randomization efficiency, avoid stop and rare codons and limit redundancy. Among them, **GLUE-IT, PEDEL-AA, Codon-Calculator, AA-Calculator** and **DRIVeR** are well adapted for construction of libraries by epPCR or site-saturation mutagenesis and they have proven to be useful for estimating library completeness and diversity[192].

CoFinder, another algorithm, was developed to determine the optimal primer mixture for the targeted amino acid distribution. It was applied to an haloalkane dehalogenase variant (DhaA31) to convert a toxic compound, 1,2,3-trichloropropane, with an enantiomery excess of about 13% for (*R*)-2,3-dichloropropan-1-ol. Through 5 rounds of evolution targeting 25 positions on this mutant, (*R*)- or (*S*)-2,3-dichloropropan-1-ol was obtained in an enantiomery excess of about 90% and 97% respectively by two variants[193].

Mena *et al*. developed in 2005 the first algorithm named **LibDesign** for whole automated optimized design of libraries. As input, a set of sequences is given and amino acid frequency at each position is evaluated. Then, degenerate codons are designed in order to represent all the most common amino acids identified[194]. In 2007, Treynor *et al.* introduced a technique for building Degenerated Codon (DC) libraries based on structure-based computational tools that evaluate the energy between two amino acids and the different rotamers available per positions using a template structure and a rotamer library. For each degenerated codon designed, average amino acid pair energy is calculated[195]. **CASTer**[196] is another tool to design and evaluate saturation library which will be further detailed later in this section.

Table 7: Oversampling of mutants to screen for a representation of about 95% of desired sequences depending on the degenerate used codons (NNK vs. NDT) (adapted from Reetz *et al.*, 2008).

Nb. of mutated positions	Degenerate codon			
	NNK		NDT	
	Nb. of represented codons	Nb. of screened transformants	Nb. of represented codons	Nb. of screened transformants
1	32	94	12	34
2	1.028	3.066	144	430
3	32.768	98.163	1.728	5.175
4	1.048.576	3.141.251	20.736	62.118

Of particular interest, several procedures have been reported that allow coupling saturation mutagenesis to recombination techniques. **Iterative-Saturation Mutagenesis** (ISM) is a powerful tool for enzyme evolution[197, 198]. The process is illustrated as a tree with various branches depending on the chosen pathway **(Figure 15)**. Briefly, the first step consists of the identification, ranking and selection of amino acid residues that will be changed. This is often performed using computational tools such as **CAST** (Combinatorial Active-site Saturation Test) to optimize catalytic properties and **B-FIT** (B-Factor Iterative Test) to enhance thermal and organic solvent stability[199, 200]. Most often, wild-type enzyme is used as starting point. Next, the selected residues are grouped into different multi-residue sites (depending on their spatial distribution in the 3D structure and their corresponding location in the 1D structure. Single mutations are then introduced and then, if beneficial, are recombined in the next step following an iterative process that enables the exploration of different evolution pathway. ISM has been applied to different types of optimizations: catalysis (which gathers catalytic efficiency, substrate specificity, stereo- and regio-selectivity), binding affinity, chemical and thermal stability.

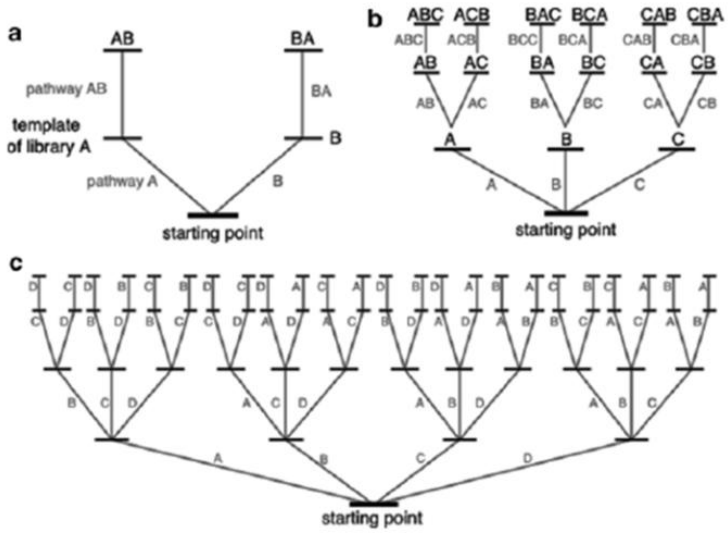


Figure 15: Illustration of ISM trees depending on the number of multisites chosen: (a) two (A and B), (b) three (A, B and C), or (c) four (A, B, C and D) multisites (from Acevedo-Roda *et al.*).

By applying ISM on 7 positions located in the active site of a phenylpropene O-methyltransferase, a novel monolignol 4-O-methyltransferase was created. One variant showed a higher catalytic efficiency (70-fold increase) and a better binding affinity (13-fold increase) to monolignols, compared to the wild-type enzyme[201]. ISM was also carried out on a homodimeric D–tagatose 3-epimerase to increase its thermostability and avoid the disintegration of the tertiary structure of one of the monomers. This semi-rational strategy led to the identification of two mutants showing improved thermostability and higher activity at elevated temperatures compared to wild-type enzyme[202].

By combining the B-FIT method and a sequence consensus approach, the thermostability of an α-amino ester hydrolase was improved by 8°C considering the temperature at which the halflife time is 30min through 4 rounds of mutagenesis leading to quadruple mutants[203]. The enantioselectivity of a lipase variant (CalA) was also improved toward the hydrolysis of different α-substituted esters. Using the CAST method originally proposed by Reetz *et al.* in 2005 and further described below, two libraries were generated targeting two different pairs of amino acids. 600 colonies were picked up and 2 variants showed a reverse enantioselectivity compared to the wild-type and improved[204].

The method **CAST** originally proposed by Reetz *et al.* in 2005 was shown to be very efficient to select the most promising positions, evaluate pair-wise mutations and expand the scope of substrate recognition by generating small focused libraries[196]. The design targets several sets of amino acid pairs, close to each other and located in or around the active site. The selection of targeted residues is usually guided by 3D structure or homology models of an enzyme with a bound or docked ligand. Amino acid residues having their side chains oriented towards the substrate are usually considered good potential targets. The method was first applied to a lipase from *Pseudomonas aeruginosa* to expand its substrate specificity by creation of 5 libraries of residue pairs. A total saturation mutagenesis was applied leading to the generation of 3,000 variants per library. 11 ester substrates were tested and 8 best mutants were identified as being able to recognize all substrates (which were not recognized at all by the native enzyme for some or better recognized than the native enzyme for others)[196].

The method was also applied to enhance the enzyme enantioselectivity toward an epoxide. A 25-fold increase was obtained in comparison to the classic single-substitution approach that led to a 7.4 increase[199]. Recently, this method was used to increase the turnover rate of a *phosphotriesterase* for malathion using the crystallographic structure of the enzyme and molecular docking of the substrate. Two variants were identified with a 5,000-fold increase in catalytic efficiency[205].

Similar to the CAST method, the **coevolving – site saturation mutagenesis** (**CCSM**) approach was developed. Cooperation between amino acid residues under selective pressure are named coevolving sites and considered as hotspots to construct focused libraries. The method was applied to an α-amylase. Six coevolving residues were targeted leading to the identification of 10 pairs of co-evolutionary interactions. Altogether, 10,010 clones were generated and one mutant was identified displaying an increase by 8°C about its half-inactivation temperature for 30 min reaction[206].

Several other procedures have been reported that allow coupling saturation mutagenesis to recombination techniques including the **ISOR** method (**Incorporating Synthetic Oligos via gene Reassembly**)[207], **synthetic shuffling**[208], **DOGS** (**Degenerate Oligonucleotide Gene Shuffling**)[209] and **ADO** (**Assembly of Designed Oligonucleotides**)[210]. Noteworthy, library size rapidly increases when mutations at multiple sites are considered.

c- Automated computational protein design

In this type of approach, we can distinguish the **redesign** of catalytic functions from existing enzymes and the ***de novo* design** of novel enzyme activities from scratch. At computational level, the main difference resides in the starting point.

For enzyme redesign, an enzyme with existing catalytic properties, often displaying promiscuous activities (ability to catalyze one major reaction plus additional side reactions involving a common catalytic mechanism), is generally used as 3D template.

For *de novo* design, there is a preliminary step that aims at screening a library of possible protein scaffolds in order to identify the most appropriate protein template to accommodate the reaction site. In parallel, a potential active site is proposed to stabilize the transition state of the targeted reaction **(theozyme)** derived from Quantum Mechanics calculations. The fit between the theozyme and the existing protein scaffolds is then investigated using different computational tools dedicated to automated computational protein design such as **RosettaMatch**[211], **RosettaDesign**[212], **Orbit**[213] or **IPRO** (**Iterative Protein Redesign and Optimization**)[214] **(Figure 16)**.

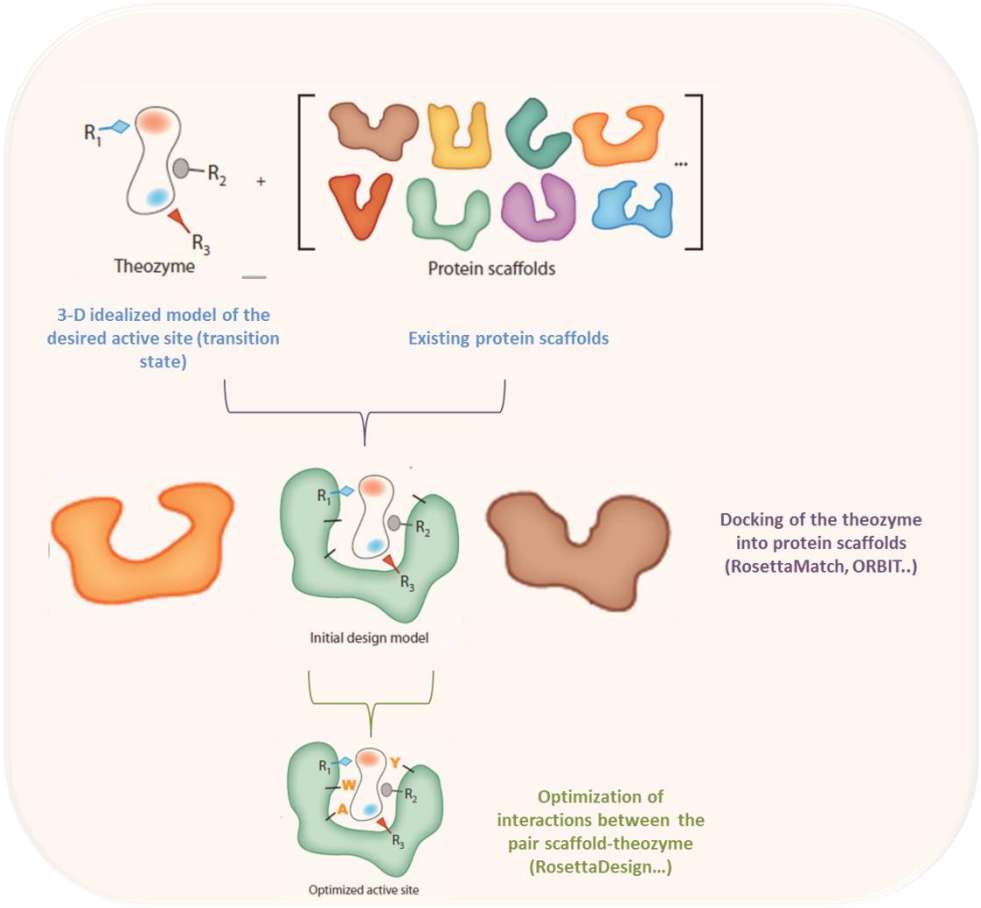


Figure 16: General strategy of automated computational protein design (adapted from Hilvert *et al* ., 2013).

Protein sequence is optimized by exploring at randomly combinations of mutations, and side-chain conformations, at selected hotspots of the enzyme active site. To adapt an active site to a specific non-natural ligand, interactions have to be optimized in order to achieve the best steric and electrostatic complementarity. Sequences are then evaluated for a given objective using an energy function (binding of a ligand, catalytic efficiency, stability, selectivity…) and ranked to provide a list of designed sequences from which the best ones can be chosen for experimental testing or used to design a sequence library of controlled diversity.

Over the past ten years, enzyme redesign and *de novo* design has been more and more applied to extend the repertoire of reactions catalyzed by enzymes and it has led to remarkable achievements which have been reported in several reviews[215, 216, 217]. In particular, enzymes able to perform Kemp-elimination, Diels-Alder and retro-aldol reactions have been computationally designed from scratch, pushing back the limits of bioengineering to propose novel catalysts for synthetic purposes[218–220]. To illustrate work performed in the field, three recent examples are briefly mentioned hereafter. The first example reports the use of a computational approach to design proteins with high affinity and selectivity for a small molecule. A ligand-binding protein was created *de novo* using RosettaDesign to incorporate five defined binding interactions with steroid digoxigenin (DIG) into a set of 401 putative protein scaffolds. Binding site sequences were optimized using RosettaDesign, leading to 17 designs that were experimentally characterized, out of which two enzymes were able to bind selectively to DIG at picomolar level, offering novel opportunities to develop small molecule

receptors for synthetic biology[221]. Another example is reported by Gordon *et al.*[222] who engineered the specificity of a serine endopeptidase called kumamolisin-As towards immunogenic portions of gluten peptides. Using RosettaDesign, a therapeutic protein containing 7 mutations was designed which displayed a 116-fold increased proteolytic activity and over 800-fold switch in substrate specificity toward immunogenic peptides[222]. Another interesting way to exploit the outcome of computational design approaches was reported by Lippow *et al.*[223]. By combining structure-based computational protein design with library-based screening, authors have gone one step further in the design of defined-sequence libraries that incorporate some designed amino acid combinations observed in computational protein design carried out with RosettaDesign. This led in only one round of mutagenesis to the engineering of a glucose 6-oxidase, improved by 400-fold compared to parental wild-type, to convert D-glucose into D-glucaric acid in a biosynthetic pathway. These results also open novel routes for metabolic engineering.

II.2. Selection and screening assays

As previously underlined, the efficiency of selection and screening assays is a pre-requisite to successfully isolate positive hits from large libraries generated by directed evolution or semi-rational design. An overview of the various methods that can be used is provided in **Figure 17**.

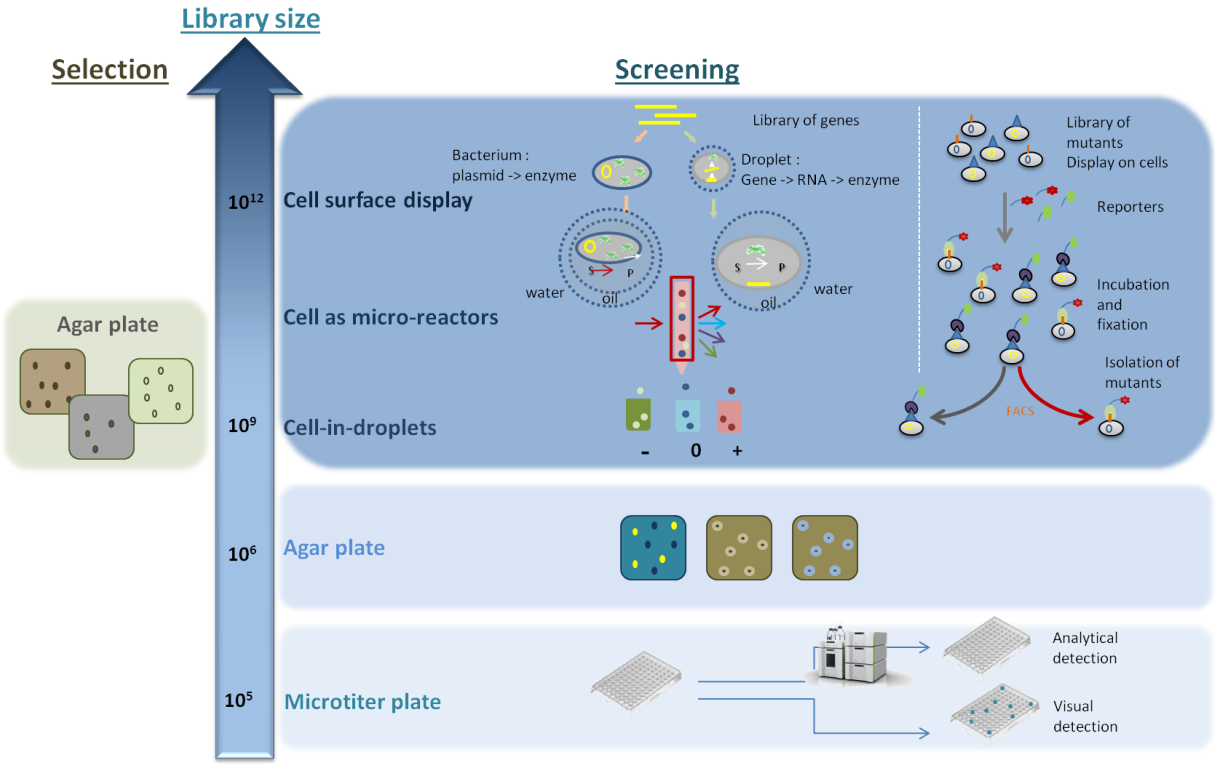


Figure 17: Overview of the different selection and screening strategies. *Selection consists of cell growth and survival on solid medium. Agar plate screening is based upon a coloration of a colony, or a halo formation. Microtiter plate screening is carried out in microplates and then clone differentiation can be done by visual determination (fluorescence, coloration) or by analytical detections (HPLC, NMR, mass spectrometry….). Then, the different compartmentalization techniques for high-throughput screening are given: cell in droplets, cell as micro-reactors and cell-surface display.*

a- Selection methods

***In vivo* selection** methods relate enzyme activity to cell survival and growth and they are fast and efficient to screen large libraries (~10^9). They can be envisaged when the targeted variants can complement an auxotrophy or neutralize lethal conditions (antibiotic resistance). Complementation of auxotrophy is limited to enzymes that synthesize products essential to survival and for which an auxotrophic host is available or can be constructed. Similarly, neutralization of lethal conditions is restricted to evolution of enzymes able to suppress the lethal factor[224,225]. Chemical complementation is another interesting alternative when the product released from enzymatic reaction can activate transcriptional factors to express genes essential to cell growth[226]. The expression host is usually *E. coli* because the bacterium is efficiently transformed and easy to cultivate. Other bacterial hosts such as *Bacillus subtilis*, *Thermus species*, *Saccharomyces cerevisiae* and *algae* as well as insect and mammalian cells were also tested. Using *in vivo* selection approaches, variants with altered catalytic efficiency[227, 103], expression[228], specificity[229, 230] or enantioselectivity[233] were isolated.

b- Screening methods

Screening methods rely on individual and fast characterization of each variant according to its biochemical and/or biophysical properties. Variants are usually produced via recombinant system of expression. Intracellular or extracellular proteins can be obtained. Alternative systems of expression such as phage, cell, and ribosome display can also be used. Then, various types of screening assays in solid, liquid medium or using microfluidic devices can be developed. Various criteria are then considered to choose the screening technology, including the property to be improved, the library size, the substrate availability, the feasibility of miniaturization etc…

Screening in solid medium

This is the simplest format of screening which relies on the isolation of colonies on agar plates containing chromogenic or fluorogenic substrates**.** Following enzymatic reaction, the colonies of interest are colored, fluorescent or surrounded by a halo. pH indicator and Schiff's reagent can also be used to allow detection of improved variants[234, 235, 236]. From 10^4 to 10^6 clones per round can be easily sorted out using automates. However, the method is usually not quantitative. Screening in solid medium is often used as primary screening to isolate active variants, which are subsequently more precisely characterized in liquid assays.

Screening in microtiter plate

After isolation of plasmidic libraries on solid medium, transformants are grown in liquid medium in 96, 384 or 1,536-well microtiter plates[234]. In order to increase the quantity of biological material, deep-well plates can also be used. Another strategy is to use a micro-Petri dish (microbial culture chip) which allows the transfer of colonies up to 10^6 wells[237]. It is a highly porous ceramic sheet with up to one million growth compartments made by microengineering. The culture medium is put in each well by capillary action. A step of cell lysis is often required to release intracellular enzymes and assay enzyme activity. Substrate consumption or product formation can be detected by colorimetry, fluorimetry, chemoluminescence, liquid chromatography (HPLC), UV/IR spectroscopy[238], mass spectrometry[239], infrared-thermographic[240,241], capillary array electrophoresis[242], circular dichroism[243], gas chromatography (GC)[244], and Nuclear Magnetic Resonance spectroscopy

(NMR)[245]. Throughput of such screening assays is well adapted to medium-size library screening containing between 10 to 10^5 variants[234].

Microfluidic-based screening assays

Microfluidic-based assays are undoubtedly the most powerful screening techniques in terms of throughput. Screening capacity is increased by 100,000-fold in comparison with assays carried out in microtiter plates[246, 247] **(Figure 17)**. Cells can be directly screened by FACS (Fluorescence-Activated Cell Sorting) using flow-cytometry to identify and separate positive hits[248, 249]. Thousands of cells per second can be analyzed. Cells can also be encapsulated in water-in-oil emulsions or in water-in-oil-in-water double emulsions to form droplets in which cells and enzyme substrates are mixed **(Figure 18)**.

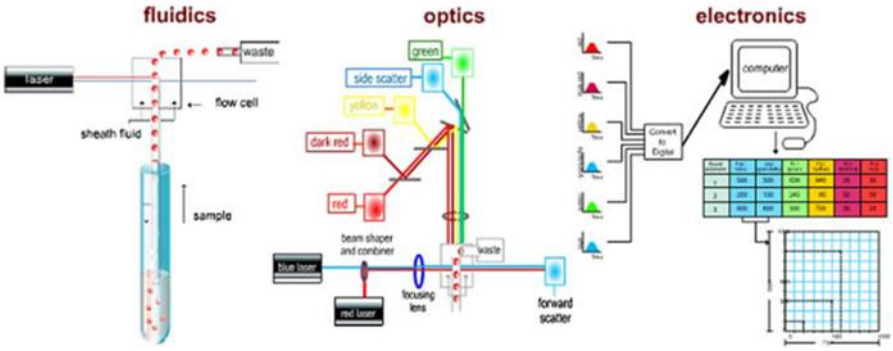


Figure 18: A schematic representation of flow cytometer using cell fluorescence: fluidic, optic and electronic components. *Cells are excited by the source, and then release light scattering information and/or fluorescence properties. They are then captured and optically separated by dichroic filters and detectors. Data are then converted and plotted in a single or two-dimensional dot plots drawing different subpopulations* (taken from Bergquist P.L. *et al*., 2009)[249].

The major drawback of FACS is related to the fluorescent labeling, which are large chemical groups that can interfere with the enzymatic reaction by steric hindrance and create a problem of transfer across the membrane. Such problems are avoided when cell surface or phage display of variant libraries are considered[234]. Alternatively, reporter-based screening assays are also very powerful methods. This method is applicable if the reaction product is a transcription factor or a molecule that can activate a reporter transcription factor[250], or a riboswitch[251]. To circumvent the limitations of FACS and of cell transformation, cell-free protein expression based on *in vitro* transcription and translation (IVTT) of a DNA (or RNA) template **(Figure 17)** in micro or nanodroplets also emerged as very powerful approaches[252, 253, 254, 255,155]. Their main advantages rely on the fact that proteins (in particular protein toxic for their host) can be directly expressed from PCR products and no background activity of contaminant native enzymes occurs **(Figure 19)**. Microfluidic system can generate up to 10,000 microdroplets per second to form a monodisperse phase. Droplets can be fused to mix substrate and enzymes. From 10 to 10^2 enzyme molecules are encapsulated per droplet and the reaction time can be finely tuned[256, 257].

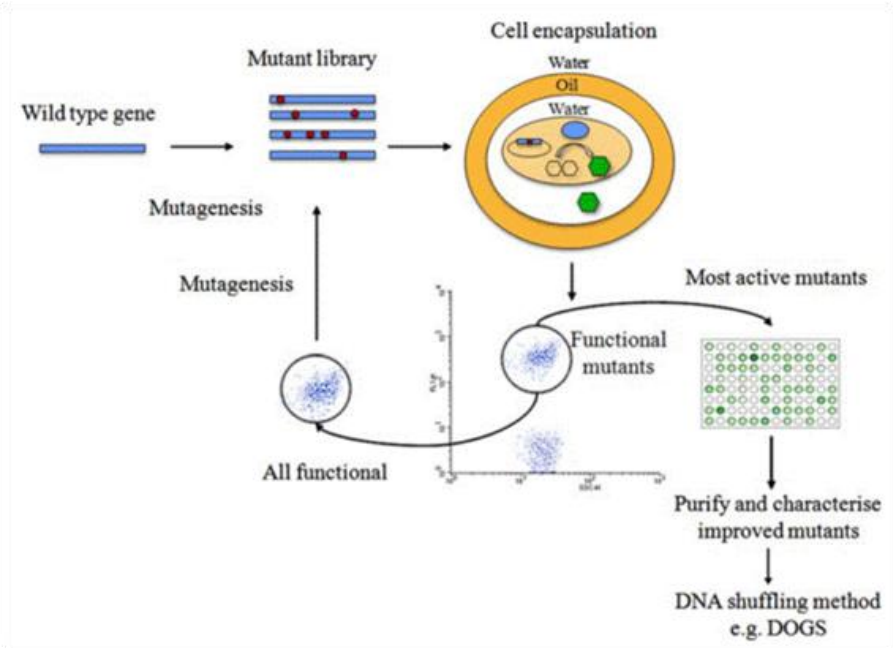


Figure 19: General process of sorting a library by encapsulation in water-in-oil-in-water droplets and FACS. *All positive hits sorted by FACS are then used for a next round of mutagenesis. The best mutants are then expressed in 96-wells microplates and then improved by DNA shuffling methods* (taken from Bergquist *et al*., 2009).

II.3. Engineering of Amylosucrases

Since the three-dimensional structure of *Np*AS was determined[32, 56, 58, 70], the enzyme has been extensively engineered by rational, semi-rational and random approaches to support structure/function studies and develop tailored catalysts with either improved performances or new properties, not exhibited by the wild-type enzyme. We will provide in the following an overview of these various studies.

a- Rational and semi-rational design

Rational engineering was first applied to determine the role of some amino acids in the reaction mechanism. Structural analysis of *Np*AS in complex with glucose and sucrose[55] allowed to map subsites -1 and +1 and the residues interacting with the substrates *ie* Glu328, Asp286, His187, Arg284, Asp393, His392, Asp144, Arg509, Asp394, Arg446, Tyr147 and Phe250 **(Figure 20)**. Among them, some are conserved in GH13 family and others not. All of them were submitted to site-directed mutagenesis to probe their role[75].

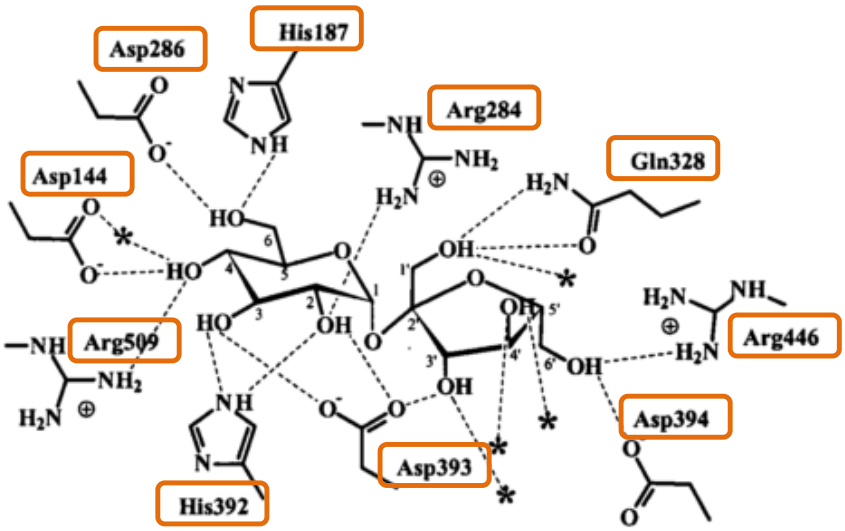


Figure 20: Schematic view of interactions between sucrose and E328Q-*Np*AS mutant into SB1 site. *Hydrogen bonds are represented in dotted lines. Waters molecules are represented by asterisks* (adapted from Mirza *et al*., 2001).

First, the 8 residues from subsite -1 which are conserved amongst GH13 enzymes were mutated by specific amino acids as described in **Table 8**.

Table 8: GH13 conserved amino acid residues from subsite -1 targeted by mutagenesis (adapted from Albenne *et al*., 2002).

Position	Wild-type amino acid	Mutation	Role
286	D	N E	Catalytic nucleophile
328	E	Q	Catalytic general acid
393	D	N E	Transition state stabilizer (Catalytic)
187	H	Q	Hydrogen bond with Glc (subsite -1)
392	H	N	Hydrogen bond with Glc (subsite -1)
284	R	V D H K	Hydrogen bond with Glc (subsite -1) Structural integrity

As expected, mutation of catalytic residues Asp286, Glu328 and Asp393 resulted in total loss of activity[62]. A drastic decrease of activity was also observed when mutations were introduced at positions 187 and 392. In wild-type *Np*AS, these two positions are occupied by histidines which have been shown to be involved in the hydrogen bonding network that stabilizes sucrose glucosyl moiety bound at -1 subsite. In similar way, mutations at position 284 from -1

subsite led to a loss of activity on sucrose, with the exception of R284K which was still active, although very poorly. This position was assumed to be crucial for structural integrity of the enzyme and to the proper positioning of the nucleophile Asp286.

Role of Y147 and F250 in sucrose specificity was also investigated **(Table 9)**. When mutated by aromatic residues, mutants were still active on sucrose but with a reduced activity: Y147F and F250Y displaying respectively 67% and 25% of the wild-type activity. Replacement by another type of residue (A or N) caused a drastic decrease of activity, indicating that both positions could contribute to sucrose activity by properly positioning the glucosyl ring in subsite -1 through hydrophobic stacking interactions with aromatic side chains.

Table 9: Aromatic amino acid residues from subsite -1 targeted by mutagenesis (adapted from Albenne *et al*., 2002).

Position	Wild-type amino acid	Mutated amino acid
147	Y	A
		N
		F
250	F	A
		N
		Y

Unconserved amino acid residues (Asp144, Arg509, Asp394 and Arg446) are specific to *Np*AS. Indeed, Asp144 and Arg509 are involved in the salt bridge located at the bottom of the catalytic pocket whereas Asp394 and Arg446 interact with the fructosyl ring of sucrose via hydrogen bonding interactions at +1 subsite. All these residues were replaced by alanine. By mutating Asp144 and Arg509, activity was totally abolished showing that the salt bridge is essential for activity and specificity. In contrast, replacement of Asp394 and Arg446 by an alanine, resulted in active enzymes but with drastically reduced activity showing that they were important but not essential residues[75].

Albenne *et al.* investigated polymerization mechanism through kinetic polymer formation and mutation of residues involved in polymer synthesis[259]. Using structural information of *Np*AS in complex with maltoheptaose (G7), four residues (Asp394, Arg446, Arg415 and Arg226) were chosen and replaced by an alanine. Arg226 forms two hydrogen bonds to the +2 or +3 glucosyl moieties and Arg415 participates to hydrophobic interactions with the +4 glucosyl unit. R415A mutation led to a drastic decrease of sucrose activity and no insoluble fraction identified indicating that Arg415 was both essential for activity but also for polymerization reaction. Mutation of Asp394 and Arg446 into alanine resulted in enzymes that were less active on sucrose (23.5% and 15%, respectively compared to the wild-type enzyme), but also less efficient to produce maltooligosaccharides of high degree of polymerization suggesting that they are important for oligosaccharide elongation.

In contrast, the introduction of an alanine at position 226 led to a lower activity on sucrose (30% of wild-type activity) and a two-fold increase of insoluble glucan formation with lower quantities of byproducts. The side chain of Arg226 was proposed to cause steric hindrance at OB1 site and to interfere with maltooligosaccharide binding and elongation[259].

Overall, this study suggested that polymerization reaction was initiated through the release of glucose from sucrose hydrolysis and that this glucose residue was subsequently used as acceptor to be elongated and form soluble maltooligosaccharides. When size of the chain was

long enough, oligosaccharides precipitated. Several amino acid residues were found important to control polymerization reaction: Arg415 for the guidance of the acceptors, Asp394 and Arg446 for the correct positioning of the acceptors. Conversely, Arg226 was found to be involved in maltooligosaccharide binding and responsible for the production of short oligosaccharide chains. Recently, Cambon *et al.* carried out saturation mutagenesis on this key position to better understand its role onto maltooligosaccharide elongation and sucrose conversion[78]. All mutants displayed an activity on sucrose enhanced by 1.5 to 6-folds compared to the wild-type *Np*AS. Some mutants, such as R226N, produced twice higher amounts of insoluble amylose compared to *Np*AS, what was attributed to an increase of the de-glycosylation rate of the reaction[78].

To generate high producers of amylose, Schneider *et al*. carried out a high-rate segmental random mutagenesis by overlap PCR toward amino acids of subsites -1 to +3[260]. 1,000 clones were generated and screened for their catalytic activity and their ability to use glycogen as acceptor. Only 6.8% of the mutants were active, among them 18% synthesized iodine-stainable products. For 82% of active mutants, only maltose and maltotriose were synthesized. The study allowed concluding that positions 394 and 396 might impact on maltooligosaccharide elongation as previously shown by Albenne *et al.*[259].

b- Directed molecular evolution

*Np*AS uses a low cost donor substrate (sucrose) and is of interest for the synthesis of amylose-like polymer and various types of gluco-derivatives **(See part I.2.e)**. Combinatorial engineering was initially applied to adapt the enzyme to industrial process and improve its thermostability and catalytic efficiency. With this objective, Van der Veen *et al*. used directed evolution to construct a library of 100,000 variants by combining random mutagenesis (with Taq polymerase and Mutazyme) and gene shuffling[103]. After selection on solid mineral medium, the mutants were characterized for increased sucrose activity and polymer formation by automated protocols. The campaign allowed isolating two mutants that were 4-times more active than the wild-type enzyme and showed a 10-fold increase in polymer synthesis. It revealed mutations far from the active site, which could not have been predicted by rational design[103].

This strategy was once again applied two years later to generate more efficient variants. Among 100,000 colonies screened for amylose production on agar plates, 90 displayed polymerase activity and showed amylose production. These active mutants were subsequently submitted to screening toward sucrose activity and thermostability. Altogether, 39 improved variants were isolated and 7 were further characterized. Among them, N387D showed a catalytic efficiency increased by 60% and E227G revealed a 2-fold increased ratio of polymerization versus hydrolysis and the synthesis of longer chains[261].

Emond *et al.* also engineered *Np*AS by directed evolution to enhance its thermostability. They applied a random mutagenesis approach based on the use of two low-fidelity DNA-polymerases (β and η human polymerases) to first replicate the parental sequence and then, amplify the gene variants by classical PCR[104]. Two libraries of 60,000 variants were generated. Altogether, 7,000 active variants were selected by growing the recombinant clones on minimum medium containing only sucrose as carbon source and screened for thermostability. Three variants showed from 3.5 to 10-folds increased half-life times at 50°C compared to the wild-type enzyme. The best mutant was a double mutant (R20C/A451T) and it was suggested that improved stability could be due to a reorganization of the salt bridge interactions and the introduction of additional hydrogen bonding[262].

In a more recent study, Daudé *et al.* applied the Neutral drift approach to evolve *Np*AS[263]. Following four successive rounds of neutral evolution, 440 neutral or nearly neutral variants, still active on sucrose substrate were isolated. From this small library, variants showing improved ability to use *p*-Nitrophenyl-α-D-glucopyranoside as donor or glucosylate various types of polyols (Xylitol, Dulcitol, Myo-inositol) were isolated. Moreover, one mutant exhibiting a totally new specificity for a non-naturally recognized acceptor (ie: methyl α-L-rhamnopyranoside) was also identified. This mutant was by 20-fold more efficient than the wild-type enzyme to use *p*-Nitrophenyl-α-D-glucopyranoside as substrate. Mutants with higher Tm (up to 5°C) were also identified by Differential Scanning Fluorimetry. This well illustrates richness of Neutral drift libraries, which may be used for very different engineering objectives.

c- Computer-aided engineering

Semi-rational engineering approaches were previously applied to *Np*AS in order to develop chemo-enzymatic pathways dedicated to the synthesis of O-antigenic oligosaccharides mimicking the lipopolysaccharide of *Shigella flexneri* pathogenic strains and that could enter in the conception of carbohydrate-based multivalent glycovaccines against Shigellosis. The *S. flexneri* O-antigen is formed by a pentasaccharide repeating motif encompassing a tetrasaccharide backbone composed of three L-rhamnopyranosyl units (A,B,C) linked by α-1,2 osidic linkages and attached to *N*-acetyl-glucosamine (D) by an α-1,3 linkage **(Figure 21)**. This backbone is usually branched with one glucosyl unit or an O-acetyl group[264, 265], the position and regiospecificity of the glucosylation being specific to each serotype. With the view of synthesizing the repeating motifs of *S. flexneri* 1b and 3a **(Figure 21)**, two chemo-enzymatic pathways were designed to avoid chemical 1,2-*cis* glucosylation, known to be very tedious and difficult, and entrusted this task to glycoenzymes. In the designed pathway, two protected sugar intermediates (ie α-allyl-*N*-acetyl-D-glucosamine and α-methyl-L-rhamnopyranoside) had to be glucosylated with the correct regiospecificity to generate α-D-glucopyranosyl-(1,4)-*N*-acetyl-β-D-glucosamine (ED) or α-D-glucopyranosyl-(1,3)-α-methyl-L-rhamnopyranoside (EA')[80,86]. A panel of native glucansucrases were tested but none of them produced EA' and only one enzyme, *Np*AS, catalyzed ED formation but with a very low yield. A computer-aided approach was thus undertaken to challenge engineering of *Np*AS variants having the requested acceptor specificity while keeping their ability to use sucrose as substrate.

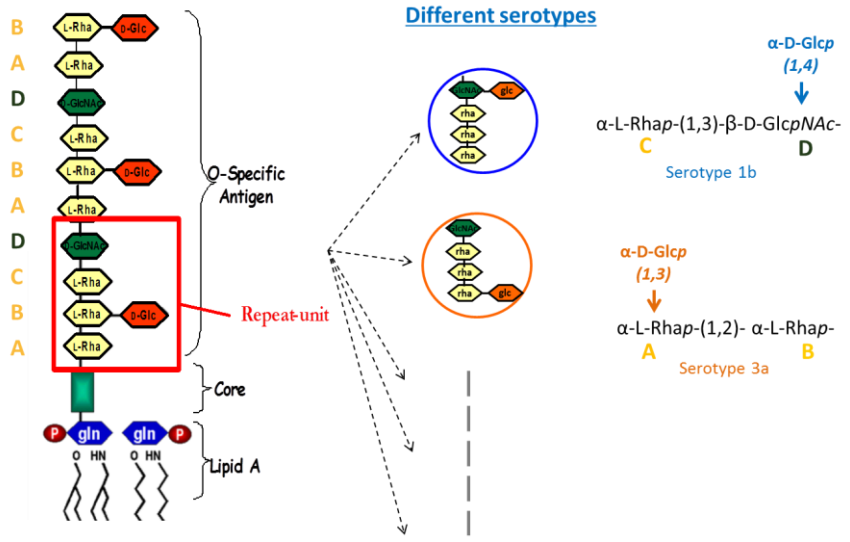


Figure 21: Representation of the lipopolysaccharide at the surface of *Shigella flexneri* bacterial surface; and the different targeted *S. flexneri* serotypes (encircled).

First, molecular docking of the two acceptor reaction products in *Np*AS active site allowed the identification of 7 positions from subsite +1 (I228, I300, D394, R446, V331, A289 and F290), potentially involved in acceptor recognition. Each position was systematically substituted by the 19 other amino acid residues to create a small library of 133 single mutants **(Figure 22, 23)**. Then, the mutants were screened for their ability to use sucrose and to glucosylate the non-natural acceptors. From this small library, 15 mutants displayed a totally new specificity towards α-methyl-L-rhamnopyranoside. Positions 228 and 290 were found the most critical for α-methyl-L-rhamnopyranoside recognition. 30 mutants were improved for α-allyl-*N*-acetyl-D-glucosamine glucosylation and showed the correct regiospecificity. In this case, position 290 was also identified to be important for acceptor recognition.

The best mutants for production of Methyl α-D-glucopyranosyl-(1→3)-α-L-rhamnopyranoside and Allyl α-D-glucopyranosyl-(1→4)-2-acetamido-2-deoxy-α-D-glucopyranoside were I228Y and F290K, respectively. They were further characterized. From 100mM sucrose, mutant I228Y and mutant F290K catalyzed the transfer of 47% and 76% of the glucosyl moieties of sucrose onto α-methyl-L-rhamnopyranoside and α-allyl-*N*-acetyl-D-glucosamine, respectively. Both mutants showed an outstanding increase in catalytic efficiency towards acceptor as revealed by their $k_{cat}/K_{macceptor}$ values **(Figure 24)**.

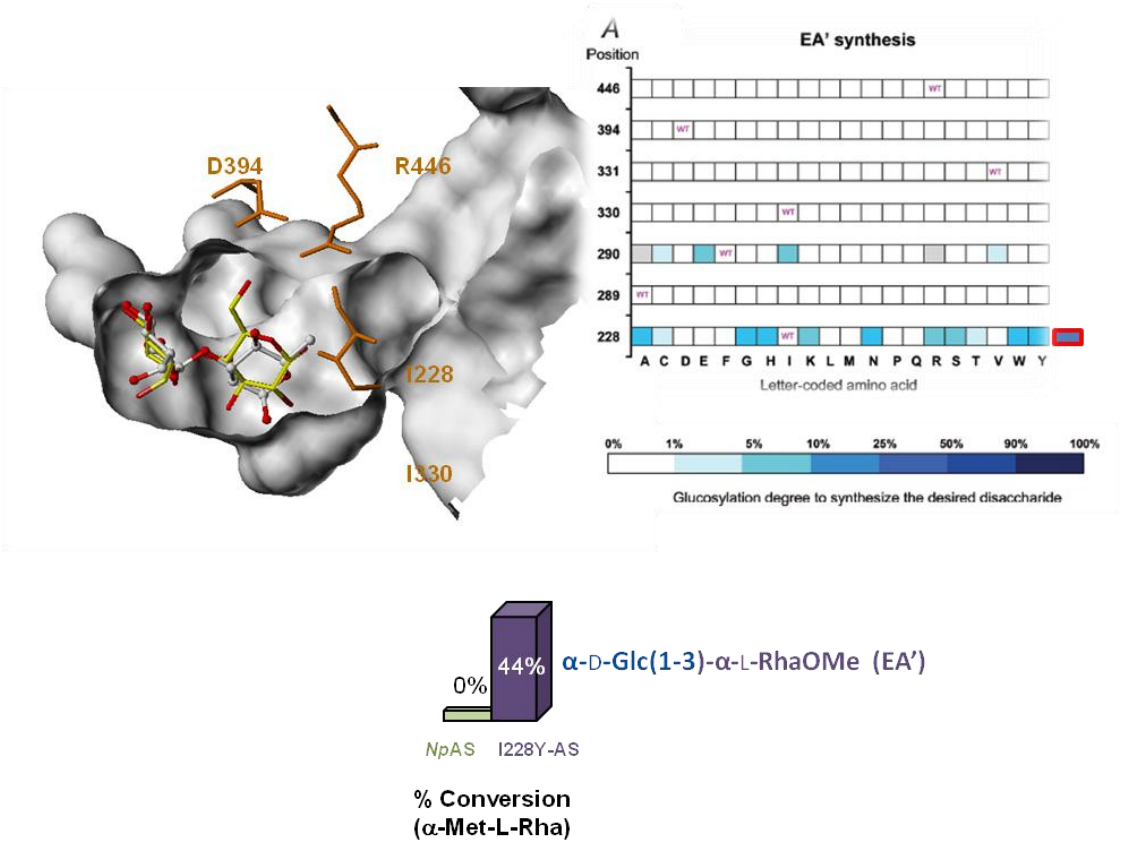


Figure 22 : A global representation of α-methyl-L-rhamnopyranoside glucosylation by mutant I228Y (from left to right): *molecular docking of the α-D-glucopyranosyl-(1,3)-α-methyl-L-rhamnopyranoside into* Np*AS active site and targeted residues; representation of the screening of the monomutant library on α-methyl-L-rhamnopyranose; percentage of acceptor conversion* (adapted from Champion *et al.*, 2009).

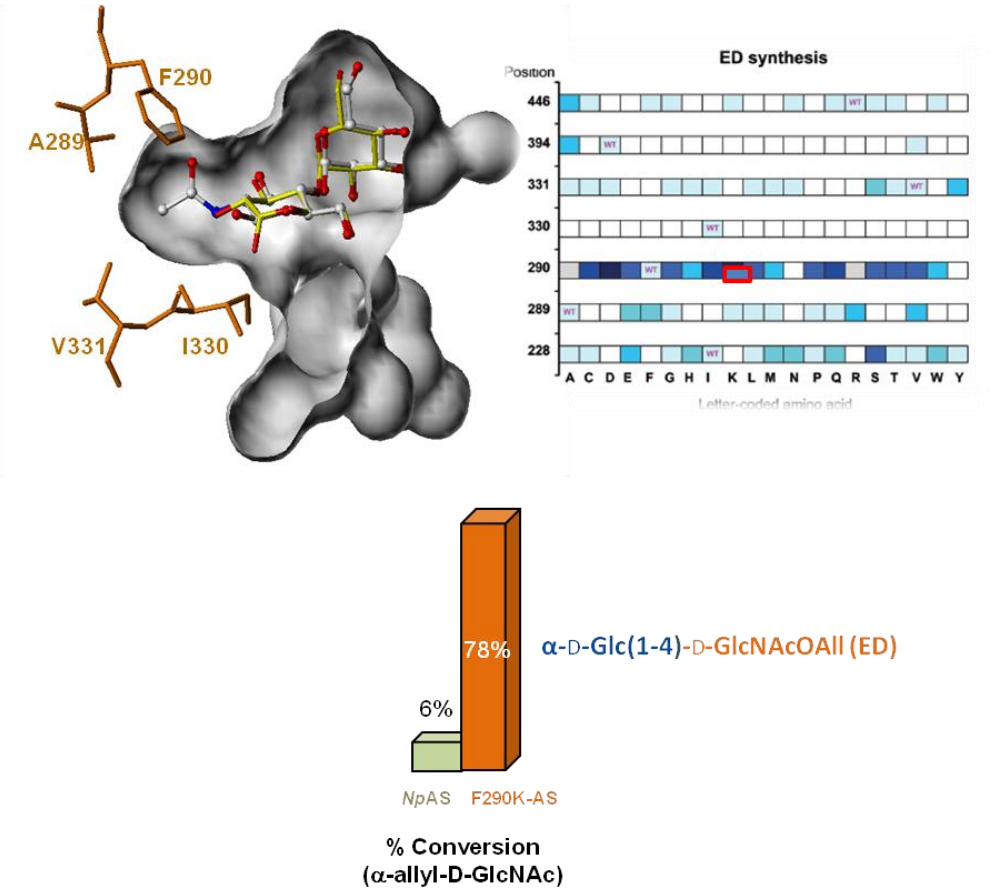


Figure 23: A global representation of α-allyl-*N*-acetyl-D-glucosamine glucosylation by mutant F290K (from left to right): *molecular docking of the* α-allyl-N-acetyl-D-glucosamine *into* Np*AS active site and targeted residues; representation of the screening of the monomutant library on* α-allyl-N-acetyl-D-glucosamine*; percentage of acceptor conversion* (adapted from Champion *et al*., 2009).

	A: Transglucosylation of A'		
		wtAS	I228Y
Characterization of acceptor transglucosylation[a]	**A'** conversion degree (%)	0	44
	% Mono-glucosylated –**A'**	0	41
	% Di-glucosylated –**A'**	0	3
	% Glc units transferred onto **A'**	0	47
Kinetic Parameters of acceptor transglucosylation	Initial rate[b] of Glc units transferred onto **A'** (µmoles of product formed/min/g_{enzyme})	<5[c]	210
	Initial rate[b] of Glc units transferred onto natural acceptors[d] (µmoles of product formed/min/ g_{enzyme})	610	65
	$k_{cat}/K_{m.\ Suc}$[e] ($mM^{-1}min^{-1}$)	$<4.10^{-4}$ [c]	0.080
	$k_{cat}/K_{m.\ A'}$[f] ($mM^{-1}min^{-1}$)	$<4.10^{-4}$ [c]	0.085

	B: Transglucosylation of D'		
		wtAS	F290K
Characterization of acceptor transglucosylation[a]	**D'**conversion degree (%)	6	78
	% Mono-glucosylated **D'**	3	56
	% Di-glucosylated **D'**	3	22
	% Glc units transferred onto **D'**	9	100
Kinetic Parameters of acceptor transglucosylation	Initial rate[b] of Glc units transferred onto **D'** (µmoles of product formed/min/g_{enzyme})	290	6690
	Initial rate[b] of Glc units transferred onto natural acceptors[d] (µmoles of product formed/min/ g_{enzyme})	630	<5[c]
	$k_{cat}/K_{m.\ Suc}$[e] ($mM^{-1}min^{-1}$)	0.074	3.630
	$k_{cat}/_{Km\ .\ D'}$[f] ($mM^{-1}min^{-1}$)	0.122	15.880

Figure 24: Kinetic Parameters determination of mutant I228Y and F290K on methyl-L-rhamnopyranoside and α-allyl-*N*-acetyl-D-glucosamine (adapted from Champion *et al*. 2009).

[a] At initial time (t0): Sucrose (Suc) = Acceptor (Acc) = 146mM. At final time (tf) = 24 h, sucrose was fully consumed. Glc = Glucosyl; Acceptor conversion degree = ([Acc]t0 - [Acc]f) / [Acc]t0 where [Acc] = Molar acceptor concentration; % Monoglucosylated Acceptor = [Monoglucosylated Acc]tf / [Acc]t0; % Diglucosylated acceptor = [Diglucosylated Acc]tf / [Acc]t0; % Glc units transferred onto acceptor derivatives = [Glucosyl units transferred onto acceptor derivatives] / [Glucosyl units transferrable from initial sucrose].

[b] Initial rate was determined at a concentration of 250mM for both donor and acceptor.

[c] The HPLC limit of detection is in the nmole range.

[d] Natural acceptors = H_2O, sucrose isomers, (Glc)n.

[e] Catalytic efficiency kcat/$Km_{,Suc}$ was determined by measuring the initial rate of monoglucosylated acceptor formation at a constant acceptor concentration (250mM).

[f] Catalytic efficiency kcat/$Km_{,Acc}$ was determined by measuring the initial rate of monoglucosylated acceptor formation at a constant sucrose concentration (250mM).

Polymerization activity was abolished for mutant I228Y and maintained for mutant F290K[185,186]. Iterative saturation mutagenesis through pairwise combinations of amino acid mutations targeting positions 290, 228 and 289 was then performed to further enhance α-allyl-*N*-acetyl-D-glucosamine glucosylation. A total of 20,000 variants was screened. Of this library, three remarkable variants, A289P-F290C, A289P-F290L and A289P-F290I, were isolated and crystallized. Their catalytic efficiency (k_{cat}/K_m) of transglucosylation was increased by 100 to 300-fold in comparison to the monomutant F290K[266].

More recently, the FoldX algorithm was used to evaluate the contribution of the amino acids forming the subsite -1 to the protein thermodynamic stability[267]. *In silico* mutagenesis of residues D144, Y147, H187, F250, R284, H392 and R509 of the subsite -1 was performed to substitute each residue by the 19 possible amino acids and determine the free energy difference between wild-type enzyme and each variant ($\Delta\Delta G$ (kcal/mol) $=\Delta G_{WT} - \Delta G_{variant}$). The T_m values of *Np*AS mutants were experimentally determined by Differential Scanning Fluorometry (DSF) and were, for the vast majority of them, consistent with computational predictions (except for positions 187 and 284). Activity using either sucrose or *p*-nitrophenyl-α-D-glucopyranoside donor substrates was also determined. Half of the mutants were still

active on sucrose **(Figure 25)**. Overall, enhancement of sucrose activity was correlated to improved stability. Analysis of transglucosylation activity further revealed 7 mutants (H187D, H392A, H392G, H392P, H392Q, H392S and Y147F), which produced shorter oligosaccharides in higher quantities. Finally, all the mutations had a negative impact on *p*NP substrate recognition. Six additional sucrose derivatives (*α*-D-galactopyranosyl-1,2-*β*-D-fructofuranoside ; *α*-D-xylopyranosyl-1,2-*β*-D-fructofuranoside ; *p*-nitrophenyl-*α*-D-glucoside ; *p*-nitrophenyl-*α*-D-galactoside ; *p*-nitrophenyl-*α*-D-mannoside ; *α*-D-galactopyranosyl-1,6-α-D-glucopyranoside-1,2-*β*-D-fructofuranoside and 1,6-dichloro-1,6-dideoxy-*β*-D-fructofuranosyl-4-chloro-4-deoxy-*α*-D-galactopyranoside) were tested as alternative substrate. None of them were used as substrate leading to the conclusion that single mutations at subsite -1 alone are not sufficient to change donor specificity[267].

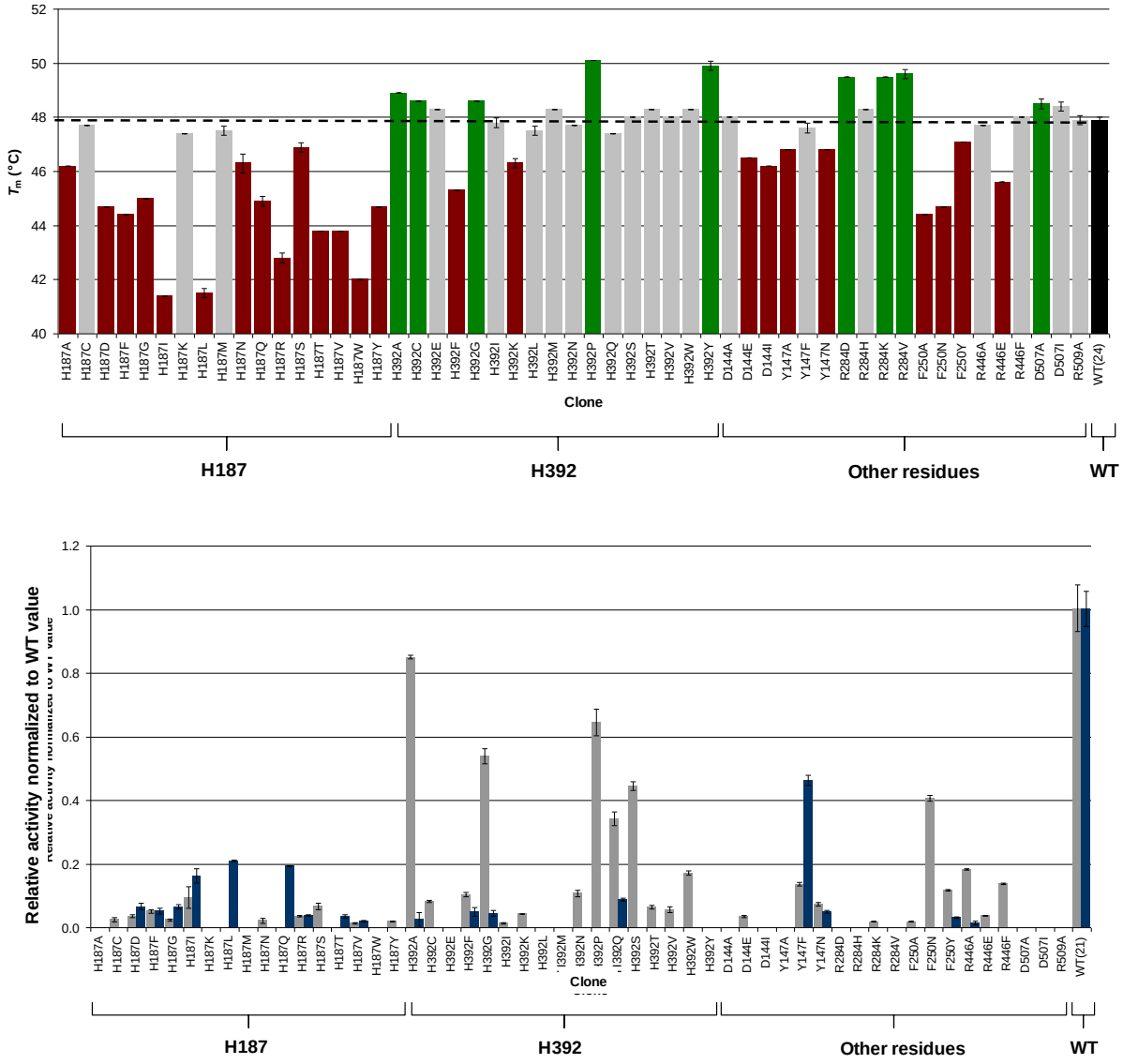


Figure 25: Analysis of purified single variants of *Np*AS obtained by staturation mutagenesis. *A: Experimental Tm values determined by Differential Scanning Fluorimetry. Variants colored in green are more stable than wild-type enzyme (> $Tm_{,WT}$ +0.6°C), variants in red are less stable than wild-type(< $Tm_{,WT}$ -0.6°C), and variants in grey are equivalent to the wild-type enzyme. B: Relative activities of variants towards sucrose (grey) and p-nitrophenyl-a-D-glucopyranoside (blue) normalized with respect to the wild-type value* (from Daude *et al.*, 2013)

References

(1) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233.

(2) Kralj, S.; van Geel-schutten, G. H.; van der Maarel, M.; Dijkhuizen, L. *Biocatal. Biotransformation* **2003**, *21*, 181.

(3) Kralj, S.; van Geel-Schutten, G. H.; Dondorff, M. M. G.; Kirsanovs, S.; van der Maarel, M. J. E. C.; Dijkhuizen, L. *Microbiology* **2004**, *150*, 3681.

(4) Tieking, M.; Korakli, M.; Ehrmann, M. A.; Gänzle, M. G.; Vogel, R. F. *Appl. Environ. Microbiol.* **2003**, *69*, 945.

(5) Tieking, M.; Kaditzky, S.; Valcheva, R.; Korakli, M.; Vogel, R. F.; Gänzle, M. G. *J. Appl. Microbiol.* **2005**, *99*, 692.

(6) Faber, E. J.; Smit, E.; Bonting, K.; Smith, R.; Brink, B. Ten; Kamerling, J. P.; Vliegenthart, J. F. G.; Dijkhuizen, L. *Appl. Environ. Microbiol.* **1999**, *65*, 3008.

(7) Monchois, V.; Remaud-Siméon, M.; Russell, R. R. B.; Monsan, P.; Willemot, R. M. *Appl. Microbiol. Biotechnol.* **1997**, *48*, 465.

(8) Shinamura, A.; Nakano, Y. J.; Mukasa, H.; Kuramitsu, H. K. *J. Bacteriol.* **1994**, *176*, 4845.

(9) Kang, H.-K.; Oh, J.-S.; Kim, D. *FEMS Microbiol. Lett.* **2009**, *292*, 33.

(10) Leemhuis, H.; Pijning, T.; Dobruchowska, J. M.; van Leeuwen, S. S.; Kralj, S.; Dijkstra, B. W.; Dijkhuizen, L. *J. Biotechnol.* **2013**, *163*, 250.

(11) Janecek, S.; Svensson, B.; Henrissat, B. *J. Mol. Evol.* **1997**, *45*, 322.

(12) Stam, M. R.; Danchin, E. G. J.; Rancurel, C.; Coutinho, P. M.; Henrissat, B. *Protein Eng. Des. Sel.* **2006**, *19*, 555.

(13) Côté, G. L.; Holt, S. M.; Miller, C. *ACS Symposium Series*; **2002**; 849, 76.

(14) Côté, G. L. *Novel Enzyme Technology for Food Applications*; Elsevier, 2007; pp. 243–269.

(15) Argüello Morales, M. A.; Remaud-Siméon, M.; Willemot, R.-M.; Vignon, M. R.; Monsan, P. *Carbohydr. Res.* **2001**, *331*, 403.

(16) KoepsellL, H. J.; Tsuchiya, H. M.; Hellman, N. N.; Kazenko, A.; Hoffman, C. A.; Sharpe, E. S.; Jackson, R. W. *J. Biochem. Chem.* **1953**, *200*, 793.

(17) Yoon, S.-H.; Robyt, J. F. *Carbohydr. Res.* **2002**, *337*, 2427.

(18) Richard, G.; Yu, S.; Monsan, P.; Remaud-Siméon, M.; Morel, S. *Carbohydr. Res.* **2005**, *340*, 395.

(19) Richard, G.; Morel, S.; Willemot, R.-M.; Monsan, P.; Remaud-Siméon, M. *Carbohydr. Res.* **2003**, *338*, 855.

(20) Demuth, K.; Jördening, H.; Buchholz, K. *Carbohydr. Res.* **2002**, *337*, 1811.

(21) Côté, G. L.; Dunlap, C. A. *Carbohydr. Res.* **2003**, *338*, 1961.

(22) Hellmuth, H.; Hillringhaus, L.; Höbbel, S.; Kralj, S.; Dijkhuizen, L.; Seibel, J. *ChemBioChem* **2007**, *8*, 273.

(23) Seibel, J.; Hellmuth, H.; Hofer, B.; Kicinska, A.-M.; Schmalbruch, B. *ChemBioChem* **2006**, *7*, 310.

(24) Nakahara, K.; Kontani, M.; Ono, H.; Kodama, T.; Tanaka, T.; Ooshima, T.; Hamada, S. *Appl. Environ. Microbiol.* **1995**, *61*, 2768.

(25) Meulenbeld, G. H.; Zuilhof, H.; Van Veldhuizen, A.; Van den Heuvel, R. H. H.; Hartmans, S. *Appl. Environ. Microbiol.* **1999**, *65*, 4141.

(26) Seo, E.-S.; Lee, J.-H.; Park, J.-Y.; Kim, D.; Han, H.-J.; Robyt, J. F. *J. Biotechnol.* **2005**, *117*, 31.

(27) Bertrand, A.; Morel, S.; Lefoulon, F.; Rolland, Y.; Monsan, P.; Remaud-Siméon, M. *Carbohydr. Res.* **2006**, *341*, 855.

(28) Rolland-Sabate, A.; Colonna, P.; Potocki-Veronese, G.; Monsan, P.; Planchot, V. *J. Cereal Sci.* **2004**, *40*, 17.

(29) Nugent, A. P. *Nutr. Bull.* **2005**, *30*, 27.

(30) Meulenbeld, G. H.; Hartmans, S. *Biotechnol. Bioeng.* **2000**, *70*, 363.

(31) Kralj, S.; van Geel-Schutten, I. G. H.; Faber, E. J.; van der Maarel, M. J. E. C.; Dijkhuizen, L. *Biochemistry* **2005**, *44*, 9206.

(32) Skov, L. K.; Mirza, O.; Henriksen, a; De Montalk, G. P.; Remaud-Siméon, M.; Sarçabal, P.; Willemot, R. M.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2001**, *276*, 25273.

(33) Machius, M.; Declerck, N.; Huber, R.; Wiegand, G. *Structure* **1998**, *6*, 281.

(34) Robert, X.; Haser, R.; Gottschalk, T. E.; Ratajczak, F.; Driguez, H.; Svensson, B.; Aghajari, N. *Structure* **2003**, *11*, 973.

(35) Brison, Y.; Pijning, T.; Malbert, Y.; Fabre, É.; Mourey, L.; Morel, S.; Potocki-Véronèse, G.; Monsan, P.; Tranier, S.; Remaud-Siméon, M.; Dijkstra, B. W. *J. Biol. Chem.* **2012**, *287*, 7915.

(36) Ito, K.; Ito, S.; Shimamura, T.; Weyand, S.; Kawarasaki, Y.; Misaka, T.; Abe, K.; Kobayashi, T.; Cameron, A. D.; Iwata, S. *J. Mol. Biol.* **2011**, *408*, 177.

(37) Pijning, T.; Vujičić-Žagar, A.; Kralj, S.; Dijkhuizen, L.; Dijkstra, B. W. *Acta Crystallogr. Sect.* F **2012**, *68*, 1448.

(38) Hehre, E. J. *J. Biol. Chem.* **1949**, *177*, 267.

(39) Hehre, E. J.; Hamilton, D. M. *J. Bacteriol.* **1948**, *55*, 197.

(40) Hehre, E. J.; Hamilton, D. M. *J. Biol. Chem.* **1946**, *166*, 777.

(41) MacKenzie, C. R.; McDonald, I. J.; Johnson, K. G. *Can. J. Microbiol.* **1978**, *24*, 357.

(42) Parker, R. B.; Creamer, H. R. *Arch. Oral Biol.* **1971**, *16*, 855.

(43) Ruby, J. D.; Shirey, R. E.; Gerencser, V. F.; Stelzig, D. A. *J. Dent. Res.* **1982**, *61*, 627.

(44) Riou, J. Y.; Guibourdenche, M.; Popoff, M. Y. *Ann. Microbiol.* **1983**, *134*, 257.

(45) Potocki de Montalk, G.; Remaud-Siméon, M.; Willemot, R. M.; Planchot, V.; Monsan, P. *J. Bacteriol.* **1999**, *181*, 375.

(46) Potocki de Montalk, G.; Remaud-Siméon, M.; Willemot, R. M.; Sarçabal, P.; Planchot, V.; Monsan, P. *FEBS Lett.* **2000**, *471*, 219.

(47) Pizzut-Serin, S.; Potocki-Veronese, G.; Van der Veen, B.; Albenne, C.; Monsan, P.; Remaud-Siméon, M. *FEBS Lett.* **2005**, *579*, 1405.

(48) Emond, S.; Mondeil, S.; Jaziri, K.; André, I.; Monsan, P.; Remaud-Siméon, M.; Potocki-Véronèse, G. *FEMS Microbiol. Lett.* **2008**, *285*, 25.

(49) Seo, D. H.; Jung, J. H.; Choi, H. C.; Cho, H. K.; Kim, H. H.; Ha, S. J.; Yoo, S. H.; Cha, J.; Park, C. S. *J. Microbiol. Biotechnol.* **2012**, *22*, 1253.

(50) Ha, S.-J.; Seo, D.-H.; Jung, J.-H.; Cha, J.; Kim, T.-J.; Kim, Y.-W.; Park, C.-S. *Biosci. Biotechnol. Biochem.* **2009**, *73*, 1505.

(51) Jeong, J.-W.; Seo, D.-H.; Jung, J.-H.; Park, J.-H.; Baek, N.-I.; Kim, M.-J.; Park, C.-S. *Appl. Biochem. Biotechnol.* **2014**, *173*, 904.

(52) Perez-Cenci, M.; Salerno, G. L. *Plant Sci.* **2014**, *224*, 95.

(53) Skov, L. K.; Pizzut-Serin, S.; Remaud-Siméon, M.; Ernst, H. A.; Gajhede, M.; Mirza, O. *Acta Crystallogr. Sect.* F **2013**, *69*, 973.

(54) Guérin, F.; Barbe, S.; Pizzut-Serin, S.; Potocki-Véronèse, G.; Guieysse, D.; Guillet, V.; Monsan, P.; Mourey, L.; Remaud-Siméon, M.; André, I.; Tranier, S. *J. Biol. Chem.* **2012**, *287*, 6642.

(55) Mirza, O.; Skov, L. K.; Remaud-Siméon, M.; Potocki De Montalk, G.; Albenne, C.; Monsan, P.; Gajhede, M. *Biochemistry* **2001**, *328*, 9032.

(56) Skov, L. K.; Mirza, O.; Sprogøe, D.; Dar, I.; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2002**, *277*, 47741.

(57) Jensen, M. H.; Mirza, O.; Albenne, C.; Remaud-Siméon, M.; Monsan, P.; Gajhede, M.; Skov, L. K. *Biochemistry* **2004**, *1991*, 3104.

(58) Skov, L. K.; Mirza, O.; Sprogøe, D.; Van der Veen, B..; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *Biocatal. Biotransformation* **2006**, *24*, 99.

(59) Buttcher, V.; Welsh, T.; Willmitzer, L.; Kossmann, J. *J. Bacteriol.* **1997**, *179*, 3324.

(60) Albenne, C.; Skov, L. K.; Mirza, O.; Gajhede, M.; Feller, G.; D'Amico, S.; André, G.; Potocki-Véronèse, G.; Van der Veen, B.; Monsan, P.; Remaud-Siméon, M. *J. Biol. Chem.* **2004**, *279*, 726.

(61) Albenne, C.; Skov, L. K.; Mirza, O.; Gajhede, M.; Potocki-Veronese, G.; Monsan, P.; Remaud-Siméon, M. *FEBS Lett.* **2002**, *527*, 67.

(62) Sarcabal, P.; Remaud-Siméon, M.; Willemot, R. M.; de Montalk, G. P.; Svensson, B.; Monsan, P. *FEBS Lett.* **2000**, *474*, 33.

(63) Thèse de doctorat, INSA de Toulouse, Albenne, C. Ingénierie rationnelle de l'amylosaccharase de *Neisseria polysaccharea* : de la structure tri-dimensionnelle aux bases moléculaires de la catalyse, 2002.

(64) Nakajima, R.; Imanaka, T.; Aiba, S. *Appl. Microbiol. Biotechnol.* **1986**, *23*, 355.

(65) Janecek, S. *Biochem. J.* **1992**, *288*, 1069.

(66) Janecek, S. *Eur. J. Biochem.* **1994**, *224*, 519.

(67) Janeček, Š. *FEBS Lett.* **1995**, *377*, 6.

(68) Janecek, S. *FEBS Lett.* **1994**, *353*, 119.

(69) Janeček, Š. *Biol. Bratislava* **2002**, *11*, 29.

(70) Skov, L. K.; Mirza, O.; Henriksen, A.; De Montalk, G. P.; Remaud-Siméon, M.; Sarcabal, P.; Willemot, R. M.; Monsan, P.; Gajhede, M. *Acta Crystallogr. Sect.* D **2000**, *56*, 203.

(71) Davies, G.; Henrissat, B. *Structure* **1995**, *15*, 853.

(72) Bozonnet, S.; Jensen, M. T.; Nielsen, M. M.; Aghajari, N.; Jensen, M. H.; Kramhøft, B.; Willemoës, M.; Tranier, S.; Haser, R.; Svensson, B. *FEBS J.* **2007**, *274*, 5055.

(73) MacGregor, E. A. *Starch-Starke* **1993**, *45*, 232.

(74) Macgregor, E. A.; Janecek, S.; Svensson, B. *Biochim. Biophys. Biochim. Biophys. Acta* **2001**, *1546*, 1.

(75) Albenne, C.; Potocki de Montalk, G.; Monsan, P.; Skov, L.; Mirza, O.; Gajhede, M.; Remaud-Siméon, M. *Biol. Bratislava* **2002**, *57*, 119.

(76) Kumar, V. *Carbohydr. Res.* **2010**, *345*, 893.

(77) McCarter, J. D.; Withers, S. G. *Curret Opin. Struct. Biol.* **1994**, *4*, 885.

(78) Cambon, E.; Barbe, S.; Pizzut-Serin, S.; Remaud-Siméon, M.; André, I. *Biotechnol. Bioeng.* **2014**, *111*, 1719.

(79) Desmet, T.; Soetaert, W.; Bojarová, P.; Křen, V.; Dijkhuizen, L.; Eastwick-Field, V.; Schiller, A. *Chemistry* **2012**, *18*, 10786.

(80) Champion, E.; Andre, I.; Moulis, C.; Boutet, J.; Descroix, K.; Morel, S.; Monsan, P.; Mulard, L. A.; Remaud-Siméon, M. *J. Am. Chem. Soc.* **2009**, *131*, 7379.

(81) Jung, J.-H.; Seo, D.-H.; Ha, S.-J.; Song, M.-C.; Cha, J.; Yoo, S.-H.; Kim, T.-J.; Baek, N.-I.; Baik, M.-Y.; Park, C.-S. *Carbohydr. Res.* **2009**, *344*, 1612.

(82) Cho, H.-K.; Kim, H.-H.; Seo, D.-H.; Jung, J.-H.; Park, J.-H.; Baek, N.-I.; Kim, M.-J.; Yoo, S.-H.; Cha, J.; Kim, Y.-R.; Park, C.-S. *Enzyme Microb. Technol.* **2011**, *49*, 246.

(83) Seo, D.-H.; Jung, J.-H.; Ha, S.-J.; Cho, H.-K.; Jung, D.-H.; Kim, T.-J.; Baek, N.-I.; Yoo, S.-H.; Park, C.-S. *Appl. Microbiol. Biotechnol.* **2012**, *94*, 1189.

(84) Putaux, J.; Potocki-Véronèse, G.; Remaud-Siméon, M.; Buleon, A. *Biomacromolecules* **2006**, *7*, 1720.

(85) Daudé, D.; Champion, E.; Morel, S.; Guieysse, D.; Remaud-Siméon, M.; André, I. *ChemCatChem* **2013**, *5*, 2288.

(86) Champion, E.; André, I.; Mulard, L. A.; Monsan, P.; Remaud-Siméon, M.; Morel, S. *J. Carbohydr. Chem.* **2009**, *28*, 142.

(87) De Montalk, G.; Remaud-Siméon, M.; Willemot, R.-M.; Monsan, P. *FEMS Microbiol. Lett.* **2000**, *186*, 103.

(88) Thèse de Doctorat, INSA de Toulouse, Potocki de Montalk, G. L'amylosaccharase de *Neisseria polysaccharea* : aspects structuraux et caracterisation de son activite catalytique, 1999.

(89) Schneider, J.; Hofer, B. *FEBS J.* **2007**, *274*, 373.

(90) Hyunsu, P.; Kim, J.; Park, J.; Baek, N.; Park, C.; Lee, H.; Cha, J. *J. Microbiol. Biotechnol.* **2012**, *22*, 1698.

(91) Seo, D.-H.; Jung, J.-H.; Ha, S.-J.; Song, M.-C.; Cha, J.; Yoo, S.-H.; Kim, T.-J.; Baek, N.-I.; Park, C.-S. *J. Mol. Catal. B Enzym.* **2009**, *60*, 113.

(92) Okada, G. *Carbohydr. Res.* **1973**, *26*, 240.

(93) Feingold, D. S.; Avidad, G.; Hestrin, S. *J. Biol. Chem.* **1957**, *224*, 295.

(94) Illanes, A.; Cauerhff, A.; Wilson, L.; Castro, G. R. *Bioresour. Technol.* **2012**, *115*, 48.

(95) Bryson, V.; Vogel, H. J. *Science* . **1965**, *147*, 68.

(96) Jensen, R. A. *Annu. Rev. Microbiol.* **1976**, *30*, 409.

(97) Moore, J. C.; Arnold, F. H. *Nat. Biotechnol.* **1996**, *14*, 458.

(98) You, L.; Arnold, F. H. *PROTEIN Eng.* **1996**, *9*, 77.

(99) Arnold, F. H. *Chem. Eng. Sci.* **1996**, *51*, 5091.

(100) Zhang, J. H.; Dawes, G.; Stemmer, W. P. C. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4504.

(101) Stemmer, W. P. C. *Nature* **1994**, *370*, 389.

(102) Eriksen, D. T.; Lian, J.; Zhao, H. *J. Struct. Biol.* **2014**, *185*, 234.

(103) Van der Veen, B.; Potocki-Véronèse, G.; Albenne, C.; Joucla, G.; Monsan, P.; Remaud-Siméon, M. *FEBS Lett.* **2004**, *560*, 91.

(104) Emond, S.; Mondon, P.; Pizzut-Serin, S.; Douchy, L.; Crozet, F.; Bouayadi, K.; Kharrat, H.; Potocki-Véronèse, G.; Monsan, P.; Remaud-Siméon, M. *Protein Eng. Des. Sel.* **2008**, *21*, 267.

(105) Singer, B.; Kusmierek, J. T. *Annu. Rev. Biochem.* **1982**, *51*, 655.

(106) Peng, W.; Shaw, B. R. *Biochemistry* **1996**, *35*, 10172.

(107) Pfeifer, G. P.; You, Y.-H.; Besaratinia, A. *Mutat. Res.* **2005**, *571*, 19.

(108) Zarif, B. R.; Azin, M.; Amirmozafari, N. *African J. Microbiol. Res.* **2011**, *5*, 651.

(109) Nguyen, A. W.; Daugherty, P. S. *Methods Mol. Biol.* **2003**, *231*, 39.

(110) Bornscheuer, U. T.; Altenbuchner, J.; Meyer, H. H. *Biotechnol. Bioeng.* **1998**, *58*, 554.

(111) Yamada, M.; Shimizu, M.; Katafuchi, A.; Gruz, P.; Fujii, S.; Usui, Y.; Fuchs, R. P.; Nohmi, T. *GENES Genet. Syst.* **2013**, *88*, 360.

(112) Wong, T. S.; Zhurina, D.; Schwaneberg, U. *Comb. Chem. High Throughput Screen.* **2006**, *9*, 271.

(113) Wong, T. S.; Roccatano, D.; Schwaneberg, U. *Biocatal. Biotransformation* **2007**, *25*, 229.

(114) Ruff, A. J.; Dennig, A.; Schwaneberg, U. *FEBS J.* **2013**, *280*, 2961.

(115) Liu, J.; Cropp, T. A. *PROTEIN Eng. Des. Sel.* **2012**, *25*, 67.

(116) Wang, Z.; Wang, H.-Y.; Feng, H. *Mol. Biotechnol.* **2013**, *53*, 49.

(117) You, C.; Zhang, Y.-H. P. *Anal. Biochem.* **2012**, *428*, 7.

(118) Pai, J. C.; Entzminger, K. C.; Maynard, J. A. *Anal. Biochem.* **2012**, *421*, 640.

(119) Kipnis, Y.; Dellus-Gur, E.; Tawfik, D. S. *Protein Eng. Des. Sel.* **2012**, *25*, 437.

(120) Luhe, A. L.; Ting, E. N. Y.; Tan, L.; Wu, J.; Zhao, H. *J. Mol. Catal. B Enzym.* **2010**, *67*, 92.

(121) Mundhada, H.; Marienhagen, J.; Scacioc, A.; Schenk, A.; Roccatano, D.; Schwaneberg, U. *ChemBioChem* **2011**, *12*, 1595.

(122) Ruff, A. J.; Marienhagen, J.; Verma, R.; Roccatano, D.; Genieser, H.-G.; Niemann, P.; Shivange, A. V; Schwaneberg, U. *J. Mol. Catal. B Enzym.* **2012**, *84*, 40.

(123) Joern, J. M.; Meinhold, P.; Arnold, F. H. *J. Mol. Biol.* **2002**, *316*, 643.

(124) Stemmer, C. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 10747.

(125) Stemmer, W. P. C. *Nature* **1994**, *370*, 389.

(126) Hiraga, K.; Arnold, F. H. *J. Mol. Biol.* **2003**, *330*, 287.

(127) Coco, W. M.; Levinson, W. E.; Crist, M. J.; Hektor, H. J.; Darzins, A.; Pienkos, P. T.; Squires, C. H.; Monticello, D. J. *Nat. Biotechnol.* **2001**, *19*, 354.

(128) Ostemeier, L. S.; Nixon, A. E.; Shim, J. H.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 3562.

(129) Patrick, W. M.; Gerth, M. L. *Methods Mol. Biol.* **2014**, *1179*, 225.

(130) Lutz, S.; Ostermeier, M.; Moore, G. L.; Maranas, C. D.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 11248.

(131) Kawarasaki, Y.; Griswold, K. E.; Stevenson, J. D.; Selzer, T.; Benkovic, S. J.; Iverson, B. L.; Georgiou, G. *Nucleic Acids Res.* **2003**, *31*, 126.

(132) Sieber, V.; Martinez, C. A.; Arnold, F. H. *Nat. Biotechnol.* **2001**, *19*, 456.

(133) Udit, A. K.; Silberg, J. J.; Sieber, V. *Methods Mol. Biol.* **2003**, *231*, 153.

(134) Kolkman, J. A.; Stemmer, W. P. C. *Nat. Biotechnol.* **2001**, *19*, 423.

(135) Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnol, F. H. *Nat. Biotechnol.* **1998**, *16*, 258.

(136) Wang, Q.; Wu, H.; Wang, A.; Du, P.; Pei, X.; Li, H.; Yin, X.; Huang, L.; Xiong, X. *J. Biol. Chem.* **2010**, *285*, 41509.

(137) Villiers, B. R. M.; Stein, V.; Hollfelder, F. *Protein Eng. Des. Sel.* **2010**, *23*, 1.

(138) Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. *PLoS One* **2009**, *4*, 1.

(139) Marienhagen, J.; Dennig, A.; Schwaneberg, U. *Biotechniques* **2012**, 1.

(140) Bikard, D.; Julie-Galau, S.; Cambray, G.; Mazel, D. *Nucleic Acids Res.* **2010**, *38*, 1.

(141) Bornscheuer, U. T. *Adv. Biochem. Eng. Biotechnol.* **2013**, *137*, 25.

(142) Kimura, M. *GENOME* **1989**, *31*, 24.

(143) Bershtein, S.; Goldin, K.; Tawfik, D. S. *J. Mol. Biol.* **2008**, *379*, 1029.

(144) Gupta, R. D.; Tawfik, D. S. *Nat. Mathods* **2008**, *5*, 939.

(145) Bloom, J. D.; Romero, P. a; Lu, Z.; Arnold, F. H. *Biol. Direct* **2007**, *2*, 17.

(146) Smith, W. S.; Hale, J. R.; Neylon, C. *BMC Res. Notes* **2011**, *4*, 138.

(147) Kimura, M. *Sci. Am.* **1979**, *241*, 98.

(148) Kimura, M. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 5969.

(149) Kimura, M. *Japanese J. Genet.* **1991**, *66*, 367.

(150) Joo, H.; Lin, Z.; Arnold, F. H. *Lett. to Nat.* **1999**, *399*, 670.

(151) Seng Wong, T.; Arnold, F. H.; Schwaneberg, U. *Biotechnol. Bioeng.* **2004**, *85*, 351.

(152) Lewis, J. C.; Arnold, F. H. *Chim. Int. J. Chem.* **2009**, *63*, 309.

(153) Amitai, G.; Gupta, R. D.; Tawfik, D. S. *HFSP J.* **2007**, *1*, 67.

(154) Draganov, D. I.; Teiber, J. F.; Speelman, A.; Osawa, Y.; Sunahara, R.; La Du, B. N. *J. Lipid Res.* **2005**, *46*, 1239.

(155) Aharoni, A.; Amitai, G.; Bernath, K.; Magdassi, S.; Tawfik, D. S. *Chem. Biol.* **2005**, *12*, 1281.

(156) Higgins, D. G.; Sharp, P. M. *Gene* **1988**, *73*, 237.

(157) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403.

(158) Edgar, R. C. *BMC Bioinformatics* **2004**, *5*, 113.

(159) Addington, T. a; Mertz, R. W.; Siegel, J. B.; Thompson, J. M.; Fisher, A. J.; Filkov, V.; Fleischman, N. M.; Suen, A. a; Zhang, C.; Toney, M. D. *J. Mol. Biol.* **2013**, *425*, 1378.

(160) Suplatov, D.; Kirilin, E.; Takhaveev, V.; Švedas, V. *J. Biomol. Struct. Dyn.* **2014**, *32*, 1752.

(161) Felsenstein, J. *J. Mol. Evol.* **1981**, *17*, 368.

(162) Guindon, S.; Gascuel, O. *Syst. Biol.* **2003**, *52*, 696.

(163) Roy, A.; Kucukural, A.; Zhang, Y. *Nat. Protoc.* **2010**, *5*, 725.

(164) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22*, 195.

(165) Sali, A.; Potterton, L.; Yuan, F.; Vanvlijmen, H.; Karplus, M. *Proteins Struct. Funct. Genet.* **1995**, *23*, 318.

(166) Hildebrand, A.; Remmert, M.; Biegert, A.; Söding, J. *Proteins Struct. Funct. Bioinforma.* **2009**, *77*, 128.

(167) Dehouck, Y.; Kwasigroch, J. M.; Gilis, D.; Rooman, M. *BMC Bioinformatics* **2011**, *12*, 151.

(168) Capriotti, E.; Fariselli, P.; Rossi, I.; Casadio, R. *BMC Bioinformatics* **2008**, *9*, S6.

(169) Damborsky, J.; Brezovsky, J. *Curr. Opin. Chem. Biol.* **2009**, *13*, 26.

(170) Kuipers, R. K.; Joosten, H.-J.; van Berkel, W. J. H.; Leferink, N. G. H.; Rooijen, E.; Ittmann, E.; van Zimmeren, F.; Jochens, H.; Bornscheuer, U.; Vriend, G.; dos Santos, V. A. P.; Schaap, P. J. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 2101.

(171) Kourist, R.; Jochens, H.; Bartsch, S.; Remko, K.; Padhi, S. K.; Gall, M.; Böttcher, D.; Henk-Jan, J.; Bornscheuer, U. T. *ChemBioChem* **2010**, *11*, 1635.

(172) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. *Nucleic Acids Res.* **2005**, *33*, W382.

(173) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.

(174) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.

(175) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R.; Uk, K. B. R. *J. Mol. Biol.* **1997**, *267*, 727.

(176) Petrek, M.; Otyepka, M.; Banás, P.; Kosinová, P.; Koca, J.; Damborský, J. *BMC Bioinformatics* **2006**, *7*, 316.

(177) Petrek, M.; Kosinová, P.; Koca, J.; Otyepka, M. *Structure* **2007**, *15*, 1357.

(178) Cortés, J.; Siméon, T.; Angulo, V. R. De; Guieysse, D. ; Remaud-Siméon M., Tran V. *Bioinformatics* **2005**, *00*, 1.

(179) Siméon, T.; Laumond, J. P.; Lamiraux, F. In *Proceedings of the 2001 IEEE International Symposium on Assembly and Task Planning: Assembly and Disassembly in the Twenty-First Century*; 2001; pp. 25–30.

(180) Biedermannova, L.; Prokop, Z.; Gora, A.; Chovancova, E.; Kovacs, M.; Damborsky, J.; Wade, R. C. *J. Biol. Chem.* **2012**, *287*, 29062.

(181) Lafaquière, V.; Barbe, S.; Puech-Guenot, S.; Guieysse, D.; Cortés, J.; Monsan, P.; Siméon, T.; André, I.; Remaud-Siméon, M. *ChemBioChem* **2009**, *10*, 2760.

(182) Kühne, T. D. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 391.

(183) Tvaroska, I.; Kozmon, S.; Wimmerova, M.; Koca, J. *J. Am. Chem. Soc.* **2012**, *134*, 15563.

(184) Ardevol, A.; Rovira, C. *Angew. Chem. Int. Ed.* **2011**, *50*, 10897.

(185) Davids, T.; Schmidt, M.; Böttcher, D.; Bornscheuer, U. T. *Curr. Opin. Chem. Biol.* **2013**, *17*, 215.

(186) Quin, M. B.; Schmidt-dannert, C. *ACS Catal.* **2011**, *1*, 1017.

(187) Oliphant, A. R.; Nussbaum, A. L.; Struhl, K. *Gene* **1986**, *44*, 177.

(188) Airaksinen, A.; Hovi, T. *Nucleic Acids Res.* **1998**, *26*, 576.

(189) Hughes, M. D.; Nagel, D. A.; Santos, A. F.; Sutherland, A. J.; Hine, A. V. *J. Mol. Biol.* **2003**, *331*, 973.

(190) Hogrefe, H. H.; Cline, J.; Youngblood, G. L.; Allen, R. M. *Biotechniques* **2002**, *33*, 1158.

(191) Dennig, A.; Shivange, A. V; Marienhagen, J.; Schwaneberg, U. *PLoS One* **2011**, *6*, 1.

(192) Patrick, W. M.; Firth, A. E. *Biomol. Eng.* **2005**, *22*, 105.

(193) Van Leeuwen, J. G. E.; Wijma, H. J.; Floor, R. J.; van der Laan, J.-M.; Janssen, D. B. *ChemBioChem* **2012**, *13*, 137.

(194) Mena, M. a; Daugherty, P. S. *Protein Eng. Des. Sel.* **2005**, *18*, 559.

(195) Treynor, T. P.; Vizcarra, C. L.; Nedelcu, D.; Mayo, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *104*, 48.

(196) Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A. *Angew. Chem. Int. Ed.* **2005**, *44*, 4192.

(197) Acevedo-Rocha, C. G.; Hoebenreich, S.; Reetz, M. T. *Directed Evolution Library Creation*; Methods in Molecular Biology; Springer New York: NY, 2014; Vol. 1179, pp. 103–128.

(198) Reetz, M. T.; Carballeira, J. D. *Nat. Protoc.* **2007**, *2*, 891.

(199) Reetz, M. T.; Wang, L. W.; Bocola, M. *Angew. Chem. Int. Ed.* **2006**, *45*, 1236.

(200) Reetz, M. T.; Soni, P.; Fernandez, L.; Gumulya, Y.; Carballeira, J. D. *Chem. Commun.* **2010**, *46*, 8657.

(201) Bhuiya, M.-W.; Liu, C.-J. *J. Biol. Chem.* **2010**, *285*, 277.

(202) Bosshart, A.; Panke, S.; Bechtold, M. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 9673.

(203) Blum, J. K.; Ricketts, M. D.; Bommarius, A. S. *J. Biotechnol.* **2012**, *160*, 214.

(204) Engstro, K.; Nyhle, J.; Sandstro, A. G. *J. Am. Chem. Soc.* **2010**, *132*, 7038.

(205) Naqvi, T.; Warden, A. C.; French, N.; Sugrue, E.; Carr, P. D.; Jackson, C. J.; Scott, C. *PLoS One* **2014**, *9*, e94177.

(206) Wang, C.; Huang, R.; He, B.; Du, Q. *BMC Bioinformatics* **2012**, *13*, 263.

(207) Herman, A.; Tawfik, D. S. *Protein Eng. Des. Sel.* **2007**, *20*, 219.

(208) Ness, J. E.; Kim, S.; Gottman, A.; Pak, R.; Krebber, A.; Borchert, T. V; Govindarajan, S.; Mundorff, E. C.; Minshull, J. *Nat. Biotechnol.* **2002**, *20*, 1251.

(209) Gibbs, M. D.; Nevalainen, K. M. H.; Bergquist, P. L. *Gene* **2001**, *271*, 13.

(210) Zha, D. X.; Eipper, A.; Reetz, M. T. *ChemBioChem* **2003**, *4*, 34.

(211) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. *Protein Sci.* **2006**, *15*, 2785.

(212) Kuhlman, B.; Baker, D. *Biophysics.* **2000**, *97*, 10383.

(213) Dahiyat, B. I.; Mayo, S. L. *Science.* **1997**, *278*, 82.

(214) Pantazes, R. J.; Grisewood, M. J.; Li, T.; Gifford, N. P.; Maranas, C. D. *J. Comput. Chem.* **2015**, *36*, 251.

(215) Kries, H.; Blomberg, R.; Hilvert, D. *Curr. Opin. Chem. Biol.* **2013**, *17*, 221.

(216) Damborsky, J.; Brezovsky, J. *Curr. Opin. Chem. Biol.* **2014**, *19*, 8.

(217) Zanghellini, A. *Curr. Opin. Biotechnol.* **2014**, *29*, 132.

(218) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Iii, C. F. B.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science.* **2008**, *319*, 1387.

(219) Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190.

(220) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L. St.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. *Science.* **2010**, *329*, 309.

(221) Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; Baker, D. *Nature* **2013**, *501*, 212.

(222) Gordon, S. R.; Stanley, E. J.; Wolf, S.; Toland, A.; Wu, S. J.; Hadidi, D.; Mills, J. H.; Baker, D.; Pultz, I. S.; Siegel, J. B. *J. Am. Soc.* **2012**, *134*, 20513.

(223) Lippow, S. M.; Moon, T. S.; Basu, S.; Yoon, S.-H.; Li, X.; Chapman, B. a; Robison, K.; Lipovšek, D.; Prather, K. L. J. *Chem. Biol.* **2010**, *17*, 1306.

(224) Boersma, Y. L.; Droge, M. J.; Quax, W. J. *FEBS J.* **2007**, *274*, 2181.

(225) Dietrich, J. a; McKee, A. E.; Keasling, J. D. *High-throughput metabolic engineering: advances in small-molecule screening and selection.*; **2010**; Vol. 79, pp. 563–590.

(226) Fiet, S. V; van Beilen, J. B.; Witholt, B. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1693.

(227) Parikh, M. R.; Greene, D. N.; Woods, K. K.; Matsumura, I. *Protein Eng. Des. Sel.* **2006**, *19*, 113.

(228) Ju, J.; Misono, H.; Ohnishi, K. *J. Biosci. Bioeng.* **2005**, *100*, 246.

(229) DeSantis, G.; Liu, J. J.; Clark, D. P.; Heine, A.; Wilson, I. A.; Wong, C. H. *Bioorg. Med. Chem.* **2003**, *11*, 43.

(230) Griffiths, J. S.; Cheriyan, M.; Corbell, J. B.; Pocivavsek, L.; Fierke, C. A.; Toone, E. J. *Bioorg. Med. Chem.* **2004**, *12*, 4067.

(231) Droge, M. J.; Boersma, Y. L.; van Pouderoyen, G.; Vrenken, T. E.; Ruggeberg, C. J.; Reetz, M. T.; Dijkstra, B. W.; Quax, W. J. *ChemBioChem* **2006**, *7*, 149.

(232) Boersma, Y. L.; Droge, M. J.; Quax, W. J. *FEBS J.* **2007**, *274*, 2181.

(233) Reetz, M. T.; Hoebenreich, H.; Soni, P.; Fernandez, L. *Chem. Commun.* **2008**, *53*, 5502.

(234) Leemhuis, H.; Kelly, R. M.; Dijkhuizen, L. *IUBMB Life* **2009**, *61*, 222.

(235) Timmis, K.; Carreno, C. A.; de Lorenzo, V. In *In Handbook of Hydrocarbon and Lipid Microbiology*; 2010; pp. 4563–4579.

(236) Champion, E.; Moulis, C.; Morel, S.; Mulard, L. A.; Monsan, P. *ChemCatChem* **2010**, *2*, 969.

(237) Ingham, C. J.; Sprenkels, A.; Bomer, J.; Molenaar, D.; Berg, A. Van Den; Vlieg, J. E. T. V. H.; Vos, W. M. De. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18217.

(238) Reetz, M. T.; Waldvogel, S. R. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 865.

(239) Reetz, M. T.; Becker, M. H.; Klein, H. W.; Stockigt, D. *Angew. Chem. Int. Ed.* **1999**, *38*, 1758.

(240) Reetz, M. T.; Hermes, M.; Becker, M. H. *Appl. Microbiol. Biotechnol.* **2001**, *55*, 531.

(241) Reetz, M. T.; Becker, M. H.; Liebl, M.; Furstner, A. *Angew. Chem. Int. Ed.* **2000**, *39*, 1236.

(242) Reetz, M. T.; Kuhling, K. M.; Deege, A.; Hinrichs, H.; Belder, D. *Angew. Chem. Int. Ed.* **2000**, *39*, 3891.

(243) Reetz, M. T.; Kuhling, K. M.; Hinrichs, H.; Deege, A. *Chirality* **2000**, *12*, 479.

(244) Reetz, M. T.; Kuhling, K. M.; Wilensek, S.; Husmann, H.; Hausig, U. W.; Hermes, M. *Catal. Today* **2001**, *67*, 389.

(245) Reetz, M. T.; Eipper, A.; Tielmann, P.; Mynott, R. *Adv. Synth. Catal.* **2002**, *344*, 1008.

(246) Kintses, B.; Hein, C.; Mohamed, M. F.; Fischlechner, M.; Courtois, F.; Lainé, C.; Hollfelder, F. *Chem. Biol.* **2012**, *19*, 1001.

(247) Griffiths, A. D.; Tawfik, D. S. *Trends Biotechnol.* **2006**, *24*, 395.

(248) Goddard, J. P.; Reymond, J. L. *Curr. Opin. Biotechnol.* **2004**, *15*, 314.

(249) Bergquist, P. L.; Hardiman, E. M.; Ferrari, B. C.; Winsley, T. *Extremophiles* **2009**, *13*, 389.

(250) Gredell, J. A.; Frei, C. S.; Cirino, P. C. *Biotechnol. J.* **2012**, *7*, 477.

(251) Desai, S. K.; Gallivan, J. P. *J. Am. Chem. Soc.* **2004**, *126*, 13247.

(252) Miller, O. J.; Bernath, K.; Agresti, J. J.; Amitai, G.; Kelly, B. T.; Mastrobattista, E.; Taly, V.; Magdassi, S.; Tawfik, D. S.; Griffiths, A. D. *Nat. Methods* **2006**, *3*, 561.

(253) Yang, G.; Withers, S. G. *ChemBioChem* **2009**, *10*, 2704.

(254) Schaerli, Y.; Hollfelder, F. *Mol. Biosyst.* **2009**, *5*, 1392.

(255) Tawfik, D. S.; Griffiths, A. D. *Nat. Biotechnol.* **1998**, *16*, 652.

(256) Theberge, A. B.; Courtois, F.; Schaerli, Y.; Fischlechner, M.; Abell, C.; Hollfelder, F.; Huck, W. T. S. *Angew. Chem. Int. Ed.* **2010**, *49*, 5846.

(257) Kintses, B.; van Vliet, L. D.; Devenish, S. R. a; Hollfelder, F. *Curr. Opin. Chem. Biol.* **2010**, *14*, 548.

(258) Skov, L. K.; Mirza, O.; Sprogøe, D.; Van der Veen, B.; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *Biocatal. Biotransformation* **2006**, *24*, 99.

(259) Albenne, C.; Skov, L. K.; Mirza, O.; Gajhede, M.; Feller, G.; D'Amico, S.; André, G.; Potocki-Véronèse, G.; Van der Veen, B.; Monsan, P.; Remaud-Siméon, M. *J. Biol. Chem.* **2004**, *279*, 726.

(260) Schneider, J.; Fricke, C.; Overwin, H.; Hofmann, B.; Hofer, B. *Appl. Environ. Microbiol.* **2009**, *75*, 7453.

(261) Van der Veen, B.; Skov, L. K.; Potocki-Veronese, G.; Gajhede, M.; Monsan, P.; Remaud-Siméon, M. *FEBS J.* **2006**, *273*, 673.

(262) Emond, S.; André, I.; Jaziri, K. *Protein Sci.* **2008**, *17*, 967.

(263) Thèse de doctorat, INSA de Toulouse, Daudé, D.; Remaud-Siméon, M.; André, I. Molecular Evolutionary Perspectives of Amylosucrases: From Substrate Promiscuity to Tailored Catalysis, **2013**.

(264) Jennison, A. V; Verma, N. K. *FEMS Microbiol. Rev.* **2004**, *28*, 43.

(265) Lindberg, A. A.; Karnell, A.; Weintraub, A. *Rev. Infect. Dis.* **1991**, *13*, S279.

(266) Champion, E.; Guérin, F.; Moulis, C.; Barbe, S.; Tran, T. H.; Morel, S.; Descroix, K.; Monsan, P.; Mourey, L.; Mulard, L. A.; Tranier, S.; Remaud-Siméon, M.; André, I. *J. Am. Chem. Soc.* **2012**, *134*, 18677.

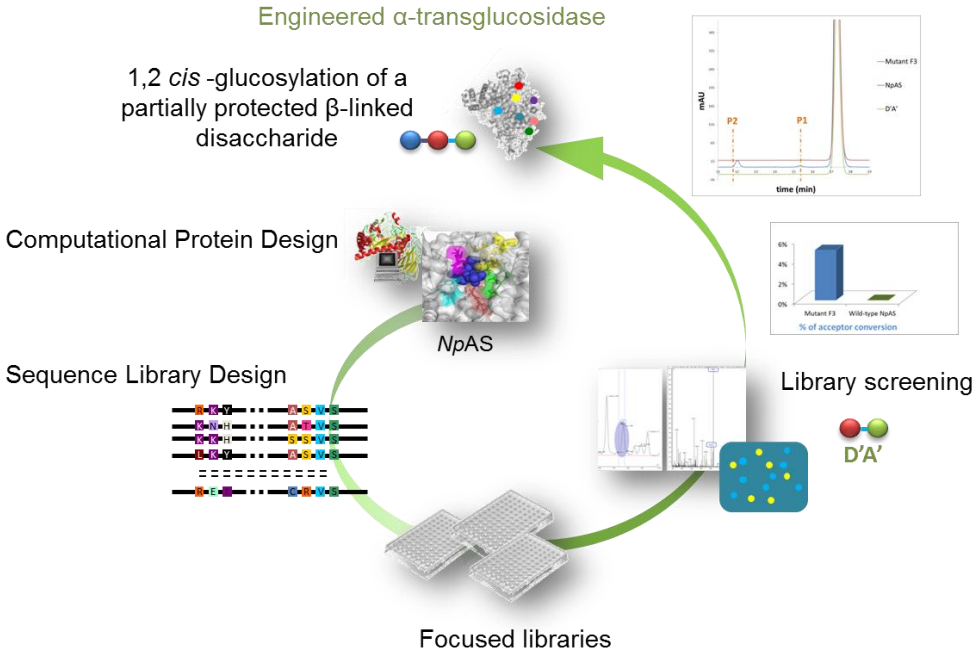
(267) Daudé, D.; Topham, C. M.; Remaud-Siméon, M.; André, I. *Protein Sci.* **2013**, *22*, 1754.

(268) Teze, D.; Hendrickx, J. ; Czjzek M., ; Ropartz, D., ; Sanejouand, Y.H. ; Vinh Tran, V. ; Tellier, C. ; Dion M. *Prot. Eng. Des. Sel.* **2014**, *27*, 13.

Chapter II:

Computer-aided engineering of a transglycosylase for the glucosylation of an unnatural disaccharide of relevance for bacterial antigen synthesis.

ACS catal., (2015), *5* (2), pp 1186–1198



The first and main objective of this work was to engineer an α-transglucosylase that would be able to glucosylate a partially protected β-linked disaccharide, allyl (2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside, a potential intermediate in the synthesis of *Shigella flexneri* cell-surface oligosaccharides. The ultimate goal of this work is to integrate this enzyme into chemo-enzymatic pathway to circumvent limitation of the chemical 1,2-*cis* glucosylation and to synthesize oligosaccharides that will compose vaccines against Shigellosis. Given the challenge to face, a computer-aided approach was undertaken to reshape the multiple sugar binding subsites of our targeted enzyme, the amylosucrase from *Neisseria polysaccharea*, in order to enable productive binding of the non-natural acceptor. A semi-rational library of $2.7x10^4$ sequences was designed and constructed using degenerate oligonucleotides encoding only for designed amino acids and overlap PCRs. Upon multi-level screenings, only one mutant reached the targeted goal of glucosylating the non-natural disaccharide without losing its specificity for sucrose donor. Upon characterization, this mutant revealed a total of 7 mutations distributed in the active site first shell and the putative molecular determinants responsible for the new acceptor specificity were discussed on the basis of a molecular modelling study. The associated results are detailed in an article published in the *ACS Catalysis* journal and are presented hereafter.

INTRODUCTION

Enzyme engineering is a major instrument of modern biocatalysis and novel enzyme-based process development.[1,2] Advances in bioinformatics, biophysics, biochemistry and screening instrumentation provided all their contribution to rationalize or accelerate the optimization, adaptation, and even the conception of enzymes.[3] Learning from natural evolution process allowed decrypting some of the natural mechanisms that shape protein fitness landscapes and serve to develop laboratory methods of evolution and tailor catalysts for synthetic purposes.[4] In particular, computer-aided approaches including sequence alignments combined or not with structural data, phylogenetic analyses, identification of mutation correlations within protein superfamilies occupy an increasing place in the rationalization of enzyme evolution to design focused libraries, guide amino acid substitutions and narrow down the size of the library.[5] Also, Computational Protein Design (CPD) is developing quickly[6–9] and despite still existing limitations in the treatment of flexibility, sequence-conformation exploration, solvent consideration and energy functions, it provides extremely useful information to create libraries of defined-sequences.[10–17] An increasing number of examples demonstrate their potential to draw evolutionary paths leading to desired catalytic functions with a good efficacy and accuracy. However, these computer-aided approaches have been scarcely applied to carbohydrate-active enzymes (CAZymes) although they could significantly accelerate the exploration of the combinatorial mutations, especially when the engineering of multiple carbohydrate binding sites is necessary to confer the desired substrate specificity to the enzyme.[18]

Although the number of natural CAZymes able to act in a regio- or stereo-specific manner on a vast range of glycosidic molecules is continuously expanding in gene databases, access to catalysts that will address a particular need is not a trivial issue. In particular, the lack of appropriate enzymatic tools with requisite substrate and bond specificities has hampered extensive exploration of enzymatic or chemo-enzymatic routes to complex carbohydrates and the design of biocatalysts, which in combination with chemistry could help bypassing some constraints encountered in glycochemistry, such as the need of protection of highly reactive hydroxyl groups. Such constraints can be integrated into the (re)design of enzymes, purposely 'tailored' to enter the chemo-enzymatic pathway at a specific and programmed stage and thus enable thorough exploration of novel pathways toward defined carbohydrate-based structures.[19,20]

This is the challenge addressed in this study where a computer-aided approach was followed to re-design several sugar binding subsites of a sucrose-utilizing transglucosylase, the amylosucrase from *Neisseria polysaccharea* (*Np*AS). The objective was to adapt the enzyme to the transglucosylation of a non-natural disaccharide, structurally highly divergent from its natural substrates, with the long term goal of developing *in vitro* chemo-enzymatic pathways to access complex oligosaccharides featuring α-D-glucopyranosyl side chains. Indeed, α-D-glucosylation is frequently encountered in natural polysaccharides either in the backbone or as side-chain substitutions. The latter situation is nicely exemplified in *Shigella flexneri*, a family of Gram negative bacteria. Most of the specific polysaccharides (O-SP) of the different *S. flexneri* types and subtypes show repeating units with the same linear tetrasaccharide backbone, and differing mainly in the site-specific α-D-glucosylation and O-acetylation of this common

tetrasaccharide.[21,22] Recently, we demonstrated that active site of the α-retaining *Np*AS could be reshaped to enable the site-selective α-D-glucosylation of two non-natural acceptor monosaccharides, namely methyl α-L-rhamnoside and allyl 2-acetamido-2-deoxy-α-D-glucopyranoside.[19,20] A more ambitious design relying on the re-engineering of several binding subsites of *Np*AS was undertaken here to render the enzyme able to glucosylate disaccharide D'A', featuring a partially protected D-glucosamine (D') β-(1,2)-linked to a partially protected L-rhamnoside (A') (Figure 1, Figure S1 of the Supporting Information). The unnatural disaccharide acceptor (D'A') was designed so as to open the way to a set of serotype-specific *S. flexneri* O-SP motifs after appropriate enzymatic α-D-glucosylation (Figure S1 of the Supporting Information). In the following, emphasis was put on the conversion of the D'A' acceptor into the ED'A' trisaccharide, featuring an α-D-glucopyranosyl residue (E) at O-$4_{D'}$ (Figure 1). This glucosylation pattern is present in the O-SP from *S. flexneri* serotypes 1a, 1b and 1d.[21,22] The CPD approach was first used for the generation of a primary sequence library, which was then rationally reduced in size by integration of additional information regarding conserved amino acids or correlated pairs of amino-acids found in homologous enzymes. Reasons for adopting such a tactic were motivated by the objectives addressed in this work: (1) conserving the original specificity toward donor substrate, while (2) creating a novel specificity for a complex acceptor substrate not recognized by the parental wild-type enzyme, (3) requiring thus redesign at multiple binding subsites of a very constrained active site, and (4) need of downsizing mutant libraries due to the availability of only low- and medium-throughput screening assays. Using the framework disclosed herein, a library of defined amino-acid sequences encoding only for the desired combinations was constructed and screened. An α-transglycosylase, whose natural specificity is the synthesis of α-(1,4)-linked glucopyranosyl polymers from sole sucrose,[23–25] was converted into an enzyme displaying a novel substrate specificity for the α-D-glucosylation at OH-$4_{D'}$ of the β-(1,2)-linked disaccharide D'A' for which there is no equivalent reaction yet reported in the literature.

RESULTS AND DISCUSSION

Candidate enzyme and selected disaccharide.

Adding to the fact that a wealth of structural information is already available on *Np*AS enabling the use of computer-aided engineering strategies, this enzyme appeared to be a promising and malleable candidate for the challenge undertaken herein, which is the conception of a mutant able to regiospecifically glucosylate, from sucrose donor, a lightly protected disaccharide acceptor. As seen in Figure 1, the structure of the disaccharide **D'A'** is very different from that of the natural acceptor (maltose). Noteworthy, this molecule is not recognized at all by the wild-type enzyme and its accommodation necessitates major re-design of both +1 and +2 subsites to allow its glucosylation during the second step of the catalytic reaction (Figure 1). In contrast to most chemo-enzymatic syntheses of complex oligosaccharides described to date,[26–29] the strategy under investigation is based on an early-stage enzymatic glucosylation, followed by chemical chain elongation of the obtained product. Toward this aim, it is critical that the products of the enzymatic glucosylation step are compatible with further chemical transformation into building blocks permitting chain elongation at both the reducing and non-reducing ends. We reasoned that the use of lightly protected acceptor substrates should in theory fulfill these criteria without

preventing enzyme remodelling. In the following, disaccharide D'A', namely allyl (2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside (Figure 1), corresponding to a protected version of part of the backbone tetrasaccharide repeating unit of several *S. flexneri* surface polysaccharides (Figure S1 of the Supporting Information),[21,22] was designed as the selected acceptor. Besides being orthogonal to most protecting groups, the allyl moiety does not bring in major steric hindrance. Moreover, it was previously shown to be compatible with *Np*AS remodelling and successful chain elongation of the disaccharide resulting from the enzymatic glucosylation step of a monosaccharide acceptor.[19] The remote effect of acetamides on chemical glycosylation reactions was reported,[30] and the synthesis of oligosaccharides bearing *N*-acetylated aminosugars usually involves building blocks whereby the acetamido function is masked. In this context, we have successfully used the trichloroacetyl moiety as a *N*-protecting group for the aminosugar building blocks involved in the chemical synthesis of oligosaccharides representative of the O-SP from different *S. flexneri* types and subtypes.[31,32] In agreement with our previous achievements, the 2_D acetamide in disaccharide DA is masked as a trichloroacetamide moiety in acceptor D'A'

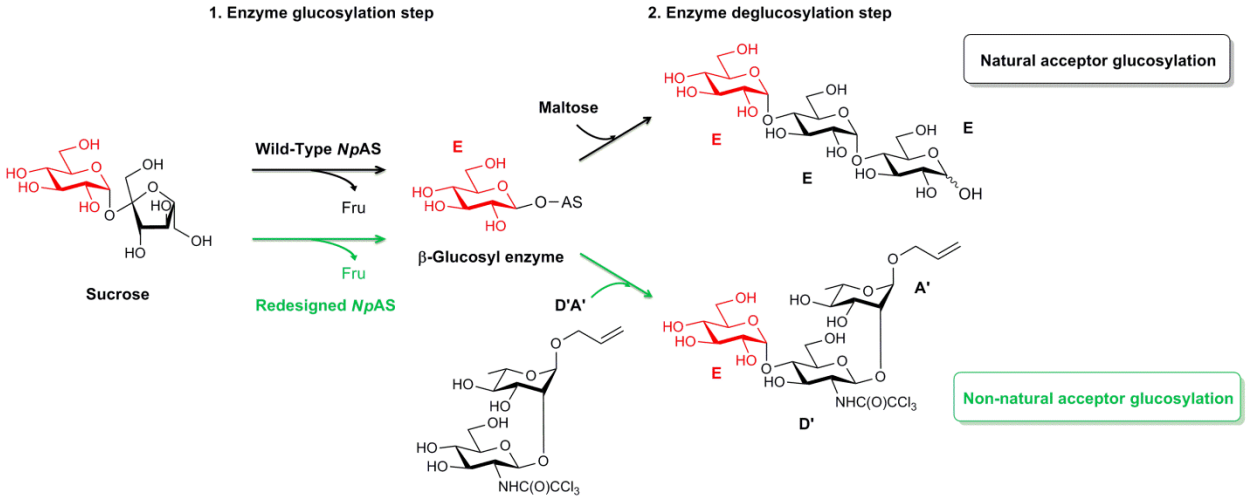


Figure 1. Reaction catalyzed by *Np*AS from sucrose in the presence of natural acceptor (maltose) and targeted reaction by redesigned amylosucrase (in green) using the non-natural acceptor D'A'. *Catalytic mechanism is decomposed in two sequential steps, the first step (enzyme glucosylation step) leads to the formation of a α-glucosyl enzyme intermediate, while the second one (enzyme de-glucosylation step) corresponds to the transglucosylation of the acceptor molecule (natural or non-natural).* Np*AS:* Neisseria polysaccharea *amylosucrase; sucrose, α-D-glucopyranosyl-(1,2)-β-D-fructofuranoside; Fru, Fructose; Maltose, Glucopyranosyl-(1-4)-α-D-glucopyranose;* ***A'****, allyl α-L-rhamnopyranoside;* ***D'****, 2-deoxy-2-trichloroacetamido-α-D-glucopyranosyl;* ***E****, α-D-Glucopyranosyl.*

Design and construction of directed sequence libraries.

Crystallographic structure of *Np*AS in complex with the maltoheptaose product[25] was used as a template to manually force docking of the desired glucosylated product (allyl α-D-glucopyranosyl-(1→4)-(2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside, ED'A') into the enzyme active site. Examination of docked ED'A' showed that all binding interactions with the glucopyranosyl moiety (E) at the -1 subsite, which is responsible for enzyme specificity toward sucrose, were preserved in comparison with crystallographic data on complexes. After excluding all essential amino acid residues for either catalysis or specific recognition of sucrose, we selected altogether 23 mutable positions from the first shell of amino acid residues of subsites +1, +2 and +3, which were likely by mutation to favor recognition of D'A'. These 23 positions are located in structurally distinct regions of the enzyme, mainly on loops 3, 4 and 7 which surround the catalytic center (Figure 2). If all 23 positions were considered to be mutable by one of the 20 possible amino acid residues, the theoretical space would be as large as $8.4 \cdot 10^{29}$, clearly out of reach of available screening methods. Prior methods focusing on the use of degenerate oligonucleotides encoding for a limited set of amino acids (such as NDT or NDK codons encoding for only 12 amino acids) have been proposed to reduce the library size.[33] However, these methods randomly introduce a reduced set of amino acids without taking into account the structural context and the mechanistic constraints. To tackle the customized construction of a small, semi-random, and diversity-controlled library, we proposed to adopt an approach enabling to focus only on the most promising mutations derived from computational predictions and that were refined by additional integration of structural and biochemical knowledge.[23-25,34] Considering the 23 selected positions and using the enzyme design protocol described in Figure 3, we sampled mutations at each position by performing 20,000 independent runs of RosettaDesign. The 20,000 generated sequences were subsequently scored and filtered using various criteria including predicted favorable enzyme-ligand binding interaction and complex stability. This filtering process yielded a subset of 1,515 designed sequences. At this stage, we decided to not directly generate this library of mutants. Indeed, in some cases, the predicted structures differed only by very small scoring differences that might have been irrelevant given all approximations made in the calculations. On the other hand, the combinatorial space of this library remains huge, around of $5.5\ 10^{11}$ sequences and still too broad to be handled experimentally. The reduction of library size was thus pursued on the basis of a specific sequence analysis consisting in determining the amino acid variability within the 1,515 designed sequences and their conservation in the evolutionary related enzymes to sub-family GH13_4 of glycoside-hydrolases[34], which share a common specificity toward sucrose substrate donor.

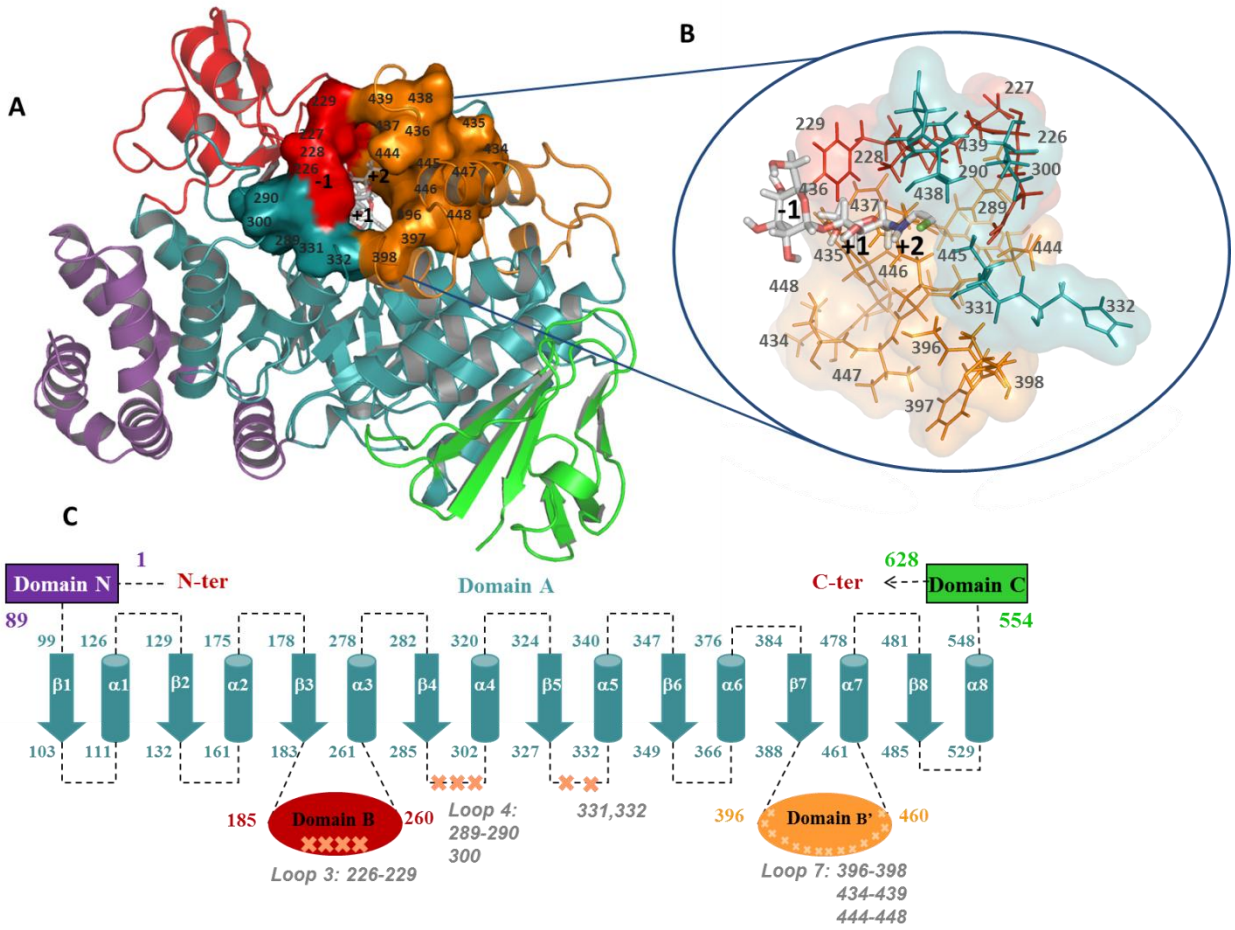


Figure 2. View of amylosucrase organization. Constitutive domains of *Np*AS are colored respectively in purple for the N-domain, in cyan for the A-domain (α/β) (catalytic barrel), in red for the B-domain (loop 3 of catalytic barrel), in orange for the B' domain (extension of the loop 7 from catalytic barrel), and green for the C-domain. *A: overall view of* Np*AS in complex with docked ED'A' showing the 23 selected positions forming the binding site; B: cross section view of the active site with ED'A' occupying subsites -1, +1 and +2; C: Schematic representation of the amylosucrase topology and location of the 23 selected position.*

On the basis of the amino acid occurrence observed for each position in the 1,515 designs, mutations that had low occurrence (≤10%) were discarded. Whenever the presence of wild-type amino acids was found over or equal to 5%, it was maintained at the given position in the final designed library. Although at some positions the wild-type amino acid was not (or only scarcely) observed in the design results (frequency of occurrence less than 5%), it was kept or re-introduced in the final library at this position when it was found highly conserved in the GH13_4 sub-family (Shannon entropy less than 0.2%). The decision to only re-introduce wild-type amino acids at very specific positions is a compromise considering that a high occurence in the sub-family GH13_4 may be critical for sucrose substrate specificity but that the preservation of wild-type signature at all the positions would have led to a too large library. Indeed, if the wild-

type amino acid had been conserved at each designable position, the size of the library would have been as large as ca. 10^7 compared to $1.2 \cdot 10^5$ using a position-dependent re-introduction of wild-type amino acid. To further reduce the size of the combinatorial library, pairwise correlation analysis of predicted mutations was carried out. On this basis, only significant inter-position correlations (residue-pair frequency $\geq$ 5%) were kept in order to only promote beneficial combinations of mutations. Such removal of correlated pair combinations displaying low residue-pair frequency allowed narrowing down further the combinatorial space size to $2.7 \cdot x10^4$ sequences. The expected outcome of this approach was to achieve a sharper exploration of the diversity in the regions of higher interest in comparison to the exorbitant size of theoretical combinatorial libraries generated if all 20 amino acid mutations had been allowed at each of the 23 positions (corresponding in theory to 20^{23} possible sequences). To do so, diversity of the library was controlled in the library construction by mixing synthetic oligonucleotides that encoded for most frequent amino acid mutations and taking into account favorable inter-residue correlations observed in sequences predicted by computational design. Considering the 23 selected positions from the first shell and using a computational approach involving the different filters above described (Figure 3), the diversity was drastically narrowed down from 20^{23} theoretical possible sequences to $2.7 \cdot x10^4$. To ensure the exploration of all the designed sequences, three sub-libraries were constructed (Table 1). In the designed libraries, the amino acid sequences were expected to contain from 9 up to 17 mutations avoiding any representation of the parental gene. To construct the libraries, the gene was first amplified using 6 couples of non-mutated primers positioned all along the sequence to generate 6 DNA wild-type fragments surrounding the 5 regions containing designed mutations (Step 1 in Figure S2 of the Supporting Information). These fragments were then purified and used as primers in subsequent PCR reactions in which both wild-type and mutated primers were used to amplify longer fragments incorporating the mutations (Step 2 in Figure S2 of the Supporting Information) with an equimolar representation of each desired mutation. The amplicons obtained from step 2 were mixed two by two (Step 3 in Figure S2 of the Supporting Information). At step 4, the five amplified long fragments incorporating the mutations were mixed together in equimolar amounts. The entire gene was finally amplified.

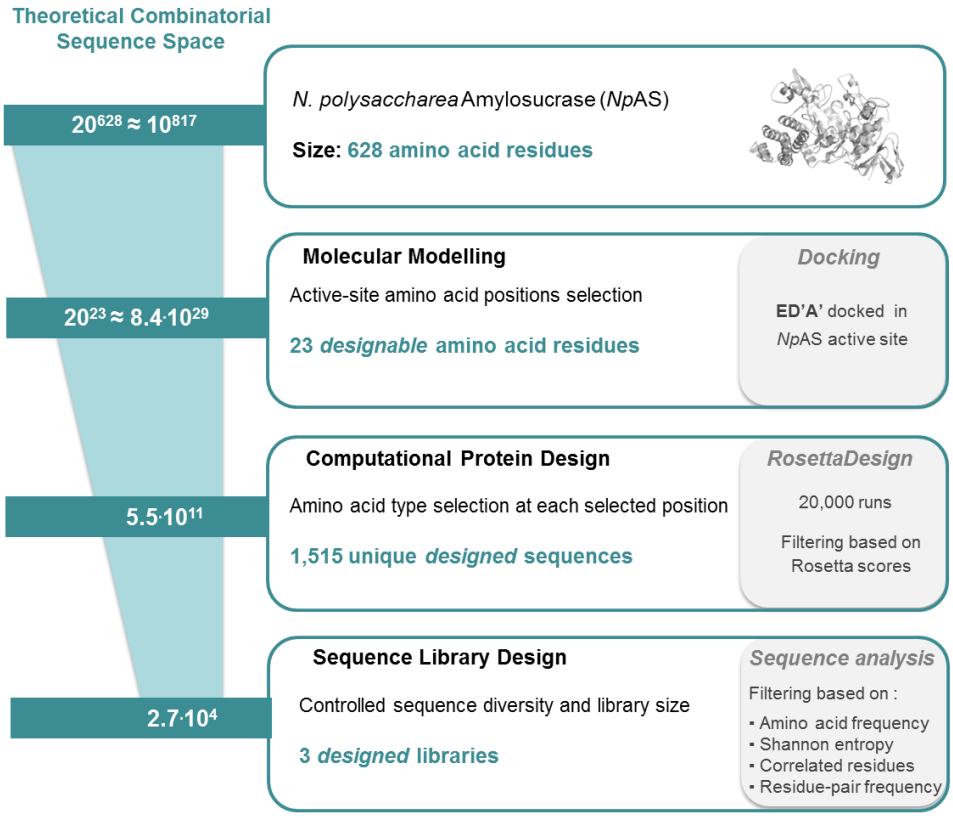


Figure 3. Computer-aided approach for enzyme and sequence library design. The size of the theoretical combinatorial sequence space is indicated at each stage of the strategy.

Table 1. Summary of the predicted mutations and authorized combinations of mutations at the 23 selected positions in the construction of sequence libraries 1-3.

Targeted positions	226	227	228	229	289	290	300	331	332	396	397	398	434	435	436	437	438	439	444	445	446	447	448
Wild-type amino acid	R	E	I	F	A	F	E	V	H	G	W	T	V	P	F	Q	Y	N	D	C	R	V	S
Authorized amino acid Library 1	K or L	K or E	H or Y	D or A or F	S	H	I or V	G or T	H	G or S	W	V	R or V	P	D or S or F	R or S	Y	D	D or E	A	A or R or S	V	S
Authorized amino acid Library 2	K or L	K or E	V or Y	D or A or F	A	Y	V	G or T	H	G or S	W	V	R or V	P	D or S or F	R or S	Y	D	D or E	A	A or R or S	V	S
Authorized amino acid Library 3	K or L	K or E	V or Y	D or A or F	I	Y	I	G or T	H	G or S	W	V	R or V	P	D or S or F	R or S	Y	D	D or E	A	A or R or S	V	S
Targeted structural domain	Domain B (Loop 3)				Domain A (Loop 4)			Domain A (Loop 5)		Domain B' (Loop 7)													

Primary screening of the library for utilization of sucrose as donor substrate.

Screening results.

The libraries were first screened on sucrose using a colorimetric pH-based assay developed earlier on solid medium,[19,20,35] which enables to detect the release of acids upon fructose uptake and decrease the number of mutants to be subsequently screened for D'A' glucosylation. Out of the 63,000 screened clones, the number of active clones represented only 0.09% of the overall library, showing the drastic impact of some mutations onto proper catalytic recognition of sucrose or enzyme viability. Altogether, 59 sucrose-active variants were identified. Of them, 55 variants were confirmed by DNS assays (Table 2).

Table 2. Experimental screening of Libraries 1, 2 and 3 on sucrose and analysis of active clones.

	Library		
	1	2	3
Theoretical number of clones	13,824	6,912	6,912
Number of screened clones	~ 23,000	~ 23,000	~ 17,000
Number of active clones on sucrose using pH-based solid medium assay	5	24	30
Number of active clones on sucrose using the DNS assay	1	24	30
% of overall hit rate	$4.3x10^{-3}$	$1.0x10^{-3}$	$1.7x10^{-3}$
Average number of mutations in active clones	13*	7.8	9.5
Minimum number of mutations in active clones	-	5	5
Maximum number of mutations in active clones	-	13	17

* One single mutant was isolated from Library 1.

Analysis of these 55 variants revealed that about 25% and 70 % of active clones from Library 2 and 3 are more active on sucrose than wild-type ones (Figure S3 of the Supporting Information). The only mutant identified from Library 1 is less active than the wild-type *Np*AS. Overall, 28 clones originating mostly from Library 3 were found with equal or higher activity on sucrose cleavage (see Figure S3 of the Supporting Information). To discriminate transglucosylases from sucrose-hydrolases, the reaction products were analyzed by HPLC. Overall, 25 of the 55 variants (45.5%) synthesized sucrose isomers and malto-oligosaccharides. 13 mutants synthesized products differing from those produced by parental wild-type *Np*AS. 17 mutants (30.9%) behaved as pure sucrose-hydrolases. The thermal stability of active mutants purified in Deep-wells was assessed by Differential Scanning Fluorimetry (DSF) in the absence of a bound ligand using the procedure previously reported[20] (Figure S4 of the Supporting Information). Overall, *T*m values of the 55 mutants ranged from 43.7±0.2°C to 51.5±0.1°C. Noteworthy, most mutants resulting from Library 2 and the mutant from

library 1 turned out to be more stable than wild-type *Np*AS (Tm of 48.6°C), although in overall, these mutants were less active on sucrose. Conversely, mutants from library 3 had lower Tm values than wild-type *Np*AS but were however more active on sucrose. Altogether these results seem to indicate that gain in mutant stability is often associated with a loss of activity on sucrose as previously observed for F290K, A289P-F290L, A289P-F290C and A289P-F290I mutants of *Np*AS.[19,20]

Conformity of constructed libraries with computational predictions.

The 55 selected clones were sequenced (Table S2 of the Supporting Information). On average, clones were found to contain 8.8 mutations per clone, with each clone displaying a different mutational rate. Among all libraries, library 3 yielded a higher number of amino acid mutations per clone (Table 2). 20% and 16.7% of sequences from libraries 2 and 3, respectively, are totally conformed to the design. The maximum number of errors in the designed sequences is 5 and they were introduced most of the time at the same positions (331, 398, 437, 439 or 445). About 73% and 76% of the encoded amino acid sequences from libraries 2 and 3, respectively, contained at the most 3 errors compared to the design. The variation at each position is reported in Figure 4. Amino acid mismatches correspond to wild-type amino acid incorporations or substitutions not allowed by the design. Indeed, several sequences from libraries 2 and 3 contained undesired wild-type amino acids at positions 228, 290, 300, 331, 437, 439 and 445 with a particularly higher frequency at positions 437, 439 and 445. These discrepancies may be attributed to the length of the oligonucleotide primers that could affect the proper hybridation of the fragments (case of the region comprised between positions 434 and 448). A higher error rate is also observed when mutations are located at primer extremities (case of mismatching found at position 398, library 3) (Figure 4).

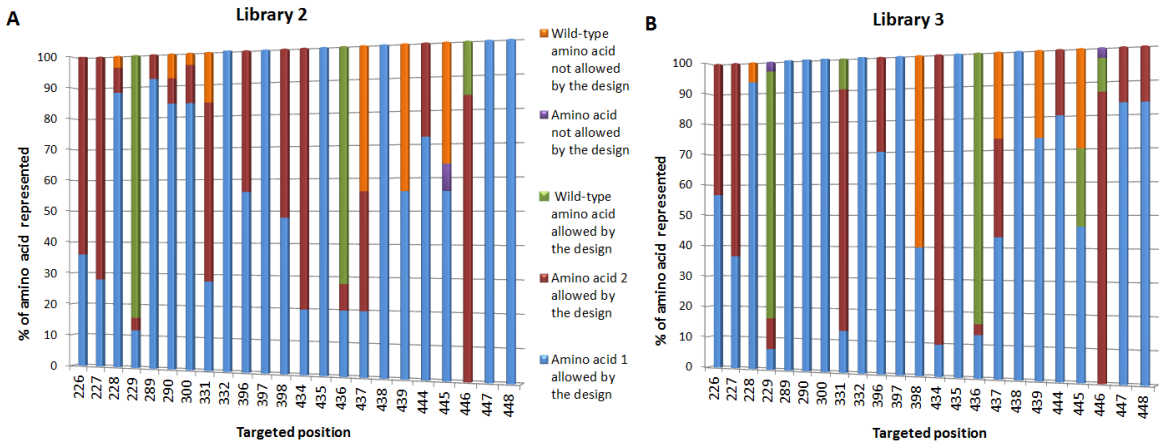


Figure 4. Amino acid composition by position of sucrose-active hits resulting from screening of designed libraries: library 2 (A), library 3 (B).

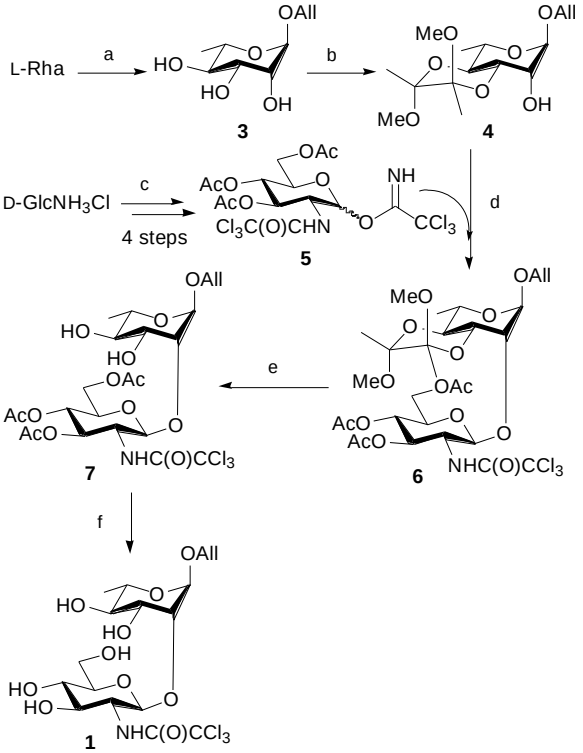
Site-selective α-D-glucosylation of disaccharide D'A' (1).

Disaccharide synthesis.

Multigram amounts of disaccharide D'A' (**1**) were necessary to allow for library screening and active mutant analysis. Toward this aim, disaccharide **1** was chemically synthesized in 9 steps starting from commercially available L-rhamnose and D-glucosamine (Scheme 1). The use of ester and acetal protecting groups was favored so as to prevent any interference with the allyl aglycon and trichloroacetamide moiety at the final deprotection step. Thus, the known allyl rhamnoside[36] **3** was protected regioselectively at position O-3 and O-4 by taking advantage of the selectivity of the 1,2-diacetal chemistry for 1,2-diequatorial diol systems.[37,38] The preliminary conversion of butane-2,3-dione into the more reactive species 2,2,3,3-tetramethoxybutane[39] was found advantageous, and alcohol **4** was isolated in a good 88% yield upon reaction of triol **3** with the latter. TMSOTf-mediated glycosylation of acceptor **4** with the known trichloroacetimidate glucosaminyl donor[40] **5** yielded the crystallized disaccharide **6** in a non-optimized 83% yield, when run on a multigram scale. Acidic hydrolysis of the acetal protecting group gave diol **7** (86%), and subsequent deacetylation of the latter furnished the target allyl glycoside **1** (92%).

Scheme 1. Chemical synthesis of disaccharide D'A' (1) from L-rhamnose and D-glucosamine hydrochloride.

Reagents and conditions: (a) See Ref 36; (b) 2,2,3,3-Tetramethoxybutane, CSA, MeCN, reflux, 30 min (88%); (c) see Ref 40; (d) TMSOTf, DCM, -15 °C → rt, 40 min (83%); (e) 90% aq TFA, rt, 50 min (86%); (f) MeONa, MeOH, rt, 2 h (92%) ; CSA: Camphorsulfonic acid, TMSOTf: trimethylsilyltrifluoromethanesulfonate, DCM: dichloromethane, rt: room temperature; TFA: trifluoroacetic acid.



Secondary screening of mutants on D'A' glucosylation and identification of hits.

All 55 sucrose-active mutants and wild-type *Np*AS were tested for their ability to glucosylate disaccharide D'A'. One mutant (F3) glucosylated D'A' with a conversion of 2 %. LC-MS analysis of the enzymatic mixtures indicated a molecular weight increase by 162 $g.mol^{-1}$ (P1) and 324 $g.mol^{-1}$ (P2) for the novel UV-detectable compounds (P1: 670 and P2: 832 $g.mol^{-1}$, respectively),

corresponding to the transfer of one or two glucosyl units onto D'A' (Figure 5).

Characterization of the mutant displaying a novel D'A' specificity and synthesized products.

Sequencing of the mutant F3 that displays a totally novel substrate specificity toward D'A', not detected for the wild-type *Np*AS, revealed the presence of 7 mutations (R226L-I228V-F290Y-E300V-V331T-G396S-T398V). This mutant has a Tm of 45.5°C instead of 48.6°C for the wild-type enzyme. Reactions in the presence of sole sucrose showed that the specific activity on sucrose of the mutant was 1.8-fold higher than that of the wild-type enzyme. As seen in Figure 6A, the mutant F3 produces a higher amount of turanose (34.2% versus 25.3% for wild-type), more maltose (14.9% vs. 6.1% for wild-type) and also shorter oligosaccharides from DP 3 to 7 (29.6% vs. 3.1% for wild-type) compared to the wild-type enzyme. Indeed, no insoluble α-glucans (DP > 25, 20%) were produced. When D'A' (146mM) was added to the reaction medium, the product profile was modified due to the glycosylation of D'A' leading to products P1 and P2, representing 0.6% and 3.7% of the transferred glucopyranosyl residue, respectively. Glucosylation of D'A' was also accompanied by an increase of small size maltooligosaccharides and trehalulose production and a pronounced decrease of turanose synthesis (twice less than that of the wild-type) (Figure 6B).

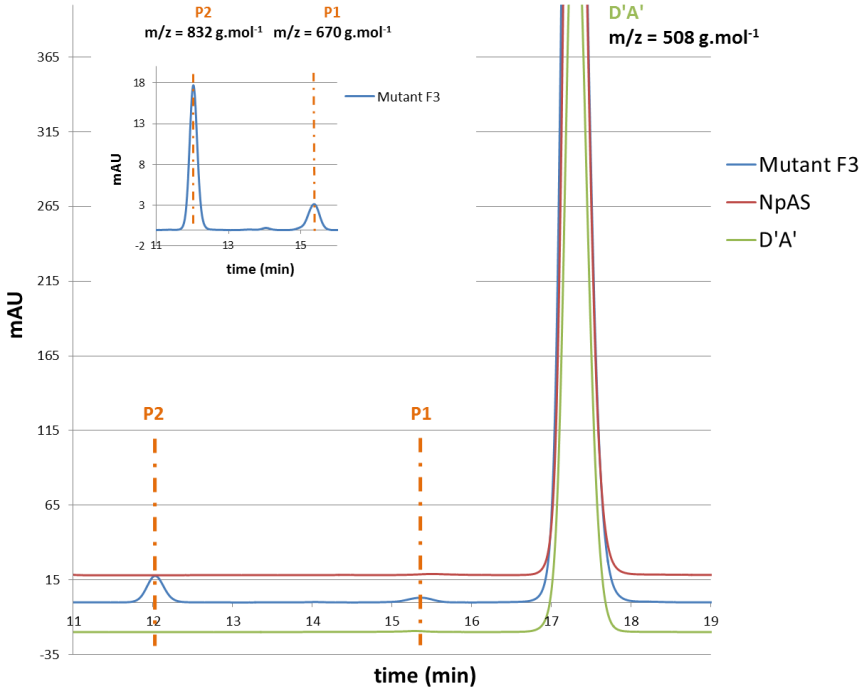


Figure 5. Comparison of product profiles obtained with wild-type *Np*AS (red) and mutant F3 (blue) after 24h using 146mM sucrose and 146mM D'A'. *HPLC analysis was performed (C18 phase column (Venusil AQ 50 mm) with acetonitrile/water (17:83, v/v) as eluent, 30°C; 1mL/min) with UV detection at 205 and 220 nm. The HPLC profile of D'A' (tr = 17.3 min) is shown in green for reference. P1 (tr = 15.3 min) and P2 (tr = 12.1 min): glucosylation reaction products of 670 and 832 g.mol^{-1}, respectively. A zoom of the product P1 and P2 is embedded in the figure.*

Following mass spectrometry analysis, P1 and P2 molecular structures were determined by NMR analysis of tetrasaccharide P2. Data showed that first glucosylation of D'A' occurred at OH-$4_{D'}$ to provide allyl α-D-glucopyranosyl-(1→4)-2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl (1→2)-α-L-rhamnopyranoside, also named ED'A' (or P1 in HPLC analysis). Subsequent chain elongation of the product of glucosylation ED'A' at OH-6_E provided allyl α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→4)-(2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside, also named EED'A' (or P2 in HPLC analysis). The production of ED'A' and EED'A' using various sucrose/D'A' ratio was investigated (Table 3). Increasing amounts of D'A' allows to produce the highest amount of glycosylated acceptors (2.2mM of ED'A', 8.9mM of EED'A' at Sucrose/Acceptor ratio 1/8) in 24 h. The best compromise between the conversion rate and yield of acceptor reaction products is obtained for the Sucrose/Acceptor ratio of 1/1, starting with 146mM sucrose. Whatever the Sucrose/Acceptor ratio used, the production level of the mono-glucosylated acceptor (up to 2.2mM) remained always lower than that of EED'A', suggesting that ED'A' is a better acceptor than disaccharide D'A'.

The steady-state kinetic parameters (k_{cat}, K_m, $k_{cat}/K_{m)}$ of transglucosylation reactions of both the wild-type *Np*AS and the mutant F3 were first determined upon varying sucrose concentration from 5 to 600mM (Table 4). The mutant F3 showed a standard saturation kinetic behavior like the wild-type enzyme. The K_m value was found diminished by 2.3-fold in comparison to the *Np*AS value, indicating a higher affinity of the mutant for sucrose. Conversely, the k_{cat} value of the mutant F3 is decreased by 1.6-fold compared to parental *Np*AS. The catalytic efficiency of the enzymes (k_{cat}/K_m values) was determined by linear regression of the initial rates of transglucosylation measured as variable substrates (sucrose donor or acceptor) (Table 3). In the presence of 200mM D'A' acceptor (considered as a saturating concentration), the catalytic efficiency for sucrose consumption decreased and this appeared to be mainly due to the increase of K_m, suggesting that the presence of disaccharide D'A' affects sucrose binding.

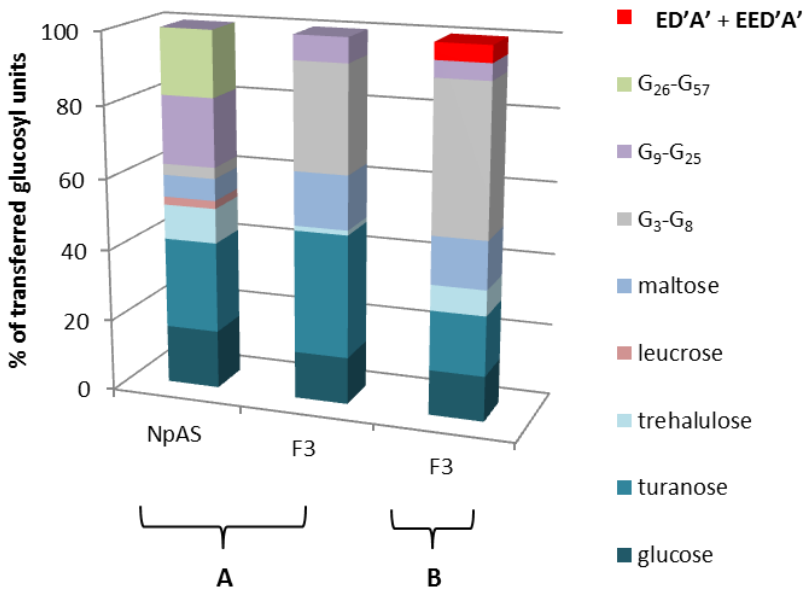


Figure 6. Yields of reaction products analyzed by HPAEC-PAD (Dionex Carbo-Pack PA100 column; 30°C; 1 mL/min; gradient of sodium acetate (from 6 to 300mM in 28 min) in 150mM NaOH) of wild-type *Np*AS and mutant F3 after 24h reaction. ***A:*** *in the presence of sole sucrose (146 mM) or* ***B:*** *in the presence of both sucrose (146mM) and* ***D'A'*** *(146mM). G: Glucose;* ***ED'A'****: allyl α-D-glucopyranosyl-(1→4)-(2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside ;* ***EED'A'****: allyl α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→4)-(2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside.*

Table 3. Yields of D'A' conversion and quantification of mono-glucosylated (P1) and di-glucosylated (P2) products by the mutant F3 using different donor/acceptor molar ratios after 24 h reaction.

	Sucrose / D'A' ratio ([suc] = 73 mM)				Sucrose/ D'A' ratio ([suc] = 146 mM)	Sucrose / D'A' ratio ([D'A'] = 73 mM)		
	1/1	1/2	1/4	1/8	1/1	2/1	4/1	8/1
D'A' conversion (%)[a]	2	2.17	1.7	1.9	3.0	3.0	4.3	5
Mono-glucosylated D'A' (ED'A' or P1)[b]**, mM**	0.3	0.5	1.1	2.2	0.6	0.3	0.3	0.3
Di-glucosylated D'A' (EED'A' or P2)[c]**, mM**	1.2	2.5	3.7	8.9	3.7	1.9	2.8	3.3

[a] conversion of **D'A'** was calculated as follows : [P1]mM + [P2]mM/[**D'A'**]mM *100.
[b] monoglucosylated **D'A'** (**ED'A', P1**) determined from HPLC analysis (C18 phase column (Venusil AQ 50mm) with acetonitrile/water (17:83, v/v)) using calibration curves established with **D'A'.**
[c] diglucosylated **D'A'** (**EED'A', P2**) determined from HPLC analysis (C18 phase column (Venusil AQ 50mm) with acetonitrile/water (17:83, v/v)) using calibration curves established with **D'A'.**

Kinetic parameters were also determined in the presence of sucrose (200mM) and D'A' acceptor (concentrations varying from 5 to 200mM). Due to the limited solubility of D'A', the enzyme could never be saturated, indicating a very low affinity of D'A' for mutant F3 leading to a poor catalytic efficiency value. Although the catalytic efficiency of the mutant is low, we can expect that additional rounds of directed evolution could help to further improve it. Nevertheless, the design, based on molecular modelling predictions, successfully conferred the novel and desired acceptor specificity without substantially affecting the original activity toward sucrose.

Table 4. Apparent kinetic parameters determination of mutant F3 for sucrose donor only (from 5 to 600mM) and in the presence of both sucrose and D'A' (sucrose concentration at 200mM and various D'A' concentrations from 5 to 200mM or 200mM D'A' and sucrose from 5 to 600mM).

Sucrose only (from 5 to 600mM)	*Np*AS	Mutant F3
K_m apparent [a] (mM)	50.2	21.58
k_{cat} [a] (s^{-1})	1.3	0.82
k_{cat}/K_m [a] (s^{-1} mM^{-1})	0.026	0.040
Sucrose (from 5 to 600 mM) + D'A' (200mM)		
$K_{m, suc}$ [b] (mM)	_ [d]	71
$k_{cat, suc}$ [b] (s^{-1})	_ [d]	1.75
$k_{cat}/K_{m, suc}$ [b] (s^{-1} mM^{-1})	_ [d]	0.024
D'A' (from 5 to 200 mM) + Sucrose (200mM)		
$k_{cat}/K_{m, D'A'}$ [c] (s^{-1} mM^{-1})	_ [d]	$3.5\ 10^{-4}$

[a] catalytic parameters (K_m, k_{cat}, k_{cat}/K_m) were determined in the presence of sucrose only from 5 to 600mM.

[b] catalytic parameters (K_m, k_{cat}, $k_{cat}/K_{m, suc}$) were determined by measuring the initial rate of fructose production in the presence of constant acceptor concentration (200mM) and sucrose from 5 to 600mM.

[c] catalytic efficiency $k_{cat}/K_{m, D'A'}$ was determined by measuring the initial rate **of** ED'A' formation from 200mM sucrose (fixed concentration) and 5 to 200mM D'A'. As D'A' was not soluble at concentrations higher than 200mM, no saturation could be observed and only $k_{cat}/K_{m, D'A'}$ was determined from the slope of the linear regression.

[d] not determined due to D'A' non recognition.

Structural insight on novel substrate specificity.

Our efforts to crystallize the mutant F3 were unsuccessful. Therefore, a three-dimensional model of the mutant was first built using as template the X-ray structure of the parental *Np*AS.[23,25,41] Both sucrose and trisaccharide ED'A', the reaction product of interest, were subsequently docked into the active site. Inspection of the models showed that nearly all 7 mutations were carried by loops 3 and 4 which have been shown to be highly flexible by molecular dynamics simulations[19,20] and assumed to play a determinant role in the catalytic mechanism of the enzyme. As a result of the introduction of these mutations, the active site of the enzyme is widened up at subsites +1 and +2 enabling the accommodation of the ED'A', and in particular its *N*-trichloroacetyl and allyl substituents, while the productive binding of sucrose is not disturbed, as confirmed by experiment (Figure 7). Indeed, detailed analysis of interactions established between the mutant F3 and ED'A' revealed that mutation of Glu300 and Arg226, which are involved in a salt bridge interaction in the parental wild-type enzyme, are mutated by smaller aliphatic residues, valine and leucine, respectively, likely enhancing the plasticity of the active site and favoring the productive binding of D'A'. Mutation of Thr398 by a valine residue enables the nesting of the allyl aglycon in a small depression, while mutation of Gly396 by a serine residue brings an additional hydrogen bonding interaction which stabilizes further the allyl group. The mutation V331T also enables to introduce a stabilizing hydrogen bonding interaction with the 2_D. *N*-trichloroacetyl group. Introduction of the Tyr residue at position 290 requires the creation of enough space, provided by the I228V mutation, to accommodate the extra hydroxyl group compared to the Phe residue present in wild-type enzyme. Altogether these 7 mutations enabled a subtle interplay between the structure, the flexibility, the stabilization and the molecular recognition, which cannot yet be predicted altogether by computational methods. As a result, the combination of these mutations enabled the reshaping of the enzyme active site that was beneficial for the glucosylation of D'A' while preserving sucrose recognition. Interestingly, the di-glucosylated molecule, EED'A', was the main glucosylation product. NMR analysis of this molecule revealed the presence of an unusual α-(1,6) linkage that might confer higher conformational flexibility to enable productive binding of ED'A' in subsites +1, +2 and+3. This could also explain to some extent why ED'A' appears to be a better acceptor than D'A'. Of note, similar observations were found in our recent work on the glucosylation of flavonoids.[42] These results also suggest a high plasticity of subsite +3 enabling accommodation of A'. Recent work on the engineering of the active site[43] also provides new directions to explore in order to better control the elongation of gluco-oligosaccharides and in particular, to avoid the formation of unwanted EED'A'.

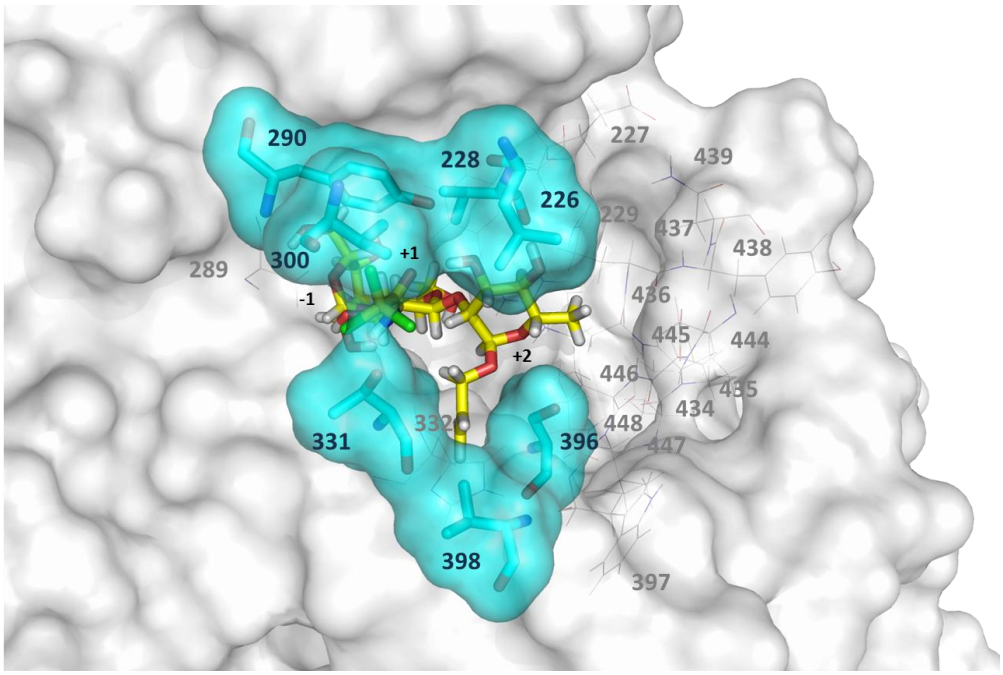


Figure 7. Molecular docking of ED'A' in the active site of a built 3D-model of the mutant F3 displaying novel acceptor specificity toward D'A'. *The 7 mutations (R226L-I228V-F290Y-E300V-V331T-G396S-T398V) are highlighted in cyan color (hydrogen atoms have been omitted for clarity purpose). In grey lines are shown the remaining positions targeted by the computational design but not mutated in the mutant F3.*

CONCLUSIONS

Computer-aided design was for the first time applied to the engineering of several subsites of an α-transglycosylase. The difficulty of this work was to target a limited number of mutations allowing to keep a donor specificity for sucrose and tailor the acceptor specificity to render *Np*AS able to recognize a β-linked and partially protected disaccharide. Noteworthy, *Np*AS has only been described up to now as being able to recognize α-linked disaccharides. The engineering approach included the use of RosettaDesign software, the selection of amino-acid sequences favoring inter-residue correlations predicted by computational design and the experimental control of the desired combinations of mutations. The medium size library was easily screened to yield, after only a single screening run, one mutant displaying the expected substrate specificity, thus showing the efficacy of the approach combining both computer-aided design and construction of sequence libraries. Indeed, out of a library of $2.7x10^4$ variants, the mutant F3 regioselectively glucosylates D'A', what was not detected for the parental wild-type enzyme. A novel transglucosylase, whose substrate specificity was purposely tailored toward a non-natural acceptor, allyl 2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside (D'A'), to yield a potential intermediate in the chemo-enzymatic synthesis of oligosaccharides representative of *S. flexneri* serotypes 1a, 1b or 1d O-SPs. Obviously, although improvements are still necessary, these new findings open the way to alternatives to ongoing developments toward chemically synthesized oligosaccharides representing fragments from the O-SP of various *S. flexneri* type 1 strains.[51,52,53] However, the low hit rate emphasizes the extreme complexity of redesigning amylosucrase, given its double-step mechanism that involves the sequential binding of a donor (sucrose) and an acceptor substrate along the catalytic reaction, and the constrained active site topology which needs to be adapted to D'A' disaccharide which presents considerable structural differences with natural acceptors. Indeed, the novel substrate specificity of the mutant was obtained upon introduction of 7 combined amino acid mutations in the first shell of the active site, enabling a drastic reshaping of the catalytic binding site to accommodate non-natural acceptor D'A' in a productive manner without perturbing too much its original specificity for donor sucrose. Not only these mutations, mainly introduced in flexible loops, enabled a change in substrate specificity but they also enhanced the stability of the enzyme in comparison to the parental wild-type *Np*AS. The catalytic efficiency of the mutant for trisaccharide production was still low compared to that observed for the natural donor sucrose. This is due to the difficulty of this design and can also be attributed to the limits of CPD approaches and experimental difficulties encountered in the construction of the library. Nonetheless, the combined approach used herein was efficient to identify in only one run of screening the evolutionary path toward a totally new and unnatural catalytic activity. Obviously, this opens now the route to further enzyme improvement via either computational methods, directed evolution or mixed strategies.

THEORETICAL AND EXPERIMENTAL PROCEDURE

Computational Modelling and Design.

Detailed procedures are provided in Supplementary Material and Methods. Three-dimensional model of *Np*AS in complex with the desired glucosylated product (ED'A') bound into the enzyme active site was constructed starting from

the crystallographic complex with maltoheptaose, a reaction product (PDB ID: 1MW0).[25] ED'A' was manually docked into the active site of the amylosucrase by superimposing the glucopyranosyl unit (E) of ED'A' onto the glucopyranosyl unit of the crystallographic maltoheptaose located at subsite −1. The molecular all-atom *ff99SB*[44], *Glycam06*[45] and *gaff* force fields[46] were used for the proteins, carbohydrates and other non-standard groups, respectively. The molecular system was then subjected to several cycles of minimizations applying different levels of restraints using the *Sander* module of the Amber 9 software package.[47] Visual inspection of the modelled complex, as well as the crystallographic complex of the amylosucrase with the donor substrate, the sucrose (PDB: 1JGI), was carried out using PyMOL software (Schrödinger, Portland, OR, USA). Based on the analysis of these minimized complexes and our knowledge on amino acid residues involved in catalysis and/or specific recognition of sucrose, 23 positions (226-229, 289-300, 331, 332, 396-398, 434-439, 444-448) of the active site were selected to be sampled for mutations. Version 3.2 of the Rosetta software suite (Rosetta Commons, UW techTransfer, Seattle, WA, USA) was used to redesign the *Np*AS active site in order to enhance the interaction with ED'A'. A protocol was used that tests mutations and side-chain conformations at the 23 selected positions and samples variations in the translation and rotation of the docked ED'A'. All amino acid types, except Cysteine, were authorized at the 23 designable positions. 60 residues surrounding the 23 designable residues were allowed to relax toward alternate side-chain rotamers. Enzyme backbone atoms, non-designed or non-repacked side-chains as well as the internal conformation of ED'A' were held fixed. The initial ED'A' conformation from which perturbations were made was taken from the minimized model of the ED'A'-enzyme complex. The design protocol based on the use of RosettaDesign and described in details in the Supplementary Material and Methods was run 20,000 independent times. The output sequences were filtered based on different Rosetta scoring and filtering schemes. The 1,515 unique sequences resulting from this filtering were used to guide library design. The library design approach developed was based on: (1) the determination from a multiple sequence alignment of 1,515 designed sequences of the position-dependent amino acid residue frequency, a Correlated Mutation score at position pairs and the residue-pair frequency; (2) the calculation of the Shannon entropy using in house software, called Sequester, to measure the position-dependent amino acid residue variability in multiple sequence alignment of GH13_4 sub-family sequences. From these computations, the amino acid types and the combinations of amino acid residues considered in the final designed libraries were selected as follows : (1) at each enzyme designed position, the wild-type amino acid residues with a frequency of occurrence ≥ 5% were maintained; (2) the wild-type amino acid residues with a frequency of occurrence < 5% or absent in the design results and with a Shannon entropy < 0.2% were introduced; (3) the other amino acid types were kept if only their position-dependent frequency was higher than 10% (3) for correlated amino acid pairs, only the combination-pairs with a frequency of occurrence ≥ 5% were preserved while all combinations were considered for non-correlated amino-acid residues. The mutant F3 of *Np*AS was constructed starting from the model of the wild-type enzyme docked with ED'A' in which mutations were introduced using the Biopolymer module of the InsightII software package (Accelrys, San Diego, CA, USA). The model was subsequently minimized following the same energy minimization procedure described above.

Chemical synthesis of disaccharide 1 (D'A').

General conditions and NMR spectra of all compounds are provided in Supplementary Information.

Allyl 3,4-*O*-(2',3'-dimethoxybutane-2',3'-diyl)-α-L-rhamnopyranoside (4).

To a solution of butane-2,3-dione (6.0 mL, 68.4 mmol) in MeOH (25 mL) were added trimethylorthoformate (22.4 mL, 204 mmol, 3.0 equiv.) and 3 drops of concentrated sulfuric acid. The solution was refluxed for 20 h. After complete conversion of the starting dione, as attested by NMR analysis, solid $NaHCO_3$ (5.2 g) was added and the suspension was concentrated to dryness under reduce pressure. The residue was dissolved in Et_2O (150 mL), washed with satd aq. $NaHCO_3$ (50 mL) and brine (50 mL), then dried over sodium sulfate. Volatiles were removed under reduced pressure and the resulting material was distilled under reduce pressure to afford 2,2,3,3-tetramethoxybutane as a colorless oil (bp 35-37°C, 4 mbar). Distilled 2,2,3,3-tetramethoxybutane (4.15 g, 23.3 mmol, 1.2 equiv.) and CSA (209 mg, 0.90 mmol, 0.05 equiv.) in anhydr. MeCN (3 mL) were added to a solution of the known rhamnoside[36] **3** (3.93 g, 19.2 mmol) in anhydr. MeCN (90 mL). The mixture was refluxed for 30 min. Et_3N was added and solvents were removed under reduced pressure. The crude oil was purified by column chromatography (Tol/EtOAc, 80:20 → 60:40) to give alcohol **4** (5.37 g, 88%) as a colorless oil. Alcohol **4** had R_f = 0.49 (cHex/EtOAc 3:2) ; $[\alpha]^{24}_D$ = - 229 ° (*c* 1.0, $CHCl_3$) ; ^{1}H NMR ($CDCl_3$) δ 5.93-5.83 (m, 1H, C*H*=CH_2), 5.29-5.23 (m, J_{trans} = 17.2 Hz, 1H, CH=C*H*$_a$H$_b$), 5.19-5.16 (m, J_{cis} = 10.4 Hz, 1H, CH=CH$_a$*H*$_b$), 4.81 (d, $J_{1,2}$ = 1.2 Hz, 1H, H-1), 4.19-4.13 (m, 1H, -OC*H*$_a$H$_{bAll}$), 4.00-3.95 (m, 2H, -OCH$_a$*H*$_{bAll}$, H-3), 3.92 (dd, $J_{2,3}$ = 3.2 Hz, 1H, H-2), 3.82 (dq, $J_{4,5}$ = 9.8 Hz, 1H, H-5), 3.70 (pt, $J_{3,4}$ = 9.9 Hz, 1H, H-4), 3.26 (s, 3H, -OCH_3), 3.23 (s, 3H, -OCH_3), 2.42 (bs, 1H, OH), 1.31 (s, 3H, CH_3), 1.28 (s, 3H, CH_3), 1.26 (d, $J_{5,6}$ = 6.2 Hz, 3H, H-6). ^{13}C NMR ($CDCl_3$) δ 134.0 (*C*H=CH_2), 117.6 (CH=*C*H_2), 100.3 ($C_{quat.}$), 100.0 ($C_{quat.}$), 99.0 (C-1), 70.0 (C-2), 68.5 (C-4), 68.4 (C-3), 68.1 (-OCH_{2All}), 66.7 (C-5), 48.2 (-OCH_3), 47.8 (-OCH_3), 18.0 (CH_3), 17.8 (CH_3), 17.7 (C-6). HR-ESI-TOF-MS *m/z* 341.1542 $[M + Na]^+$ (Calcd. for $C_{15}H_{26}O_7Na$, 341.1576).

Allyl (3,4,6-tri-*O*-acetyl-2-deoxy-2-trichloroacetamido-α-D-glucopyranosyl)-(1→2)-3,4-*O*-(2',3'-dimethoxybutane-2',3'-diyl)-α-L-rhamnopyranoside (6).

TMSOTf (200 µL, 1.1 mmol, 0.1 equiv.) was added to a solution of acceptor 4 (3.66 g, 11.5 mmol) and of the known donor[40] 5 (7.52 g, 12.6 mmol, 1.1 equiv.) in anhydr. DCM (180 mL) containing 4Å MS (10 g), stirred at -15°C under an Ar atmosphere. After 30 min, Et_3N was added and the suspension was filtered over a pad of Celite. The filtrate was concentrated to dryness. Crystallization of the crude material from EtOAc/Pentane gave disaccharide 6 (7.17 g, 9.54 mmol) as white crystals in a non optimized 83% yield. The fully protected compound 6 had R_f = 0.31 (cHex/EtOAc 3:2) ; $[\alpha]^{24}_D$ = - 92.0 ° (*c* 1.0, $CHCl_3$) ; mp = 186°C (Et_2O/Pentane) ; ^{1}H NMR ($CDCl_3$) δ 6.83 (d, $J_{2,NH}$ = 7.2 Hz, 1H, NH), 5.91-5.82 (m, 1H, C*H*=CH_2), 5.31-5.22 (m, J_{trans} = 17.2 Hz, 2H, CH=C*H*$_2$, H-3$_{D'}$), 5.17-5.13 (m, J_{cis} = 10.4 Hz, 1H, CH=C*H*$_2$), 5.08 (pt, *J* = 9.5 Hz, 1H, H-4$_{D'}$), 4.98 (d, $J_{1,2}$ = 8.2 Hz, $^1J_{CH}$ = 164.4 Hz, 1H, H-1$_{D'}$), 4.86 (d, $J_{1,2}$ = 1.2 Hz, 1H, H-1$_{A'}$), 4.24 (dd, $J_{6a,6b}$ = 12.1 Hz, $J_{5,6a}$ = 5.1 Hz, 1H, H-6a$_{D'}$), 4.18-4.10 (m, 2H, H-6b$_{D'}$, H$_{All}$), 4.08-4.00 (m, 2H, H-2$_{D'}$, H-3$_{A'}$), 3.98-3.92 (m, 2H, H$_{All}$, H-2$_{A'}$), 3.76-3.67 (m, 2H, H-5$_{A'}$, H-5$_{D'}$), 3.56 (pt, *J* = 9.9 Hz, 1H, H-4$_{A'}$), 3.23 (s, 3H, -OCH_3), 3.17 (s, 3H, -OCH_3), 2.08 (s, 3H, CH_{3Ac}), 2.02 (s, 3H, CH_{3Ac}), 2.01 (s, 3H, CH_{3Ac}), 1.25 (s, 3H, CH_3), 1.24 (s, 3H, CH_3), 1.21 (d, $J_{5,6}$ = 6.2 Hz, 3H, H-6$_{A'}$) ;

^{13}C NMR ($CDCl_3$) δ170.8, 170.7, 169.4 (3C, CO_{Ac}), 161.9 (NHCO), 134.0 (*C*H=CH_2), 117.3 (CH=*C*H_2), 100.5 ($C_{quat.}$), 100.2 ($C_{quat.}$), 99.7 (C-$1_{D'}$, $^1J_{C,H}$ = 164.1 Hz), 99.0 (C-$1_{A'}$, $^1J_{C,H}$ = 173.8 Hz), 92.4 (CCl_3), 76.1 (C-$2_{A'}$), 72.2 (2C, C-$3_{D'}$, C-$5_{D'}$), 68.7 (2C, C-$4_{A'}$, C-$4_{D'}$), 68.1 (2C, C-$3_{A'}$, C_{All}), 66.8 (C-$5_{A'}$), 62.2 (C-$6_{D'}$), 56.0 (C-$2_{D'}$), 48.0 (-OCH_3), 47.8 (-OCH_3), 20.8 (CH_{3Ac}), 20.7 (2C, CH_{3Ac}), 18.0 (2C, CH_3), 16.7 (C-$6_{A'}$). HR-ESI-TOF-MS *m/z* 772.1539 (Calcd. for $C_{29}H_{42}Cl_3NO_{15}Na$, 772.1518).

Allyl (3,4,6-tri-*O*-acetyl-2-deoxy-2-trichloroacetamido-α-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside (7).

Disaccharide **6** (1.94 g, 2.58 mmol) was dissolved in 90% aq. TFA (50 mL) and the solution was stirred for 50 min at rt, then repeatedly coevaporated with toluene. The residue was purified by column chromatography (cHex/EtOAc 50:50 → 40:60) to give diol **7** (1.41 g, 2.21 mmol, 86%) as a white solid. Disaccharide **7** had R_f = 0.38 (cHex/EtOAc 2:3) ; $[\alpha]^{24}_D$ = - 24.4 ° (*c* 1.0, $CHCl_3$) ; ^{1}H NMR ($CDCl_3$) δ7.40 (d, $J_{2,NH}$ = 8.4 Hz, 1H, NH), 5.92-5.83 (m, 1H, C*H*=CH_2), 5.29-5.24 (m, J_{trans} = 17.2 Hz, 2H, CH=CH_2, H-$3_{D'}$), 5.19-5.16 (m, J_{cis} = 10.4 Hz, 1H, CH=CH_2), 5.09 (pt, J = 9.8 Hz, 1H, H-$4_{D'}$), 4.94 (d, $J_{1,2}$ = 8.4 Hz, 1H, H-$1_{D'}$), 4.89 (d, $J_{1,2}$ = 1.3 Hz, 1H, H-$1_{A'}$), 4.24 (dd, $J_{6a,6b}$ = 12.4 Hz, $J_{5,6a}$ = 5.0 Hz, 1H, H-$6a_{D'}$), 4.17-4.06 (m, 3H, H-$6b_{D'}$, H_{All}, H-$2_{D'}$), 3.97-3.93 (m, 2H, H_{All}, H-$2_{A'}$), 3.85 (dd, $J_{3,4}$ = 9.5 Hz, $J_{2,3}$ = 3.3 Hz, 1H, H-$3_{A'}$), 3.76-3.72 (m, 1H, H-$5_{D'}$), 3.56-3.36 (dq, $J_{4,5}$ = 9.4 Hz, 1H, H-$5_{A'}$), 3.41 (pt, J = 9.5 Hz, 1H, H-$4_{A'}$), 2.11 (s, 3H, CH_{3Ac}), 2.06 (s, 3H, CH_{3Ac}), 2.05 (s, 3H, CH_{3Ac}), 1.30 (d, $J_{5,6}$ = 6.2 Hz, 3H, H-$6_{A'}$) ; ^{13}C NMR ($CDCl_3$) δ171.0, 170.8, 169.5 (3C, CO_{Ac}), 162.4 (NHCO), 133.9 (*C*H=CH_2), 117.4 (CH=*C*H_2), 102.0 (C-$1_{D'}$, $^1J_{C,H}$ = 163.4 Hz), 98.3 (C-$1_{A'}$, $^1J_{C,H}$ = 173.2 Hz), 92.5 (CCl_3), 79.2 (C-$2_{A'}$), 73.7 (C-$4_{A'}$), 72.3 (2C, C-$3_{D'}$, C-$5_{D'}$), 71.8 (C-$3_{A'}$), 68.5 (C-$4_{D'}$), 68.2 (C-$5_{A'}$), 68.1 (C_{All}), 62.2 (C-$6_{D'}$), 56.2 (C-$2_{D'}$), 20.9, 20.7, 20.6 (3C, CH_{3Ac}), 17.5 (C-$6_{A'}$). HR-ESI-TOF-MS *m/z* 658.0876 (Calcd. for $C_{23}H_{32}Cl_3NO_{13}Na$, 658.0837).

Allyl (2-deoxy-2-trichloroacetamido-α-D-glucopyranosyl)-(1→2)-β-L-rhamnopyranoside (1 or D'A').

To a solution of diol **7** (1.27 g, 2.00 mmol) in anhydr. MeOH (30 mL) stirred under an Ar atmosphere, was added MeONa (25% wt in MeOH, 275 µL, 1.20 mmol, 0.6 equiv.). The reaction mixture was stirred at rt for 2 h then Dowex 50Wx8-200 was added. The suspension was filtered over a pad of Celite. The solvent was removed under vacuum and the residue was purified by column chromatography (DCM/MeOH, 90:10 → 80:20) to give pentaol **1** (937 mg, 1.83 mmol, 92%) as a white solid. Disaccharide **1** had R_f = 0.11 (DCM/MeOH 9:1) ; $[\alpha]^{24}_D$ = - 29.3 ° (*c* 1.0, water) ; ^{1}H NMR (D_2O) δ 6.08–5.98 (m, 1H, C*H*=CH_2), 5.45-5.39 (m, J_{trans} = 17.4 Hz, 1H, CH=CH_2), 5.37-5.33 (m, J_{cis} = 10.4 Hz, 1H, CH=CH_2), 5.06 (d, $J_{1,2}$ = 1.7 Hz, 1H, H-$1_{A'}$), 4.98 (d, $J_{1,2}$ = 8.1 Hz, 1H, H-$1_{D'}$), 4.30-4.24 (m, 1H, H_{All}), 4.16-4.11 (m, 2H, H_{All}, H-$2_{A'}$), 4.00 (bd, $J_{6a,6b}$ = 12.5 Hz, 1H, H-$6a_{D'}$), 3.91 (dd, $J_{3,4}$ = 9.6 Hz, $J_{2,3}$ = 3.1 Hz, 1H, H-$3_{A'}$), 3.88-3.80 (m, 3H, H-$6b_{D'}$, H-$2_{D'}$, H-$3_{D'}$), 3.80-3.72 (m, 1H, H-$5_{A'}$), 3.57-3.55 (m, 2H, H-$4_{D'}$, H-$5_{D'}$), 3.50 (pt, J = 9.6 Hz, 1H, H-$4_{A'}$), 1.33 (d, $J_{5,6}$ = 6.3 Hz, 1H, H-$6_{A'}$). ^{13}C NMR (D_2O) δ164.9 (NHCO), 133.3 (*C*H=CH_2), 118.5 (CH=*C*H_2), 101.4 (C-$1_{D'}$, $^1J_{C,H}$ = 164.4 Hz), 97.9 (C-$1_{A'}$, $^1J_{C,H}$ = 172.9 Hz), 91.6 (CCl_3), 78.0 (C-$2_{A'}$), 75.9 (C-$4_{D'}$), 73.2 (C-$3_{D'}$), 72.4 (C-$4_{A'}$), 70.1 (C-$3_{A'}$), 70.0 (C-$5_{D'}$), 68.8 (C-$5_{A'}$), 68.3 (C_{All}), 60.6 (C-$6_{D'}$), 57.8 (C-$2_{D'}$), 16.6 (C-$6_{A'}$) ; HR-ESI-TOF-MS *m/z* 532.0504 (Calcd. for $C_{17}H_{26}Cl_3NO_{10}Na$, 532.0520).

Bacterial strains, plasmids & chemicals.

Details are provided in Supplementary Material and Methods.

Library construction.

The gene library was constructed in 5 PCR steps using Phusion^{R} polymerase (New England Biolabs, Ipswich, MA, USA) (Figure S2 of the Supporting Information). In the first step, the 6 DNA wild-type fragments surrounding the 5 regions containing designed mutations were amplified (Figure S2 of the Supporting Information).

Amplified products were purified by a gel extraction kit. In a second step, PCR reactions were divided in 10 pools (Figure S2 of the Supporting Information). In this way, in pool 1 , the fragment F1 (10 ng) was mixed with 226-229_rev (0.5 µM), the reverse primer containing mutated codons for 226, 227, 228 and 229 positions. In pool 2, the fragment F2 was mixed with 226-229_for, the forward primer containing mutated codons for 226, 227, 228 and 229 positions. In pool 3, the fragment F2 was mixed with 289-300_rev, the reverse primer containing mutated codons for 289, 290 and 300 positions. And so on for the 10 pools. PCR was conducted in a total volume of 20 µL in the following conditions : one step at 98°C for 30 sec, followed by 20 cycles at 98°C-10 sec, 52°C-30 sec (pool 1 to 10), 72°C-20 sec. DNA fragments and primers used in each pool contained overlapping regions of at least 20 nucleotides. The primers containing degenerated codons were specific to each library (Table S1 of the Supporting Information). In a third step, amplicons obtained from step 2 were mixed two by two to perform 1 cycle at 98°C for 30 sec and 10 cycles at 98°C for 10 sec and 72°C for 40 sec. In a fourth step, 10 µL of each mixture coming from the third step were mixed together to perform 1 cycle at 98°C for 30 sec and 10 cycles at 98°C for 10 sec and 72°C for 1 min. No Phusion polymerase was added in steps 3 and 4. In a fifth step, the gene library obtained was finally amplified by adding outer primers (see Supplementary Material and Methods provided in Supporting Information).

Amplified products were purified by a gel extraction kit. Amylosucrase libraries were cloned into pGEX-6P-3 (GE Healthcare Biosciences) using EcoRI HF and NotI HF restriction sites. Ligation products were transformed in *E.coli* TOP 10 electrocompetent cells (Invitrogen, Carlsbad, U.S.A.). Freshly transformed cells were plated on LB agar supplemented with 100 µg/mL ampicillin for 24 h at 37°C.

Library screening on sucrose in solid media.

A pH-based high-throughput screening assay earlier described [35] was used to isolate sucrose active clones (see Supplementary Material and Methods provided in Supporting Information).

Expression of active clones in microplate format.

Active mutants were replicated to inoculate a 96-well microplates containing LB medium (200µl), ampicillin (100 µg/mL^{-1}) and glycerol (12% w/v) and incubated at 30°C for 24h under agitation at 450 Rpm (orbital diameter of 3 mm). Plates were then duplicated into 96-Deep Well plates containing in each well 1 mL of autoinducing media ZYM-5052 [48] supplemented with ampicillin (100 µg/mL) for culture and gene expression. These plates were incubated 24 hours at 30°C with agitation at 700 Rpm. To evaluate the growth homogeneity, absorbance was measured at 600 nm for each culture diluted 10 times in water (Sunrise spectrophotometer, Tecan, Männedorf, Switzerland). Plates were then centrifuged (20 min, 3,000 g, 4°C), supernatants were removed and the cell pellets were resuspended in 200 µL of 2X PBS buffer

containing Lyzozyme (0.5 mg/mL) and DNAse and frozen at -80°C overnight. After thawing at room temperature, microplates were centrifuged (20 min, 3,000 g, 4°C); soluble fractions were then transferred in new microplates for reaction assays. Reaction assays in the presence of sucrose were carried out as previously described. [19,20] All 55 positive hits were sequenced on the entire gene by GATC Biotech GA (Constance, Germany).

Reaction assays with sucrose and D'A' acceptor in microplate format.

Reactions were carried out in microplates with 100 µL of soluble fractions and 50 µL of sucrose solution (73mM final concentration) and 50 µL of **D'A'** acceptor solution (73mM final concentration). The enzymatic reaction was incubated at 30°C for 24 h under agitation (450 Rpm). The reactions were then stopped by heating at 95°C for 5 min, centrifugated (20 min at 3,700 Rpm), filtered and analyzed by HPLC using conditions described hereafter.

Carbohydrate analysis.

Sucrose consumption was evaluated by ion exchange chromatography at 65°C using an Aminex HP87K column (300 x 7.8 mm, Biorad) with ultra-pure water at 0.6 mL/min as eluent. The detection was performed using a shodex RI 101 series refractometer. D'A' consumption and ED'A' and EED'A' production were identified using a reversed C18 phase column (Venusil AQ 50 mm) with MeCN/water (17:83, v/v) as eluent. The detection was performed using a UV detector (Dionex, UVD 340)) at 210 and 220 nm. The glucosylated molecule was identified by LC-MS with the above conditions. Using a MSQ® Plus Mass Spectrometer (Dionex - Thermoscientific equipment), molecules were ionised by electrospray ionisation technique with a source at 450°C. Separation was achieved with a quadripole and detected using the negative mode.

Production and purification of isolated mutants.

E. coli TOP10 electrocompetent cells containing pGST-AS mutated gene were produced as previously described[49] but with a preculture step at 30°C overnight instead of 37°C and then cultures at 23°C. After sonication, the crude extract was centrifugated (10,000 Rpm, 20 min, 4°C) and the supernatant was harvested for the purification step. *Np*AS mutants were purified on an affinity chromatography column via affinity of the GST tag to the Glutathion Sepharose 4B support (Dutscher, Brumath, France) into Tris buffer (50mM Tris, 150mM NaCl, 1mM EDTA, 1mM DTT, pH 7.0). GST Tag was removed by PreScission protease (Dutscher, Brumath, France). Protein concentration was estimated using Nanodrop ND-1000 spectrophotometer and specific activity was determined at 30°C, in 50 mM Tris-HCl buffer pH 7.0, using 50 g/L sucrose during 20 min by dinitrosalicylic acid assay. One unit of amylosucrase activity corresponds to the amount of enzyme that catalyses the release of 1 µmol of reducing sugars per minute in the assay conditions.

Characterization of product profiles of the active mutant on D'A'.

Reactions were performed at 30°C during 24 h with 146mM sucrose alone or supplemented with 146mM acceptor molecule. Wild-type enzyme and mutants were employed at 1 U/mL and reaction mixes were stopped by heating at 95°C during 5 min. Product profiles were analyzed by HPAEC-PAD using a Dionex Carbo-Pack PA100 column at 30°C. Soluble fraction was analyzed with a gradient of sodium acetate (from 6 to 300mM in 28 min) in 150mM NaOH and the crude extract was solubilized in 1M aqueous KOH. Detection was performed

using a Dionex ED40 module with a gold working electrode and Ag/AgCl reference electrode.

Kinetic parameters determination of active mutant.

Enzyme kinetics were carried out with pure enzymes (0.2 and 0.1 U/ml) at 30°C, under agitation. Catalytic efficiency of wild-type *Np*AS and mutants was determined with sucrose alone from 0 to 600mM and with both sucrose (146mM) and acceptor molecule (from 5mM to 500mM) as variable substrates. Catalytic efficiencies were determined by measuring the release of fructose monitored for sucrose alone ($k_{cat}/K_{m,\ suc}$), and by measuring the rate of formation ($k_{cat}/K_{m,suc}$; $k_{cat}/K_{m,\mathbf{D'A'}}$) of the monoglucosylated disaccharide. After heating at 95°C, and centrifugation, aliquots were analyzed by HPLC RI with a Biorad HPX-87K column.

Semi-preparative synthesis and identification of allyl α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→4)-2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside (EED'A').

A mixture of 146mM D'A', 292mM sucrose, and 1 U/mL mutant F3 (2 mL) was gently stirred at 30°C. After 24 h, the reaction mix was heated at 95°C for 5 min, and then centrifugated (6,000 Rpm, 4°C, 20 min). The glucosylated products were separated by reverse phase semi-preparative chromatography using a C18 column (Bishoff Chromatography, 250mm X 8mm), eluting with 17% acetonitrile in pure water at a constant flow rate of 2 mL/min at 30°C. The major product was lyophilized repeatedly to give tetrasaccharide EED'A'. This product had ^{1}H NMR (D_2O): δ 5.99-5.89 (m, 1H, C*H*=CH_2), 5.43 (d, $J_{1,2}$ = 3.8 Hz, 1H, H-1_E), 5.35-5.31 (m, J_{trans} = 17.3 Hz, 1H, CH=CH_2), 5.27-5.25 (m, J_{cis} = 10.5 Hz, 1H, CH=CH_2), 4.98 (bs, 1H, H-$1_{A'}$), 4.94 (d, $J_{1,2}$ = 3.6 Hz, 1H, H-$1_{E'}$), 4.90 (d, $J_{1,2}$ = 8.4 Hz, 1H, H-$1_{D'}$), 4.19 (m, 1H, H_{All}), 4.07-4.01 (m, 3H, H-$2_{A'}$, H-$3_{D'}$, H_{All}), 3.99-3.91 (m, 2H, H-$6a_{E'}$, H-$6a_{D'}$), 5.88 (m, 1H, H-$5_{E'}$), 3.84-3.77 (m, 3H, H-$6a_{D'}$, H-$3_{A'}$, H-5_E), 3.75-3.68 (m, 5H, H-$6a_E$, H-3_E, H-$6b_E$, H-$6b_{E'}$, H-$4_{D'}$), 3.66-3.63 (m, 2H, H-$3_{E'}$, H-$5_{A'}$), 3.60-3.55 (m, 2H, H-$2_{E'}$, H-$5_{A'}$), 3.51 (dd_{po}, 1H, $J_{2,3}$ = 10.4 Hz, 1H, H-2_E), 3.48 (pt_{po}, 1H, $J_{3,4}$ = $J_{4,5}$ = 9.3 Hz, H-$4_{E'}$), 3.41 (pt_{po}, 1H, $J_{3,4}$ = $J_{4,5}$ = 9.2 Hz, H-4_E), 3.40 (pt_{po}, 1H, $J_{3,4}$ = $J_{4,5}$ = 9.6 Hz, H-$4_{A'}$), 1.23 (d, $J_{5,6}$ = 6.5 Hz, 1H, H-$6_{A'}$) ; ^{13}C NMR (D_2O): δ 167.5 (NHCO), 136.0 (*C*H=CH_2), 121.1 (CH=*C*H_2), 104.0 (C-$1_{D'}$, $^1J_{C,H}$ = 164.1 Hz), 102.1 (C-$1_{E'}$, $^1J_{C,H}$ = 172.7 Hz), 100.7 (C-1_E, $^1J_{C,H}$ = 171.8 Hz), 100.5 (C-$1_{A'}$, $^1J_{C,H}$ = 173.7 Hz), 94.2 (CCl_3), 80.7 (C-$2_{A'}$), 79.3 (C-3_E), 77.2 (C-$5_{D'}$), 76.3 (C-$3_{D'}$), 75.8 (C-$3_{A'}$), 75.7 (C-$4_{D'}$), 75.0 (C-$4_{A'}$), 74.5 (C-3_E), 74.1 (2C, C-$2_{E'}$, C-2_E), 74.0 (C-$5_{E'}$), 72.7 (C-5_E), 72.2 (C-$4_{E'}$), 72.0 (C-4_E), 71.4 (C-$5_{A'}$), 70.9 (C_{All}), 68.5 (C-$6_{E'}$), 63.3 (C-$6_{D'}$), 63.1 (C-6_E), 60.4 (C-$2_{D'}$), 19.2 (C-$6_{A'}$). *m/z* (ES-) for EED'A', m/z 832 [M - H^+].

ACKNOWLEDGMENT

The authors are grateful to C. Topham for providing his help with Sequester analysis. They also thank G. Cioci and S. Tranier in using the PICT facility for protein biophysical characterization, N. Monties for her help in the product analysis and purification by HPAEC-PAD and LC-MS as well as S. Bozonnet and S. Pizzut-Serin in using the ICEO high-throughput facility, devoted to the engineering and screening of new and original enzymes, of the Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (Toulouse, France). The authors wish to thank C. Guerreiro (Unité de Chimie des Biomolécules, Institut Pasteur, Paris, France) for her technical contribution to the synthetic aspects of this work, and F. Bonhomme (CNRS, UMR 3523, Institut Pasteur, Paris, France) for providing the HRMS data. This work was granted access to the HPC resources of the Computing Center of Region Midi-Pyrénées (CALMIP, Toulouse, France).

ABBREVIATIONS

*Np*AS, *Neisseria polysaccharea* amylosucrase; LPS, lipopolysaccharide; O-SP: O-specific polysaccharide; GH, glycoside-hydrolase; GST-AS, glutathione-S-transferase fusioned to amylosucrase; DNA; deoxyribonucleic acid; IPTG, isopropyl β-D-thiogalactopyranoside; PCR, polymerase chain reaction; LB, lysogeny broth; DNS, DiNitroSalicylic acid; HPLC, high performance liquid chromatography; RI, refractive index; UV, ultra violet; DSF, Differential Scanning Fluorimetry; LC-MS, Liquid Chromatography-Mass Spectrometry; DP, Degree of Polymerization; GH, Glycoside-Hydrolase; Tm, melting temperature; mp, Melting Point; t_r, retention time; bp, boiling point; **E**, α-D-Glucopyranosyl; **A'** , allyl α-L-rhamnopyranoside; **D'**, 2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl.

REFERENCES

(1) Turner, N. J.; Truppo, M. D. *Curr. Opin. Chem. Biol.* **2013**, *17*, 212-214.

(2) Turner, N. J.; O'Reilly, E. *Nat. Chem. Biol.* **2013**, *9*, 285-288.

(3) Davids, T.; Schmidt, M.; Böttcher, D.; Bornscheuer, U. T. *Curr. Opin. Chem. Biol.* **2013**, *17*, 215-220.

(4) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. *Nature* **2012**, *485*, 185-194.

(5) Höhne, M.; Schätzle, S.; Jochens, H.; Robins, K.; Bornscheuer, U. T. *Nat. Chem. Biol.* **2010**, *6*, 807-813.

(6) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 5700-5725.

(7) Hilvert, D. *Annu. Rev. Biochem.* **2013**, *82*, 447-470.

(8) Kries, H.; Blomberg, R.; Hilvert, D. *Curr. Opin. Chem. Biol.* **2013**, *17*, 221-228.

(9) Wijma, H. J.; Janssen, D. B. *FEBS J.* **2013**, *280*, 2948-2960.

(10) Voigt, C. A.; Mayo, S. L.; Arnold, F. H.; Wang, Z.-G. *Proc. Natl. Acad. Sci.* **2001**, *98*, 3778-3783.

(11) Voigt, C. A.; Mayo, S. L.; Arnold, F. H.; Wang, Z. *J. Cell. Biochem. Suppl.* **2002**, *37*, 58-63.

(12) Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B. *Protein Eng. Des. Sel.* **2014**, *27*, 49-58.

(13) Goldsmith, M.; Tawfik, D. S. In *Enzyme Engineering by Targeted Libraries*; Keating A.E., Eds; Methods in Enzymology; Vol 523, Academic Press, **2013**; 523, 257–283.

(14) Allen, B. D.; Nisthal, A.; Mayo, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19838-19843.

(15) Lippow, S. M.; Moon, T. S.; Basu, S.; Yoon, S.; Li, X.; Chapman, B. A. *Chem. Biol.* **2010**, 17, 1306-1315.

(16) Chaparro-riggers, J. F.; Polizzi, K. M.; Bommarius, A. S. *Biotechnol. J.* **2007**, 2, 180-191.

(17) Floor, R.J.; Wijma, H.J.; Colpa, D.I.; Ramos-Silva, A.; Jekel, P.A.; Szymański, W.; Feringa, B.L.; Marrink, S.J.; Janssen, D. B. *ChemBioChem* **2014**, *15*, 1660-1672.

(18) André, I.; Potocki-Véronèse, G.; Barbe, S.; Moulis, C.; Remaud-Siméon, M. *Curr. Opin. Chem. Biol.* **2014**, *19*, 17-24.

(19) Champion, E.; André, I.; Moulis, C.; Boutet, J.; Descroix, K.; Morel, S.; Monsan, P.; Mulard, L. A.; Remaud-Siméon, M. *J. Am. Chem. Soc.* **2009**, *131*, 7379-7389.

(20) Champion, E.; Guérin, F.; Moulis, C.; Barbe, S.; Tran, T. H.; Morel, S.; Descroix, K.; Monsan, P.; Mourey, L.; Mulard, L. A.; Tranier, S.; Remaud-Siméon, M.; André, I. *J. Am. Chem. Soc.* **2012**, *134*, 18677-18688.

(21) Perepelov, A. V.; Shekht, M. E.; Liu, B.; Shevelev, S. D.; Ledov, V. A.; Senchenkova, S. N.; L'vov, V. L.; Shashkov, A. S.; Feng, L.; Aparin, P. G.; Wang, L.; Knirel, Y. A. *FEMS Immunol. Med. Microbiol.* **2012**, *66*, 201-210.

(22) Shashkov, A.S.; Senchenkova, S.N.; Sun, Q.; Lan, R.; Wang, J.; Perepelov, A.V.; Knirel, Y.A.; Xu, J. *Carbohydr. Res.* **2013**, *24*, 93-96.

(23) Skov, L. K.; Mirza, O.; Henriksen, A.; Potocki de Montalk, G.; Remaud-Siméon, M.; Sarçabal, P.; Willemot, R.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2001**, *276*, 25273-25278.

(24) Skov, L. K.; Mirza, O.; Gajhede, M.; Feller, G.; D'Amico, S.; André, G.; Potocki-Veronese, G.; Van der Veen, B.; Monsan, P.; Remaud-Siméon, M. *J. Biol. Chem.* **2004**, *279*, 726-734.

(25) Skov, L. K.; Mirza, O.; Sprogøe, D.; Dar, I.; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2002**, *277*, 47741-47747.

(26) Hanson, S.; Best, M.; Bryan, M. C.; Wong, C-H. *Trends Biochem. Sci.* **2004**, *29, 656-663*.

(27) Wang, Z.; Chinoy, Z.; Ambre, S.;Peng, W; McBride, R.; de Vries, R.P.; Glushka, J.; Paulson, J.C.; Boons, G.J. *Science* **2013**, 341,379-383.

(28) Armstrong, Z.; Withers, S.G. *Biopolymers*, **2009**, *99*, 666-674.

(29) Muthana, S.; Cao, H.; Chen, X. *Curr. Opin. Chem. Biol.* **2009**, *13*, 573-581.

(30) Lucas, R.; Hamza, D.; Lubineau, A.; Bonnaffé, D. *Eur. J. Org. Chem.* **2004**, 10, 2107-2117.

(31) Boutet, J.; Guerreiro, C.; Mulard, L. A. *J Org Chem* **2009**, *74*, 2651-2670.

(32) Chassagne, P.; Raibaut, L.; Guerreiro, C.; Mulard, L. A. *Tetrahedron* **2013**, *69*, 10337-10350.

(33) Reetz, M. T.; Kahakeaw, D.; Lohmer, R. *ChemBioChem* **2008**, *9*, 1797-1804.

(34) Stam, M. R.; Danchin, E. G. J.; Rancurel, C.; Coutinho, P. M.; Henrissat, B. *Protein Eng. Des. Sel.* **2006**, *19*, 555-562.

(35) Champion, E.; Moulis, C.; Morel, S.; Mulard, L. A.; Monsan, P.; Remaud-Siméon, M.; André, I. *ChemCatChem* **2010**, *2*, 969-975.

(36) Pinto B.M., Morisette D.G., B. D. R. *J. Chem. Soc. Perkin Trans.* **1987**, *1*, 9-14.

(37) Ley, S. V.; Priepke, H.; W. M.; Warriner, S. L. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 2290-2292.

(38) Ley, S. V.; Polara, A. *J. Org. Chem.* **2007**, *72*, 5943-5959.

(39) Montchamp, J.-L.; Tian, F.; Hart, M. E.; Frost, J. W. *J. Org. Chem.* **1996**, *61*, 3897-3899.

(40) Boutet, J.; Kim, T. H.; Guerreiro, C.; Mulard, L. A. *Tetrahedron Lett.* **2008**, *49*, 5339-5342.

(41) Mirza, O.; Skov, L. K.; Remaud-Siméon, M.; Potocki de Montalk, G. P.; Albenne, C.; Monsan, P.; Gajhede, M. *Biochemistry* **2001**, *40*, 9032-9039.

(42) Malbert, Y.; Pizzut-Serin, S.; Massou, S.; Cambon, E.; Laguerre, S.; Monsan, P.; Lefoulon, F.; Morel, S.; André, I.; Remaud-Siméon, M. *ChemCatChem* **2014**, 16, 2282-2291.

(43) Cambon, E.; Barbe, S.; Pizzut-Serin, S.; Remaud-Siméon, M.; André, I. *Biotechnol. Bioeng.* **2014**, 111, 1719-1728.

(44) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins Struct. Funct. Genet.* **2006**, *65*, 712-725.

(45) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622-655.

(46) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(47) Case, D. A.; Darde, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollmann, P. A. 2006, AMBER 9, *Univ. California, San Francisco, CA, USA.*

(48) Studier, F. W. *Protein Expr. Purif.* **2005**, *41*, 207-234.

(49) Potocki de Montalk, G.; Remaud-Siméon, M.; Willemot, R. M.; Planchot, V.; Monsan, P. *J. Bacteriol.* **1999**, *181*, 375-381.

(50) Lemanski, G.; Ziegler, T. *Eur. J. Org. Chem.*, **2006**, 2618-2630.

(51) Dhara, D.; Kar, R.K.; Bhunia, A.; Misra, A.K. *Eur. J. Org. Chem.* **2014**, 4577–4584

(52) Hargreaves, J.; Le Guen, Y.; Guerreiro, C.; Descroix, K.; Mulard, L. A., *Org. Biomol. Chem.* **2014**, *12*, 7728-7749.

Supporting Information:

showing design of libraries, construction, screening and analysis, general methods for chemical, 1H and 13C NMR spectra for novel compounds.

Figure S1. Representation of the variety of α-D-glucosylation (E) patterns found in the DA component of the linear ABCD repeating unit of serotype-specific *S. flexneri* O-SPs [S11](a) and lightly protected D'A' acceptor selected as precursor to the DA disaccharide component from the ABCD backbone repeating unit of several *S. flexneri* O-SPs (b).

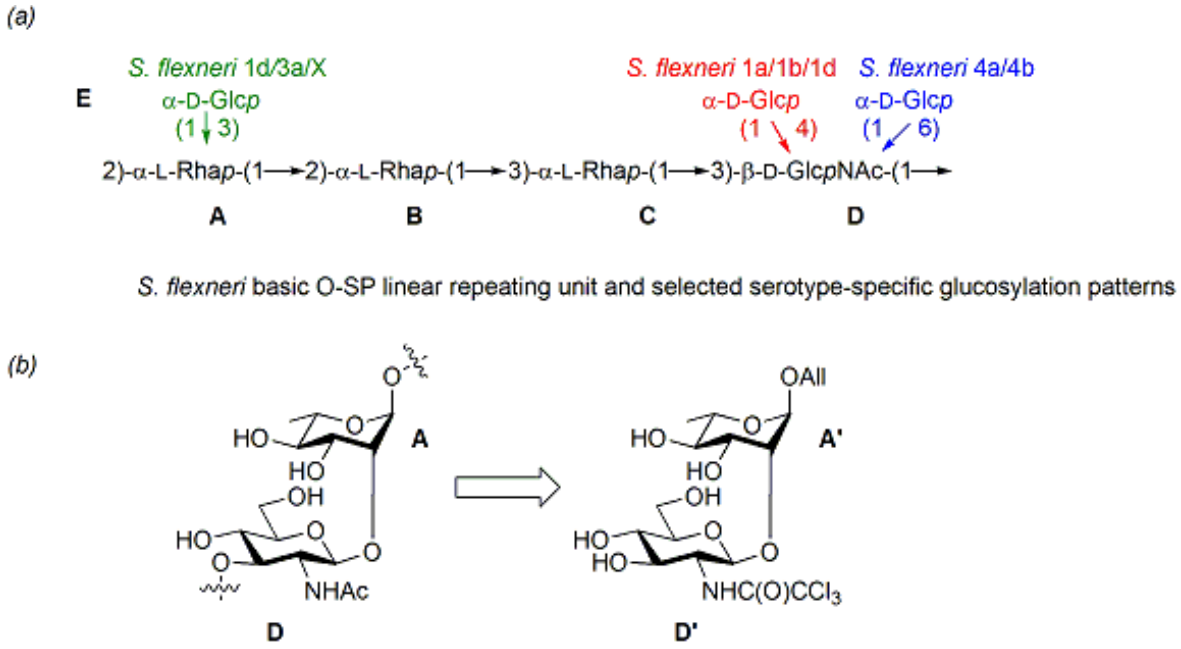
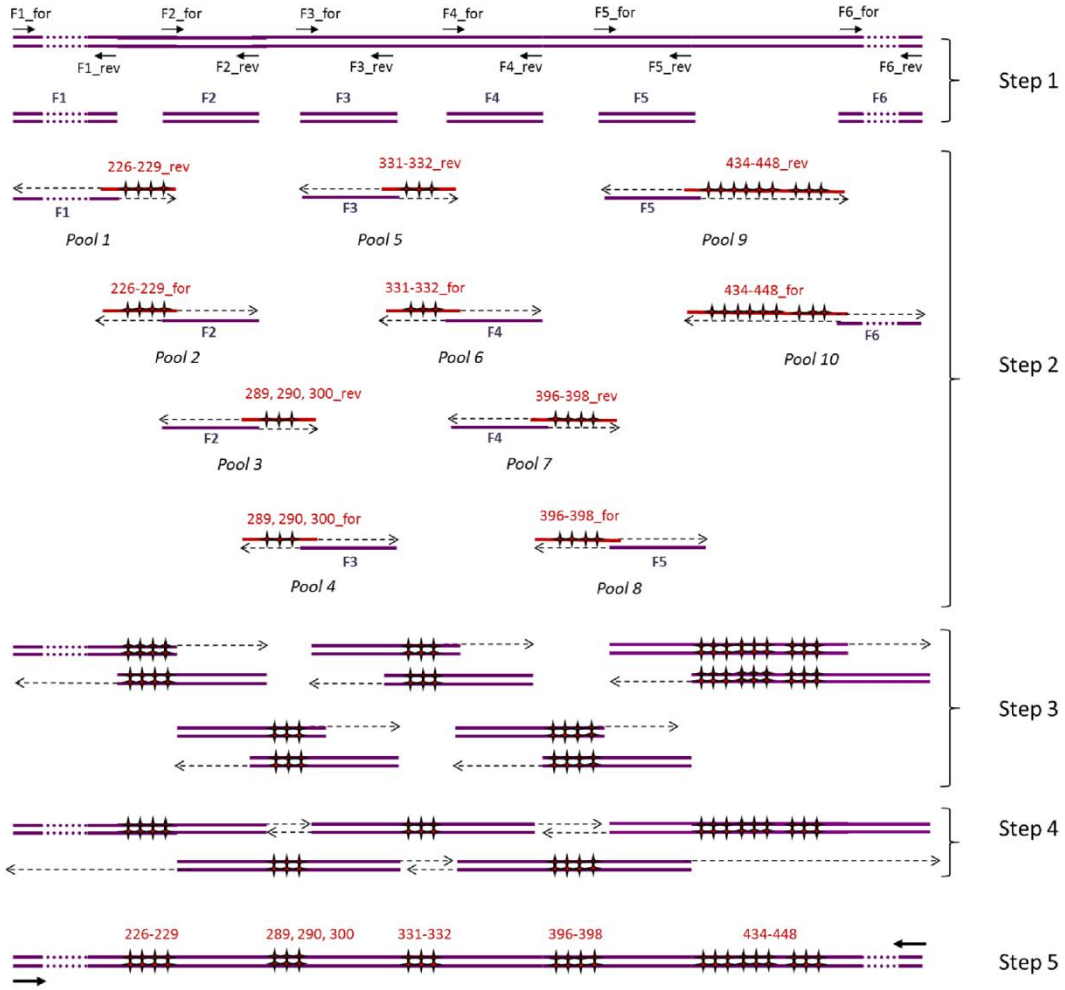


Figure S2. Strategy of construction of the gene library for simultaneous incorporation of 23 mutations authorized by the computational design



Library construction: The gene library was constructed in 5 PCR steps using Phusion^{R} polymerase (New England Biolabs, Ipswich, MA, USA). In the first step, the 6 DNA wild-type fragments surrounding the 5 regions containing designed mutations were amplified. In the first step, the 6 DNA wild-type fragments surrounding the 5 regions containing designed mutations were amplified (98°C, 30 sec; 1 cycle; 98°C, 10 sec; 55°C, 30 sec; 72°C, 15 sec; 30 cycles; 72°C, 10 min; 1 cycle) with the 6 following primer couples: F1-for (GAA GTT CTG TTC CAG GGG CCC CTG), F1-rev (CAG GGT GCG GTC GTA TTG GTC G); F2-for (ACC CGG GCG GCT TCT CGC AAC T), F2-rev (GCA GGA TGT CAA CGC CCA AGT TG); F3-for (CGC AGG CGC ACG CCC TCA TC), F3-rev (CGG ATT TGA AGA ACA CGG CGG G); F4-for (GAC CAA GTC GTC CAA TAC ATC GG), F4-rev (GTC GTG GCT GCG GAC GTA GTT G); F5-for (CCG ATG AAG ACG CGG CAT ATC), F5-rev (GAG CGA AGC TGC CGT CGA AAC); F6-for (GGT ACA GCC GCG GCA TTG GTC), F6-rev (CGC GAG GCA GAT CGT CAG TCA GTC).

Amplified products were purified by a gel extraction kit. In a second step, PCR reactions were divided into 10 pools. In this way, in pool 1, the fragment F1 (10 ng) was mixed with 226-229_rev (0.5µM), the reverse primer containing mutated codons for 226, 227, 228 and 229 positions. In pool 2, the fragment F2 was mixed with 226-229_for, the forward primer containing mutated codons for 226, 227, 228 and 229 positions. In pool 3, the fragment F2 was mixed with 289-300_rev, the reverse primer containing mutated codons for 289, 290 and 300 positions. And so on for the 10 pools. PCR was conducted in a total volume of 20 µL in the following conditions: one step at 98°C for 30 sec, followed by 20 cycles at 98°C-10 sec, 52°C-30 sec (pool 1 to 10), 72°C-20 sec. DNA fragments and primers used in each pool contained overlapping regions of at least 20 nucleotides. The primers containing degenerated codons were specific to each library (Table S1). In a third step, amplicons obtained from step 2 were mixed two by two to perform 1 cycle at 98°C for 30 sec and 10 cycles at 98°C for 10 sec and 72°C for 40 sec. In a fourth step, 10 µL of each mixture coming from the third step were mixed together to perform 1 cycle at 98°C for 30 sec and 10 cycles at 98°C for 10 sec and 72°C for 1 min. No Phusion^{R} polymerase was added in steps 3 and 4. In a fifth step, the gene library obtained was finally amplified by adding outer primers (see supplementary material).

Primers and conditions for the final amplication of the entire genes (forward_TCC CCG AAT TCA CAG TAC CTC**;** reverse_TCA CGA TGC GGC CGC CTG AAA CGG TTC A) and performing the following program: (98°C, 30 sec; 1 cycle; 98°C, 10 sec; 65°C, 15 sec; 72°C, 1 min; 30 cycles; 98°C, 10 sec; 72°C, 1 min; 10 cycles; 72°C, 6 min; 1 cycle).

Amplified products were purified by a gel extraction kit. Amylosucrase libraries were cloned into pGEX-6P-3 (GE Healthcare Biosciences) using EcoRI and NotI restriction sites. Ligation products were transformed in *E.coli* TOP 10 electrocompetent cells (Invitrogen, Carlsbad, U.S.A.). Freshly transformed cells were plated on LB agar supplemented with 100µg/mL ampicillin for 24h at 37°C.

Figure S3. Distribution of the sucrose activity observed from primary screening of the 55 isolated mutants. DNS activity was normalized with respect to the wild-type enzyme (WT) = (DNS value of the mutant / OD_{600nm} of the mutant) / (DNS value WT / OD_{600nm} WT); all activity assays were carried out in triplicates and standard are represented by bars.

Close inspection of the sucrose-active mutant sequences revealed the presence of specific combinations of amino acid mutations at several positions, in particular at positions 289, 300 and 331 belonging to loops 4 and 5 of *Np*AS, which have been shown to play a key role in the definition of the catalytic pocket topology where sucrose binds. In particular, residue E300 which is described to be involved in a salt bridge interaction with R226 in the wild-type *Np*AS has been found frequently mutated by an Ile amino acid among the sequences of most active variants. The disengagement of I300 residue from such strong interaction enables a local reorganization of interactions between residues from the various loops. Also carried by loop 4, A289 is found often mutated by Ile amino acid while I331 from loop 5 is found mutated by a Thr residue. In close vicinity, these 3 mutations (A289I, E300I, and V331T) favor the formation of van der Waals interactions which are likely to play a role on the hinge motion of loop 4 that occurs during sucrose accessibility and recognition.

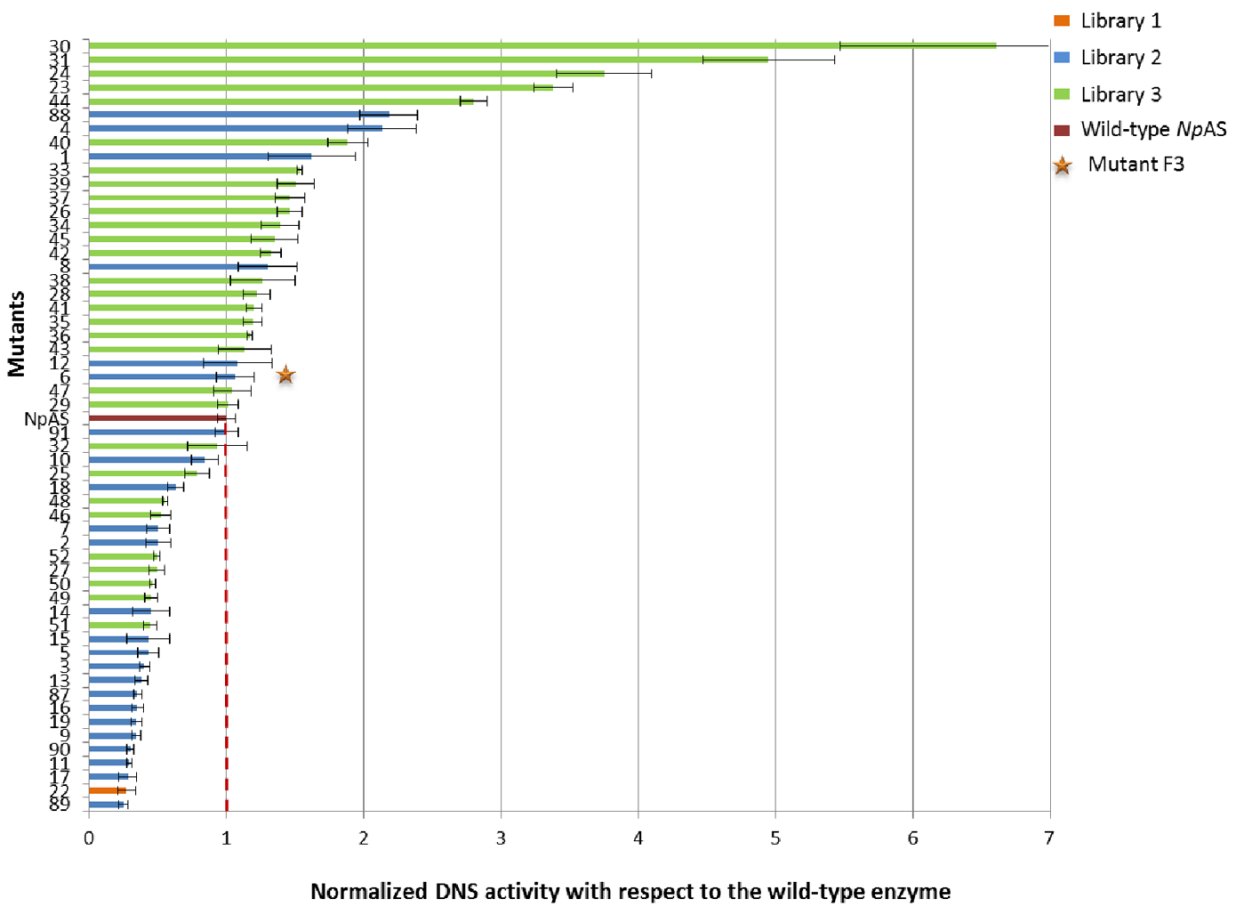
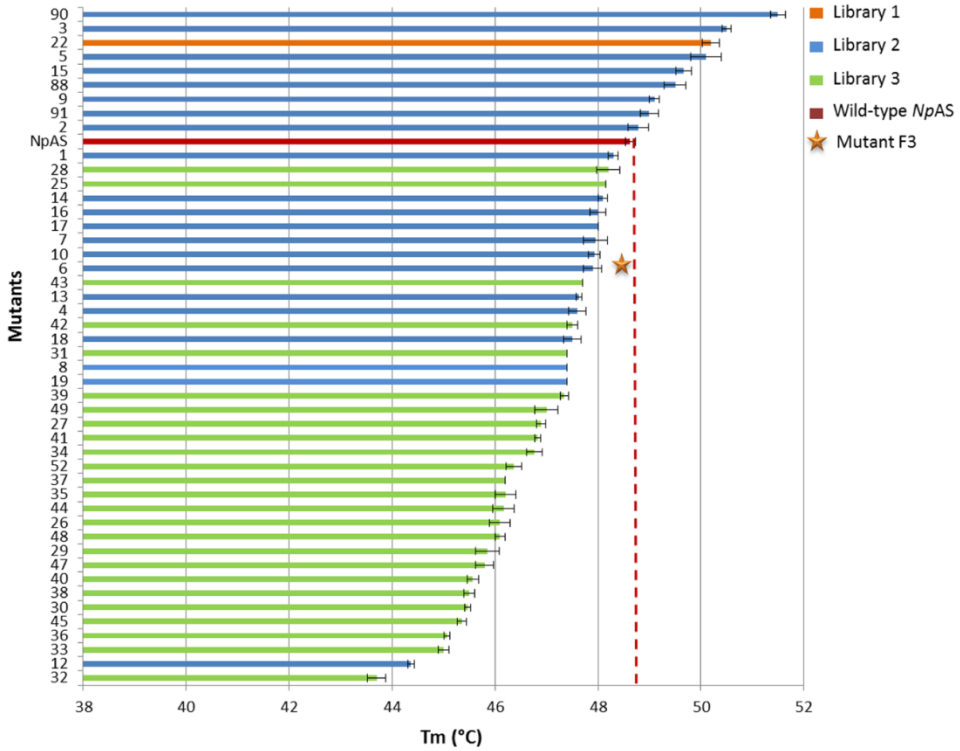


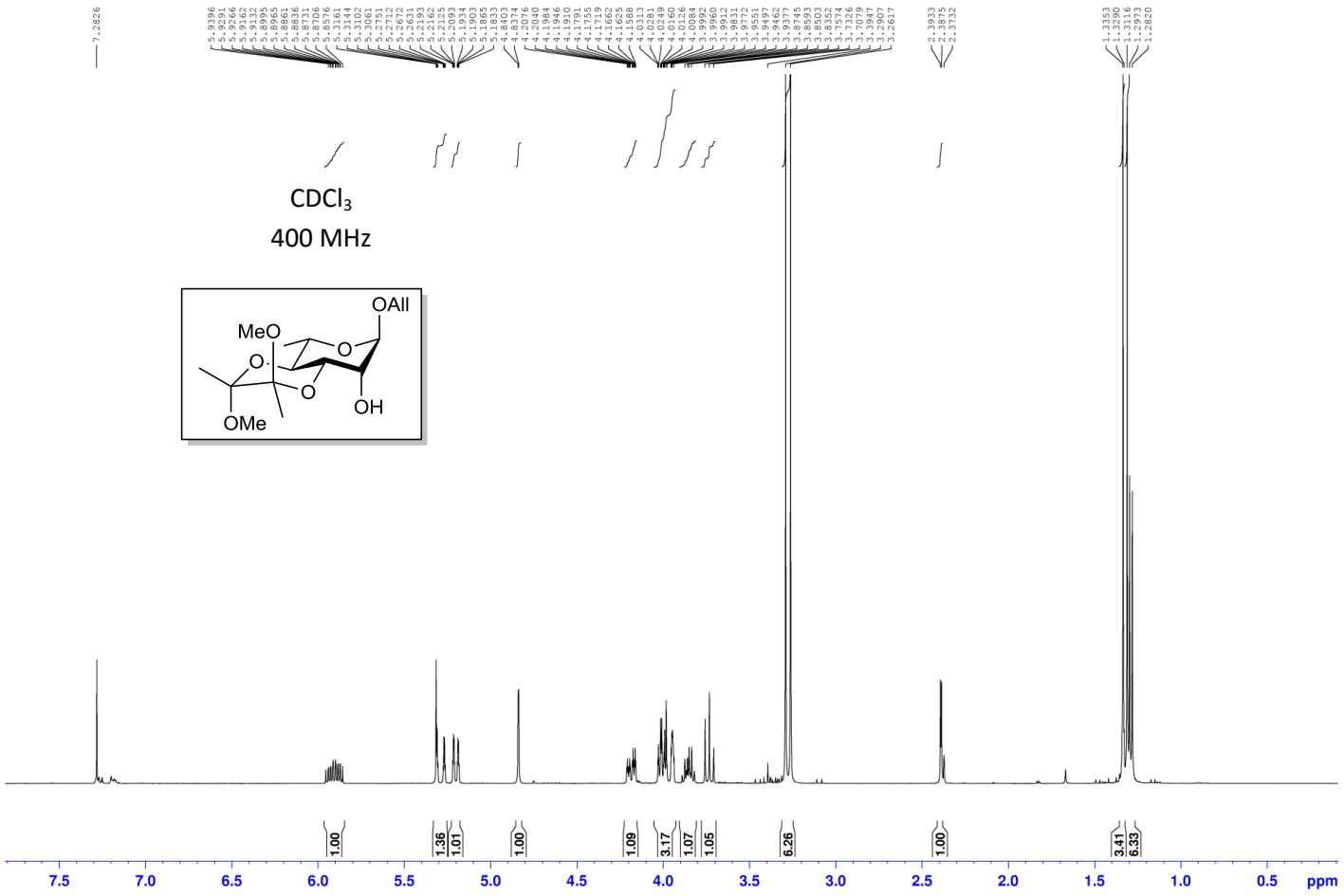
Figure S4. Thermal stability profiles determined by differential scanning fluorimetry on purified enzymes, Melting temperatures (Tm in degrees Celsius) were determined after triplicates experiments.



The melting point (Tm) of amylosucrase mutants was calculated by circular dichroism analyses on a JASCO J815 spectropolarimeter, with a Peltier cell temperature controller and by DSF with a thermal transition monitored using a Q-PCR CFX96 real-time System (Biorad, Marnes-la-Coquette, France) from 20 to 80°C. Purified enzymes were diluted at 2µM in Tris buffer (50mM Tris, 150mM NaCl, 1mM DTT, 1mM EDTA, pH7.0. DSF experiments were run at the Integrated Screening Platform of Toulouse (PICT) located at the Institute of Pharmacology and Structural Biology (Toulouse, France). Comparison of sequences corresponding to the most stable mutants with the least stable ones revealed important differences in the amino acids found at positions 289, 300 and 331, also shown to be important for the activity of the enzyme on sucrose[S9]. Most striking differences resided in the presence of an Ala residue at position 289 (unchanged with respect to wild-type *Np*AS), the presence of either Glu or Val at position 300 (when mutated in Ile, the stability was affected) and the presence of either Gly or Val at position 331 (when mutated in Thr, the stability was also affecte

NMR characterization of different compounds

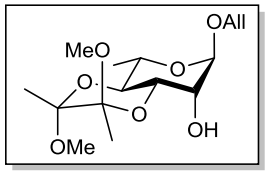
Compound **4**: Allyl 3,4-*O*-(2',3'-dimethoxybutane-2',3'-diyl)-α-L-rhamnopyranoside

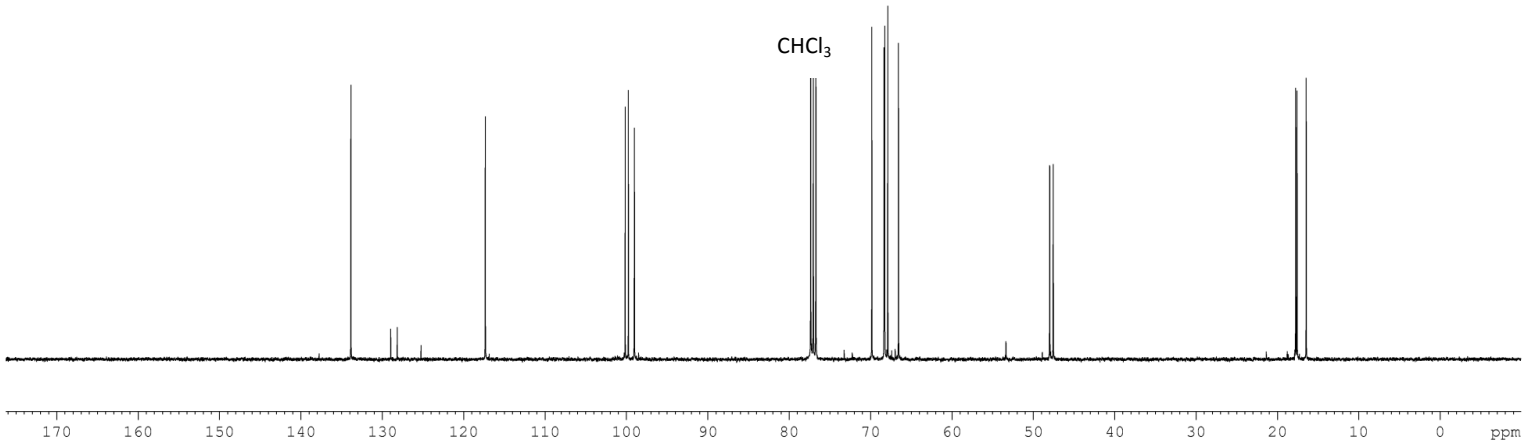


133.876
117.360
100.170
99.780
99.035
77.367
77.049
76.731
69.862
68.353
68.274
67.892
66.582
47.998
47.570
17.774
17.608
16.479

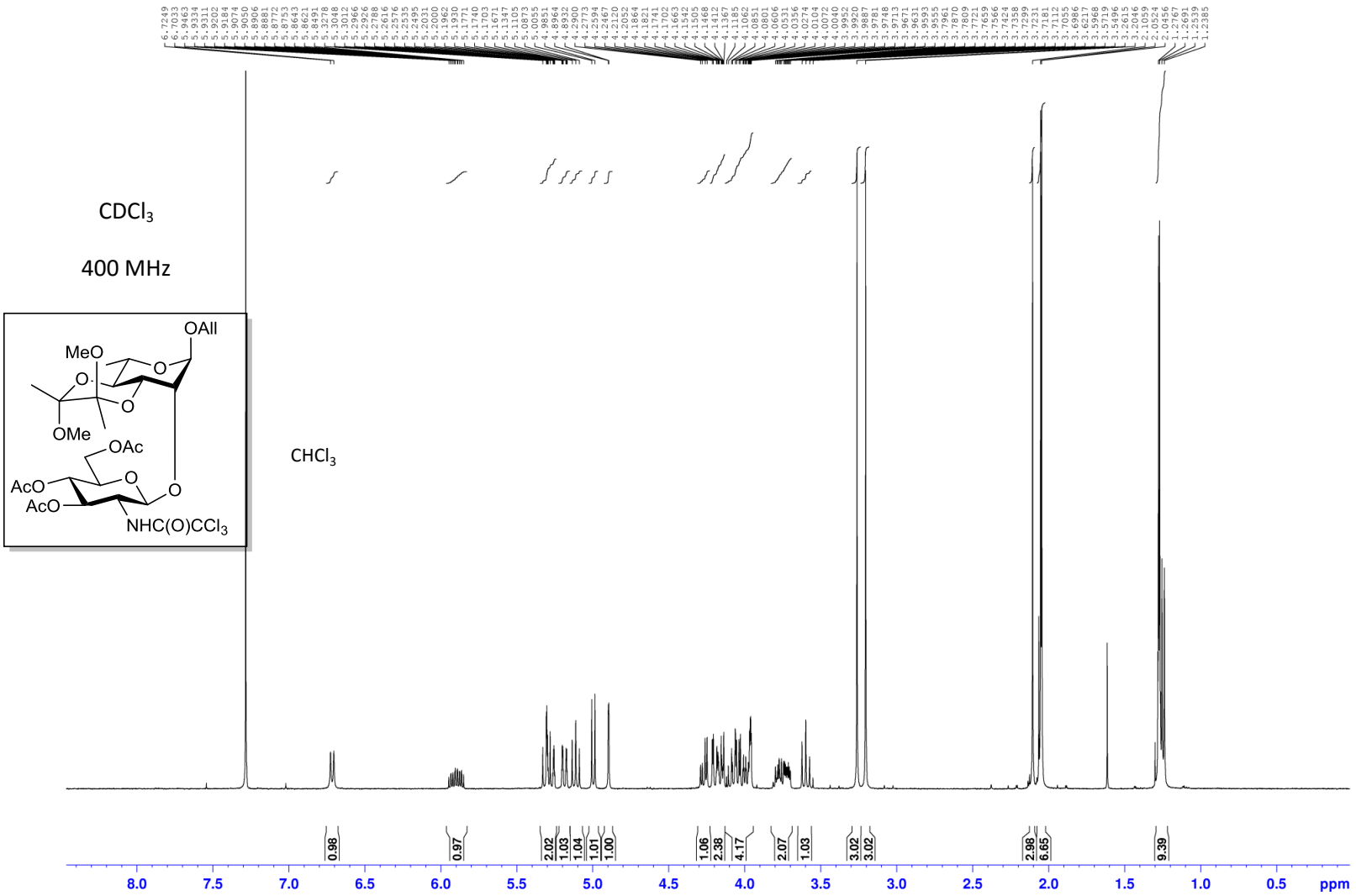
$CDCl_3$

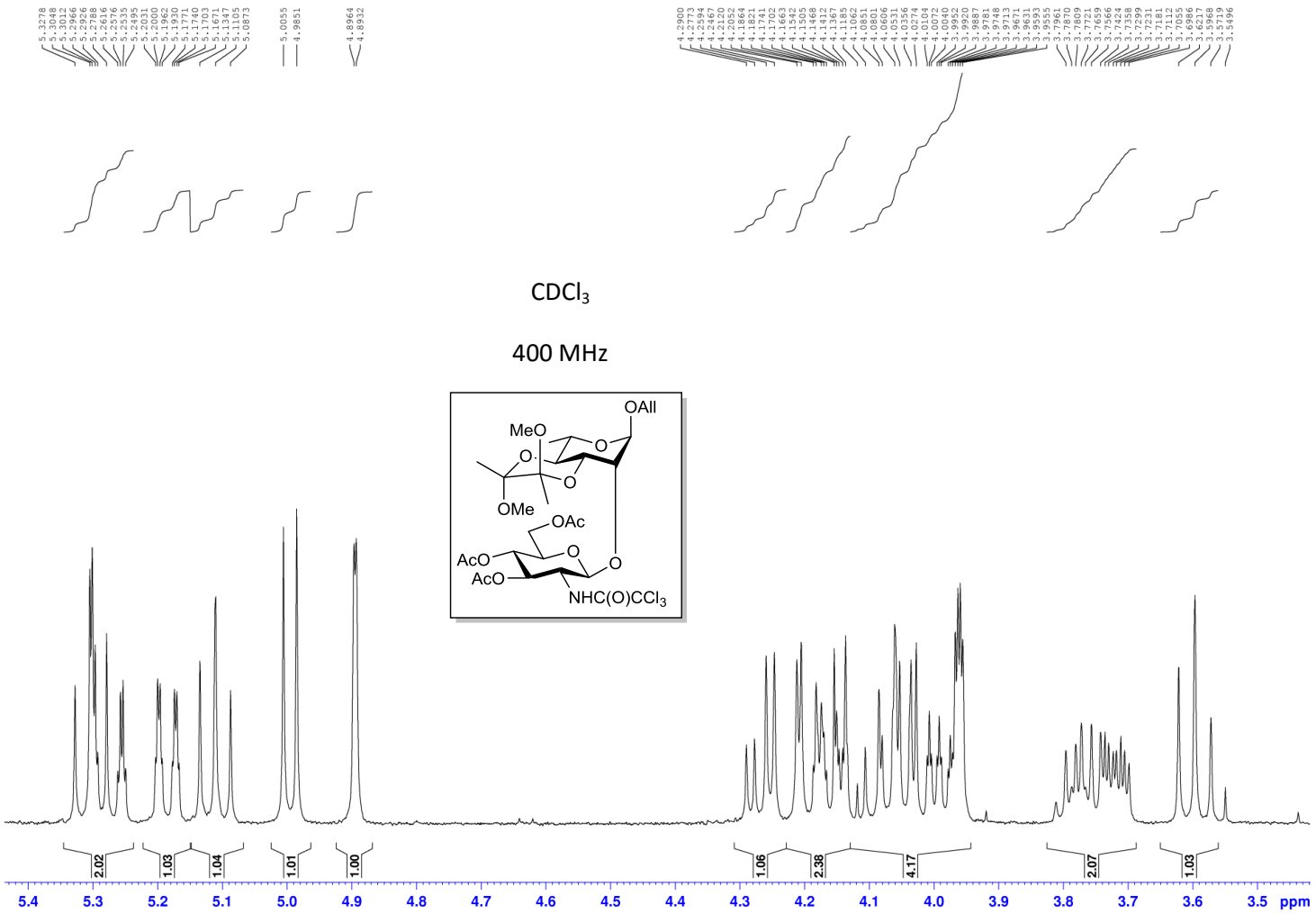
100 MHz

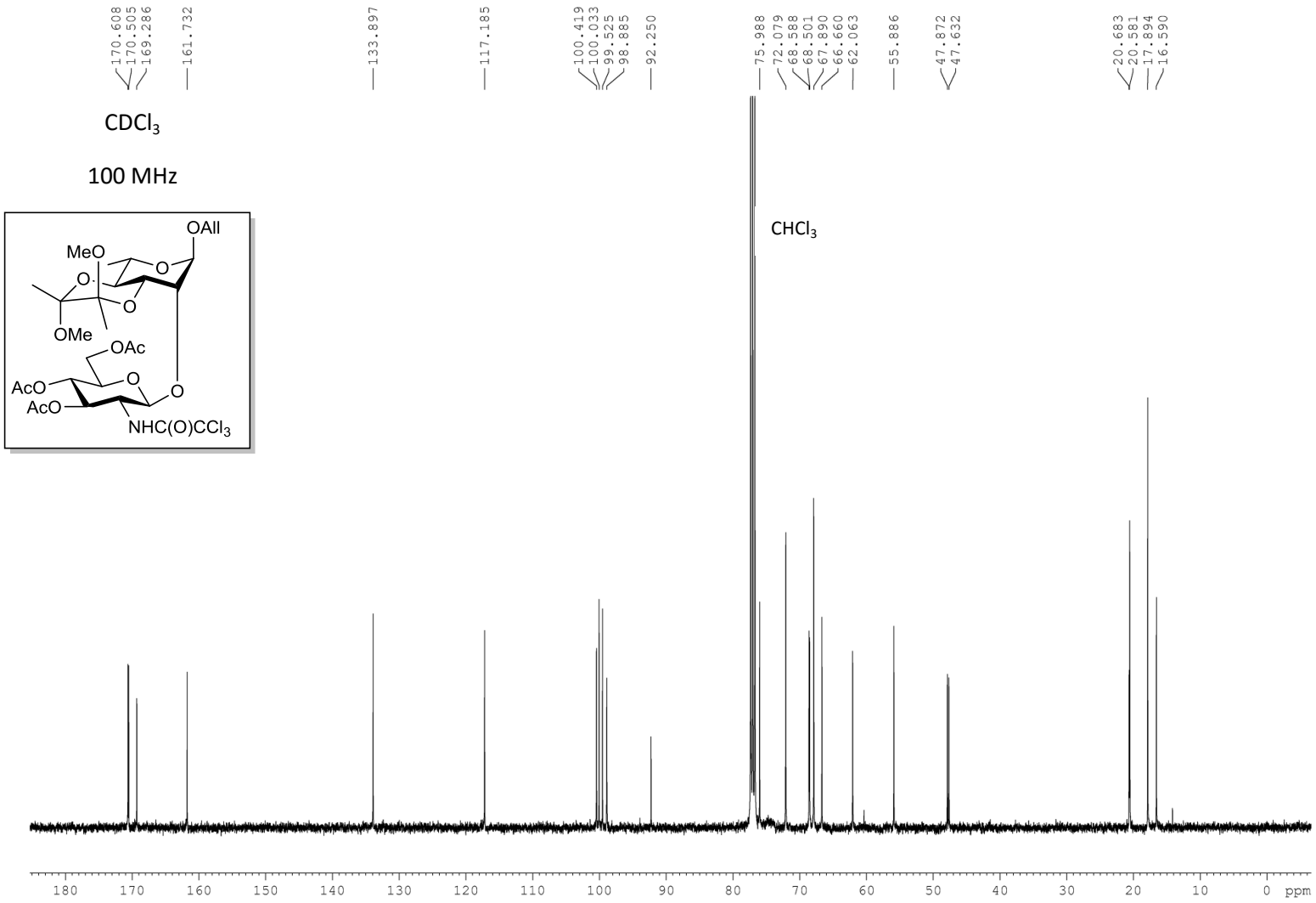


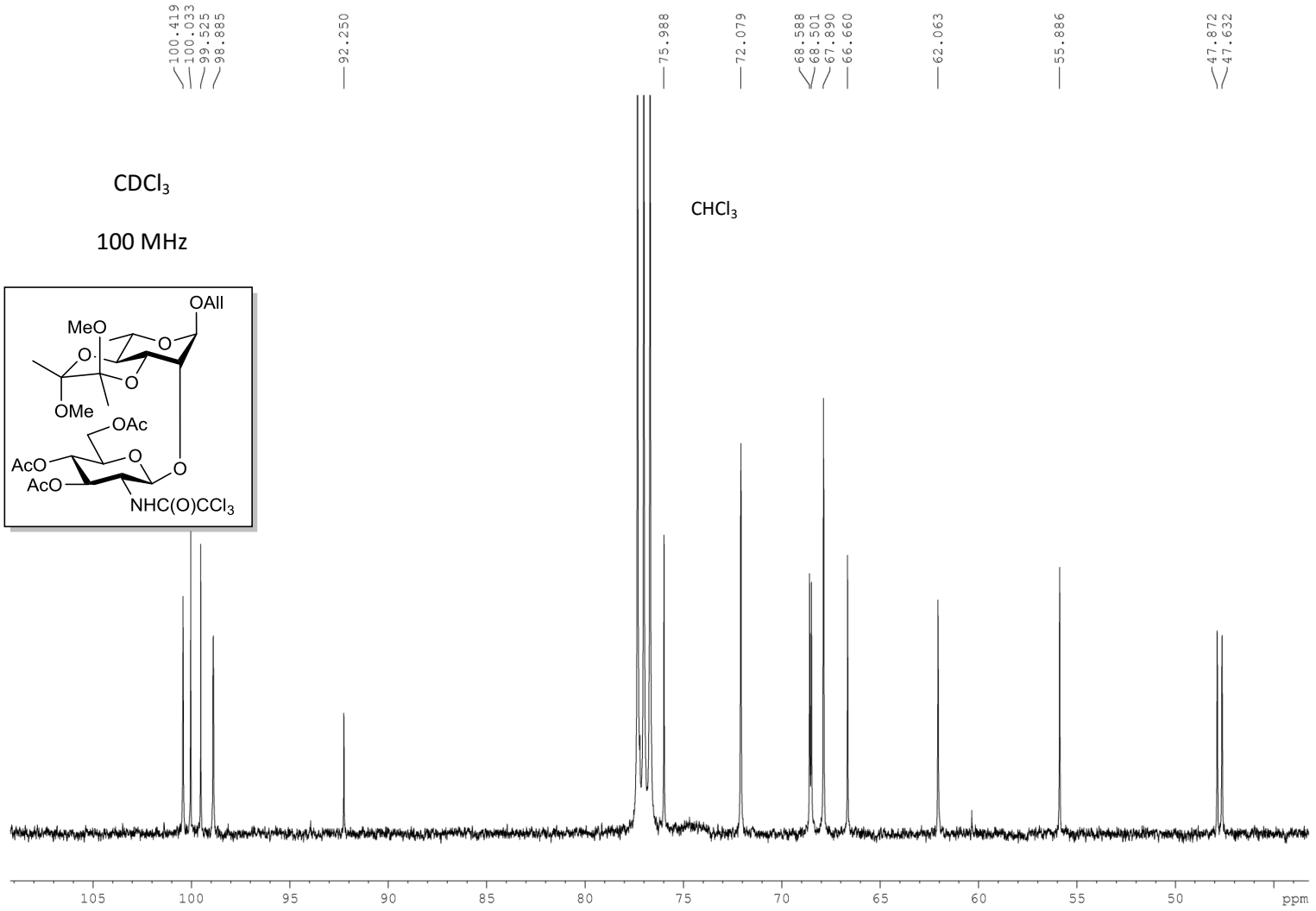


Compound **6**: Allyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl-(1→2)-3,4-*O*-(2',3'-dimethoxybutane-2',3'-diyl)- α-L-rhamnopyranoside

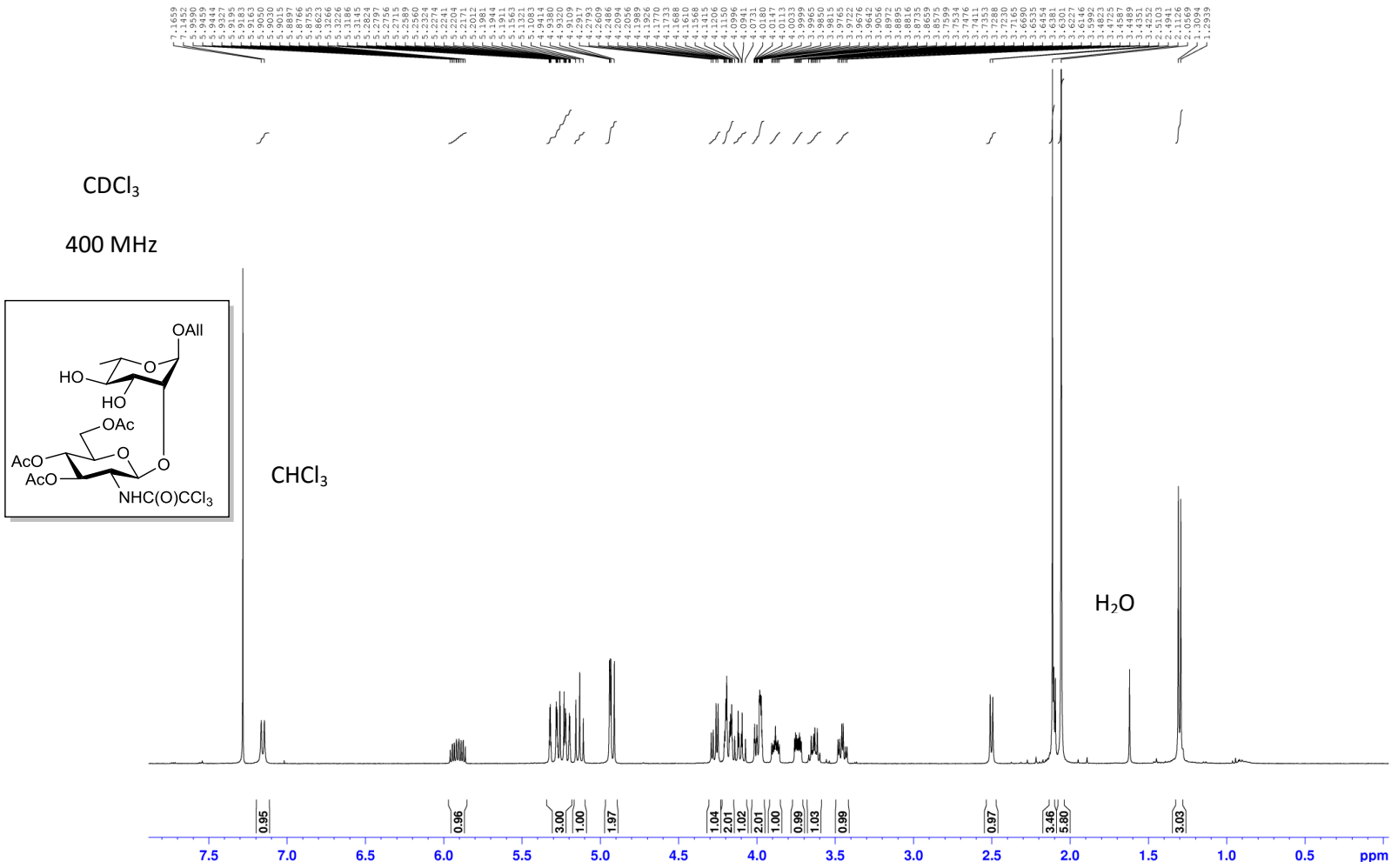


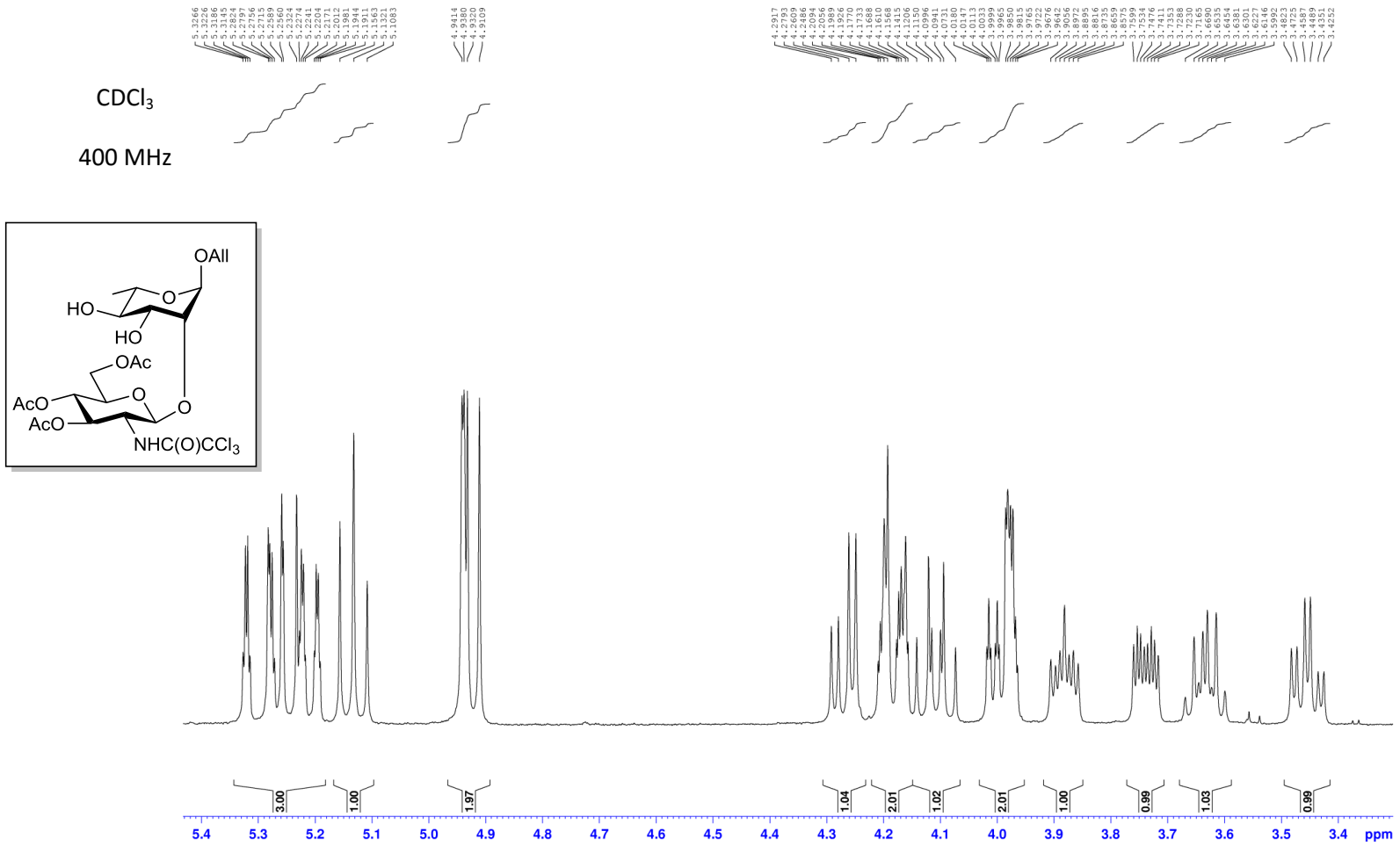

CDCl3
400 MHz
OAll
MeO
OMe
OAc
AcO
AcO
NHC(O)CCl3
2.02
1.03
1.04
1.01
1.00
1.06
2.38
4.17
2.07
1.03
5.4 5.3 5.2 5.1 5.0 4.9 4.8 4.7 4.6 4.5 4.4 4.3 4.2 4.1 4.0 3.9 3.8 3.7 3.6 3.5 ppm

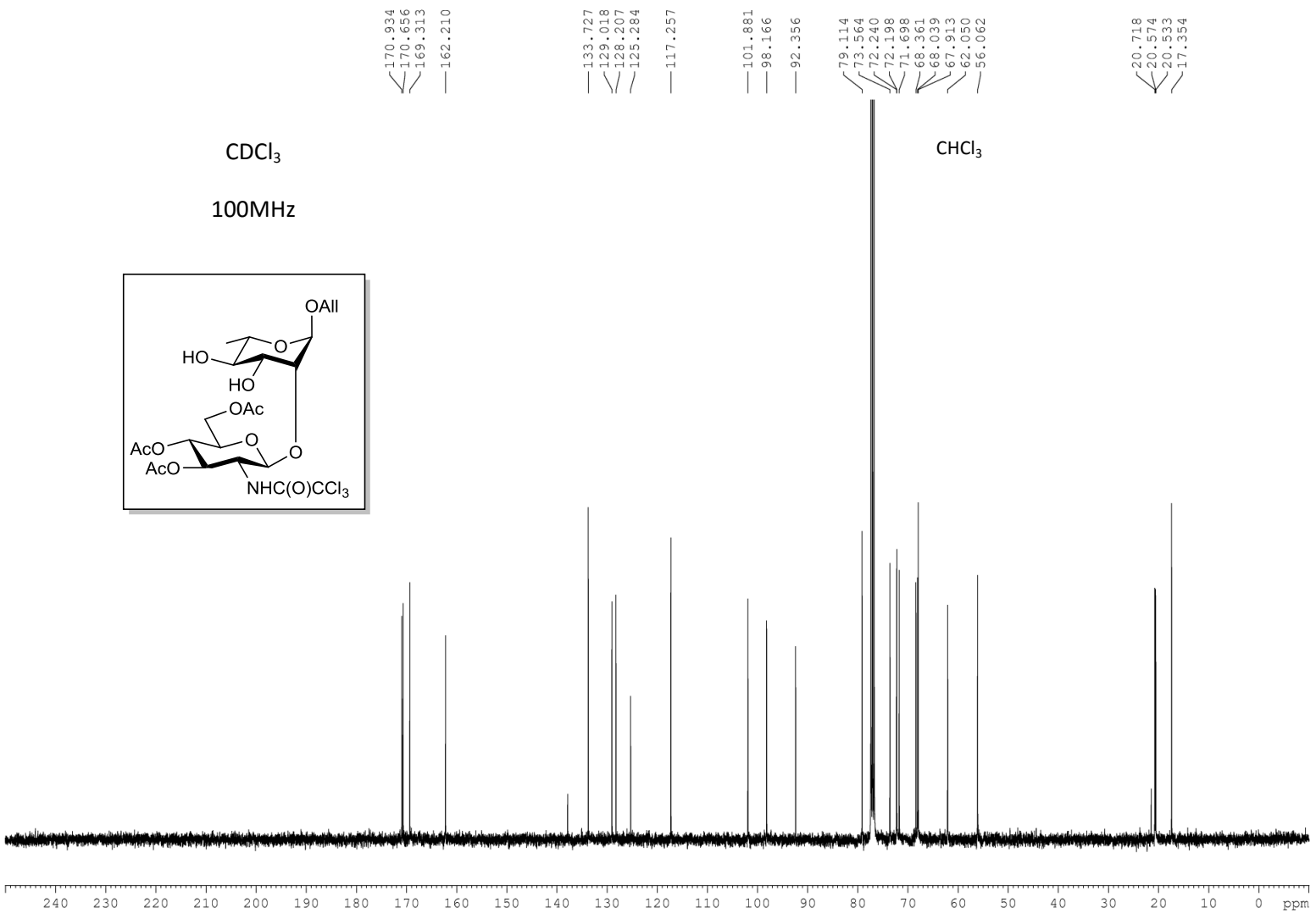

170.608
170.505
169.286
161.732
133.897
117.185
100.419
100.033
99.525
98.885
92.250
75.988
72.079
68.588
68.501
67.890
66.660
62.063
55.886
47.872
47.632
20.683
20.581
17.894
16.590
CDCl3
100 MHz
CHCl3
OAll
MeO
O
O
O
OMe
OAc
O
O
AcO
AcO
NHC(O)CCl3
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
ppm

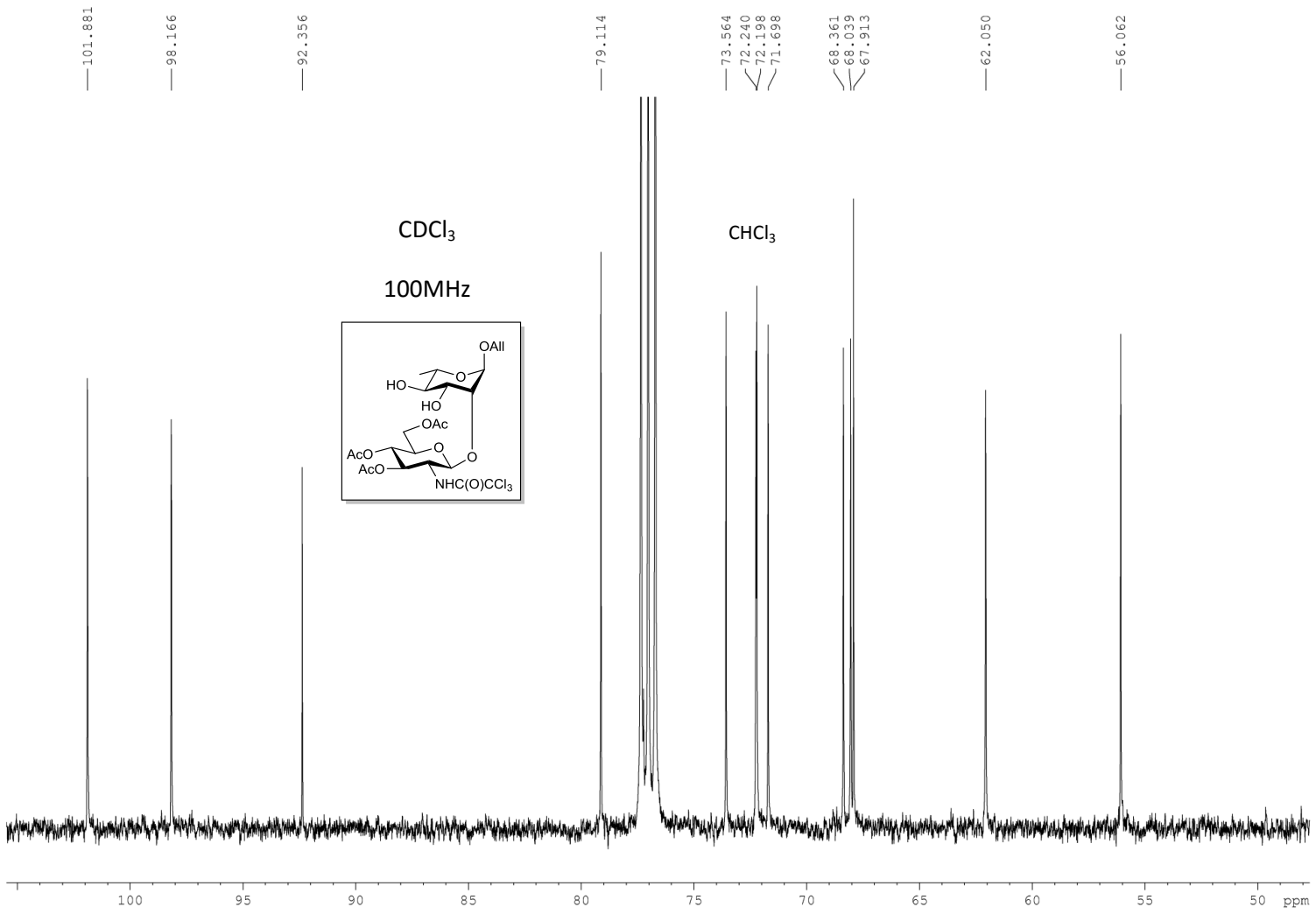

100.419
100.033
99.525
98.885
92.250
75.988
72.079
68.588
68.501
67.890
66.660
62.063
55.886
47.872
47.632
CDCl3
100 MHz
CHCl3
OAll
MeO
O
O
O
OMe
OAc
AcO
O
O
AcO
NHC(O)CCl3
105
100
95
90
85
80
75
70
65
60
55
50
ppm

Compound **7**: Allyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside

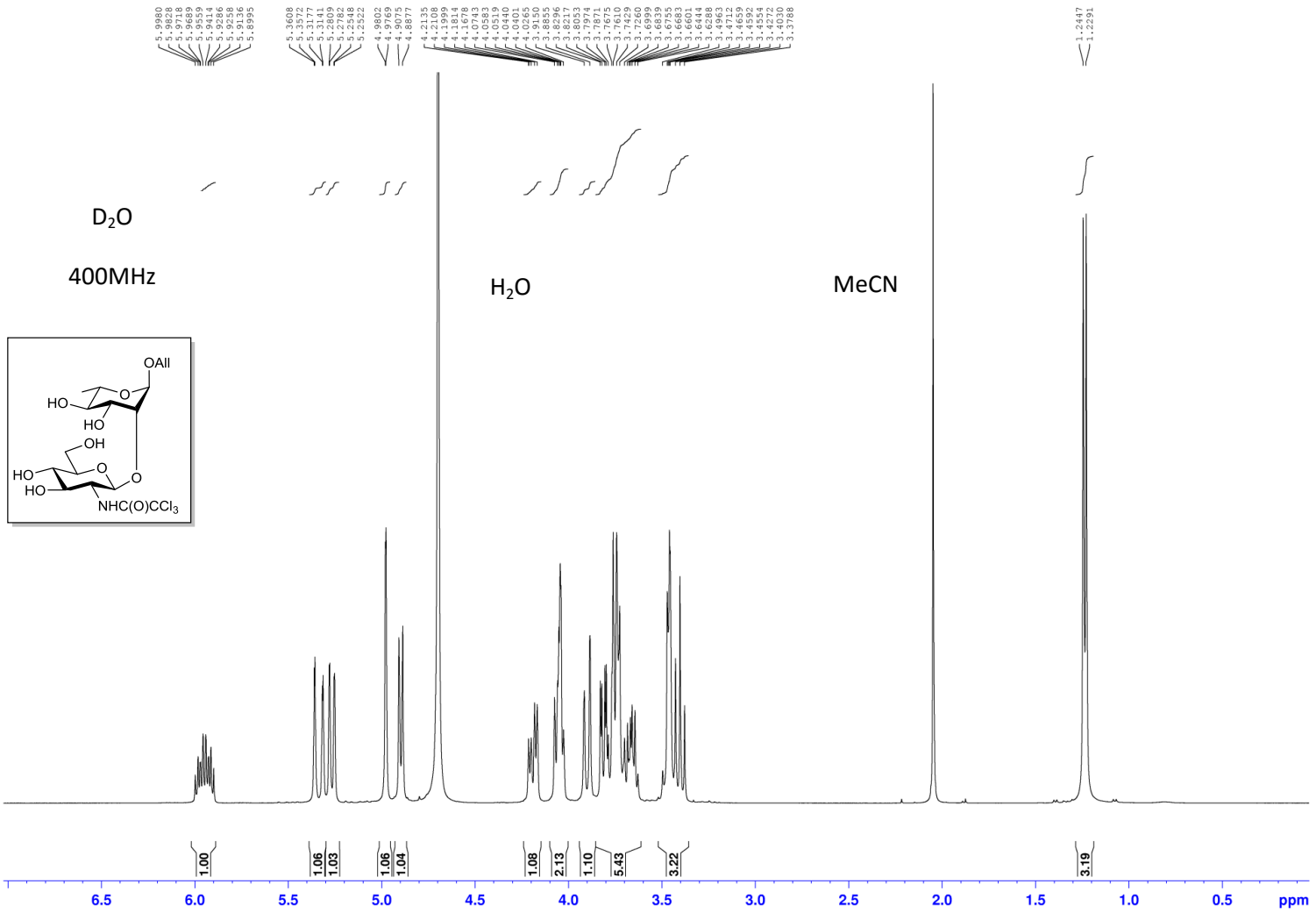


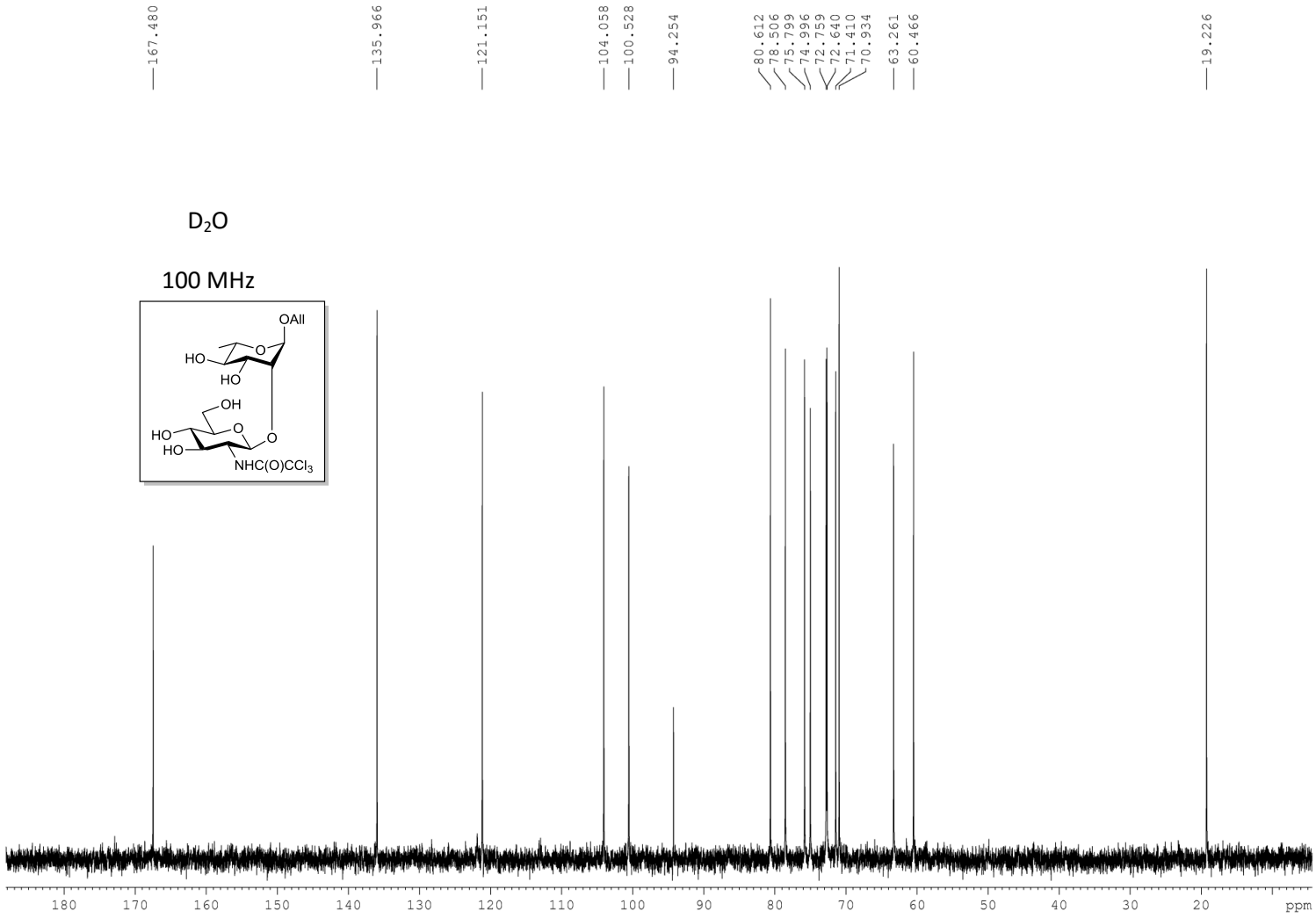

CDCl3
400 MHz
OAll
HO
O
HO
OAc
AcO
O
O
AcO
NHC(O)CCl3
3.00
1.00
1.97
1.04
2.01
1.02
2.01
1.00
0.99
1.03
0.99
5.4
5.3
5.2
5.1
5.0
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
3.9
3.8
3.7
3.6
3.5
3.4
ppm


170.934
170.656
169.313
162.210
133.727
129.018
128.207
125.284
117.257
101.881
98.166
92.356
79.114
73.564
72.240
72.198
71.698
68.361
68.039
67.913
62.050
56.062
20.718
20.574
20.533
17.354
CDCl3
100MHz
CHCl3
OAll
HO
HO
O
OAc
AcO
AcO
O
O
NHC(O)CCl3
240 230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

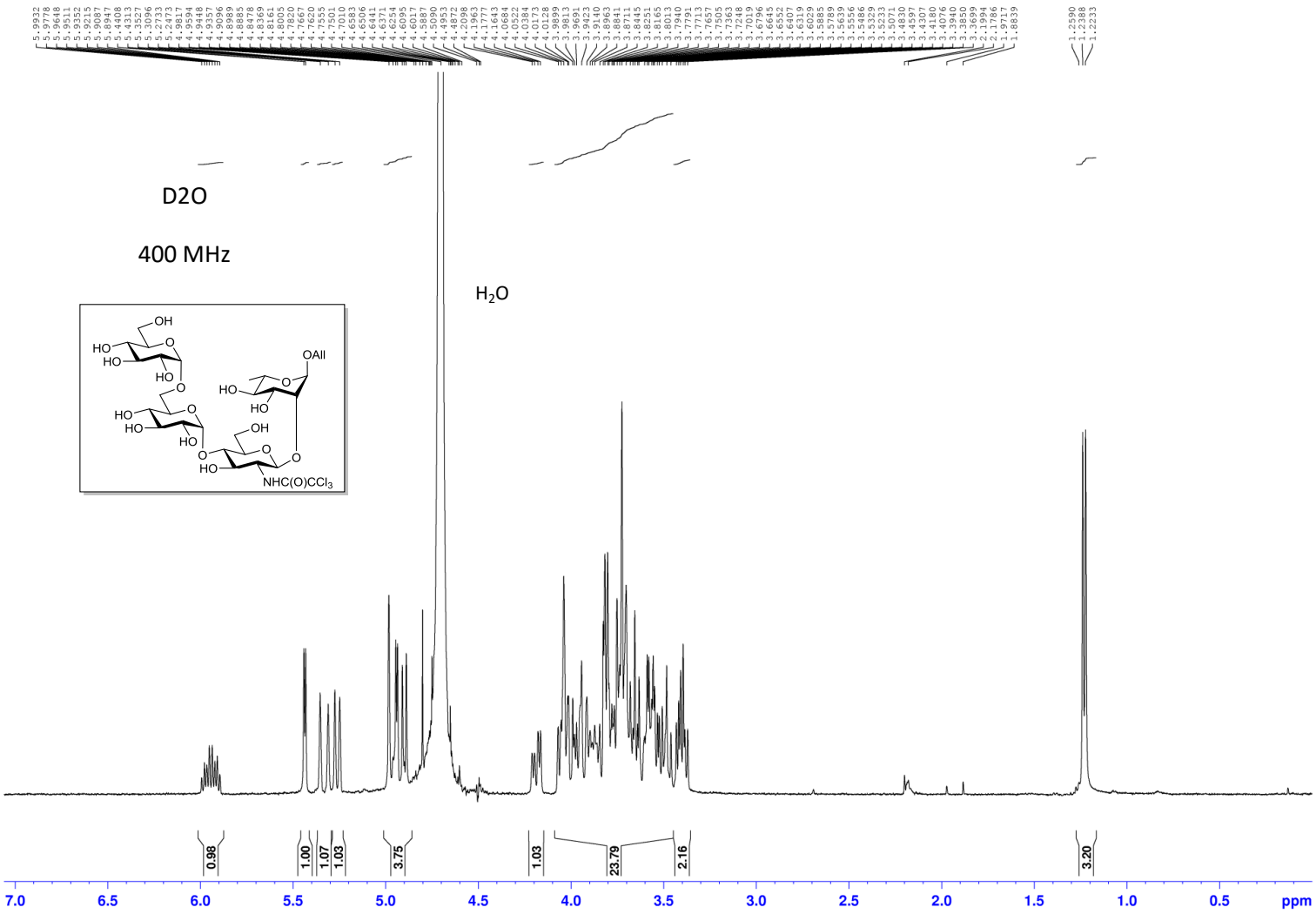

101.881
98.166
92.356
79.114
73.564
72.240
72.198
71.698
68.361
68.039
67.913
62.050
56.062
CDCl3
100MHz
CHCl3
OAll
HO
HO
OAc
AcO
AcO
NHC(O)CCl3
100
95
90
85
80
75
70
65
60
55
50
ppm

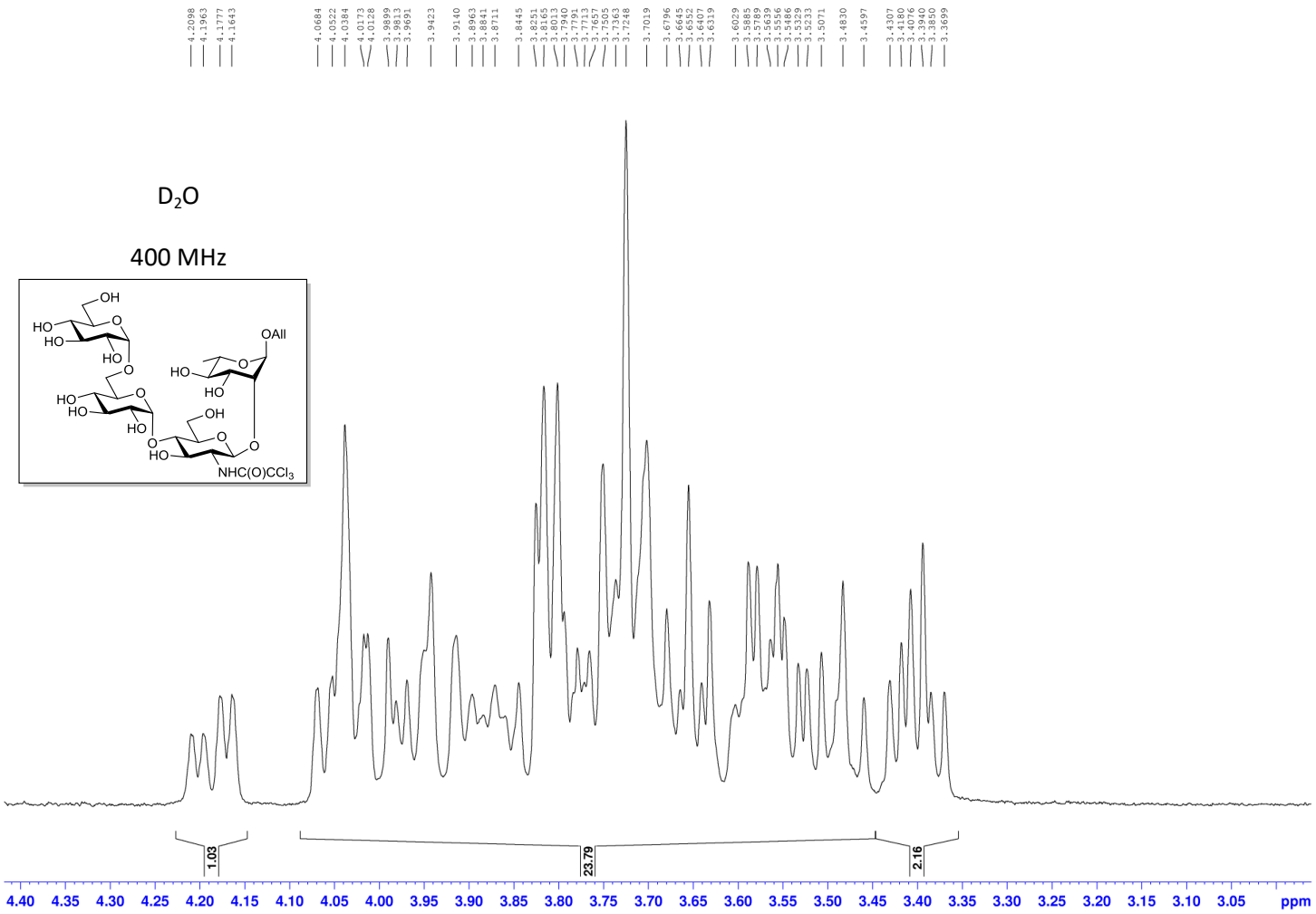
Compound **1**: Allyl 2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside

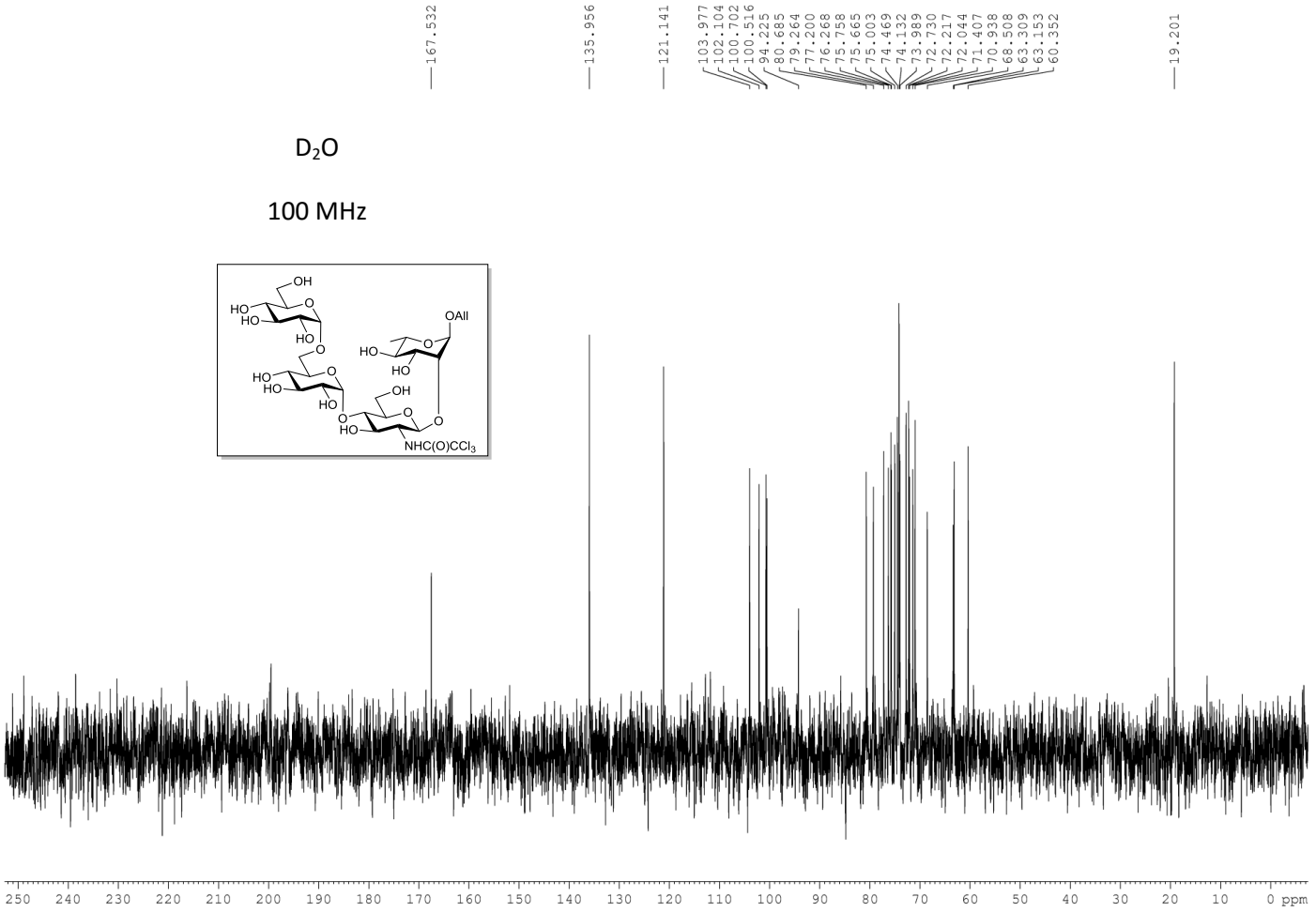



167.480
135.966
121.151
104.058
100.528
94.254
80.612
78.506
75.799
74.996
72.759
72.640
71.410
70.934
63.261
60.466
19.226
D2O
100 MHz
OAll
HO
HO
O
OH
HO
HO
O
O
NHC(O)CCl3
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
ppm

Compound **8**: Allyl α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→4)-2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside (**EED'A'**)




4.2098
4.1963
4.1777
4.1643
4.0684
4.0522
4.0384
4.0173
4.0128
3.9899
3.9813
3.9691
3.9423
3.9140
3.8963
3.8841
3.8711
3.8445
3.8251
3.8165
3.8013
3.7940
3.7791
3.7713
3.7657
3.7505
3.7363
3.7248
3.7019
3.6796
3.6655
3.6552
3.6407
3.6319
3.6029
3.5885
3.5789
3.5689
3.5556
3.5486
3.5329
3.5233
3.5071
3.4830
3.4597
3.4307
3.4180
3.4076
3.3940
3.3805
3.3699
D2O
400 MHz
OAll
NHC(O)CCl3
1.03
23.79
2.16
4.40 4.35 4.30 4.25 4.20 4.15 4.10 4.05 4.00 3.95 3.90 3.85 3.80 3.75 3.70 3.65 3.60 3.55 3.50 3.45 3.40 3.35 3.30 3.25 3.20 3.15 3.10 3.05 ppm


167.532
135.956
121.141
103.977
102.104
100.702
100.516
94.225
80.685
79.264
77.200
76.268
75.758
75.665
75.003
74.469
74.132
73.989
72.730
72.217
72.044
71.407
70.938
68.508
63.309
63.153
60.352
19.201
D2O
100 MHz
OAll
NHC(O)CCl3
250 240 230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

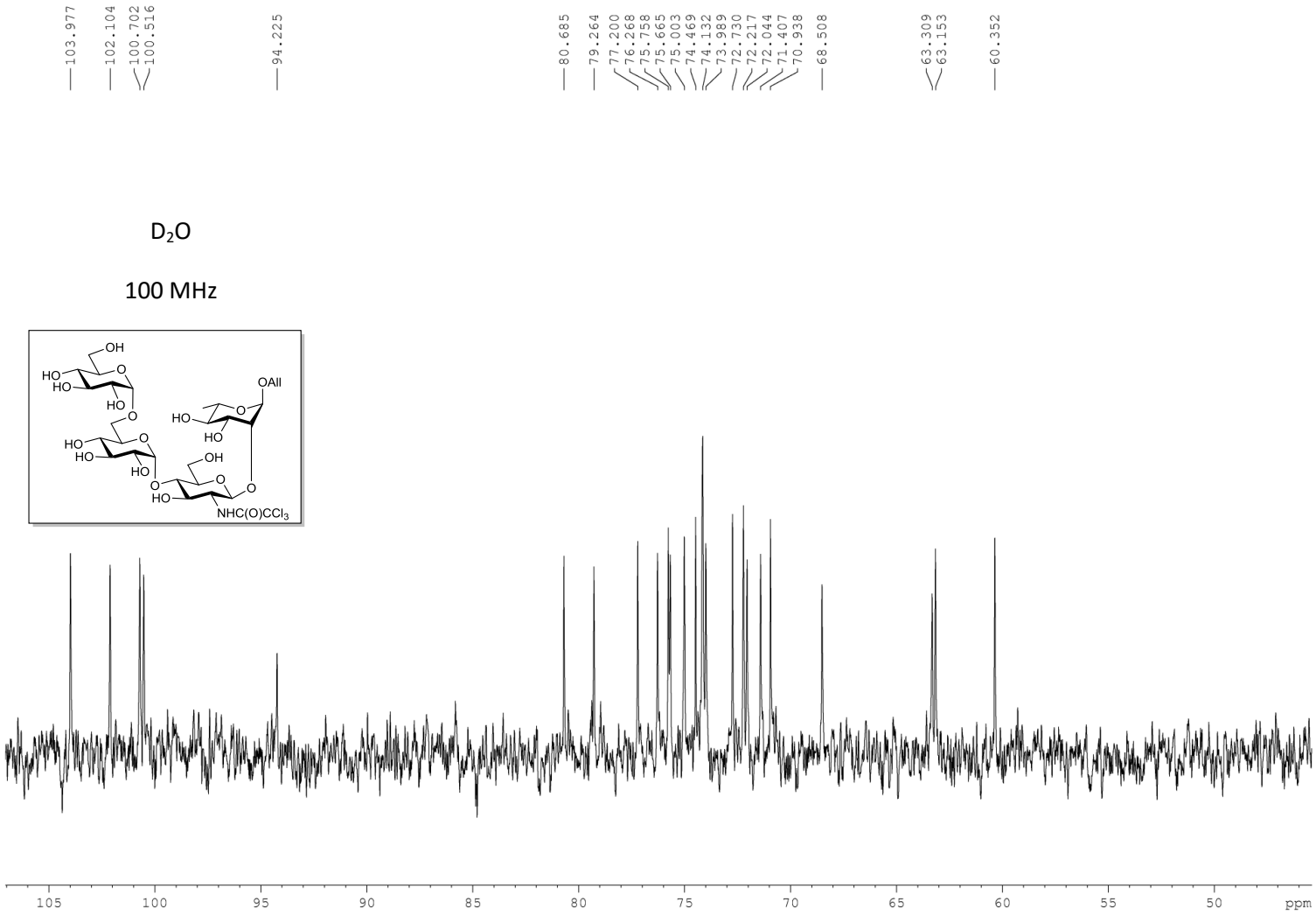

103.977
102.104
100.702
100.516
94.225
80.685
79.264
77.200
76.268
75.758
75.665
75.003
74.469
74.132
73.989
72.730
72.217
72.044
71.407
70.938
68.508
63.309
63.153
60.352
D2O
100 MHz
OAll
NHC(O)CCl3
105
100
95
90
85
80
75
70
65
60
55
50
ppm

Table S1. Primers used for the construction of the 3 sequence libraries (libraries 1, 2, 3)

Library 1	Primer sequences
226-229_for	AC CAA TAC GAC CGC ACC CTG AAA RAA CAT GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA CAT GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA TAT GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA TAT GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA CAT TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA CAT TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA TAT TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA TAT TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
226-229_rev	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC ATG TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC ATG TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC ATA TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CG G GTG CTG GTC GGG GKC ATA TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA ATG TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA ATG TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA ATA TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA ATA TTY CAG CAG GGT GCG GTC GTA TTG GT
289, 290, 300_for	C TTG GGC GTT GAC ATC CTG CGT ATG GAT GCG GTT AGC CAT ATT TGG AAA CAA ATG GGG ACA AGC TGC RTA AAC CTG CCG CAG GCG CAC GCC CTC ATC
289, 290, 300_rev	GAT GAG GGC GTG CGC CTG CGG CAG GTT TAY GCA GCT TGT CCC CAT TTG TTT CCA AAT ATG GCT AAC CGC ATC CAT ACG CAG GAT GTC AAC GCC CAA G
331-332_for	C GCC GTG TTC TTC AAA TCC GAA GCC ATC ACC CAC CCC GAC CAA GTC GTC CAA TAC ATC G
	C GCC GTG TTC TTC AAA TCC GAA GCC ATC GGC CAC CCC GAC CAA GTC GTC CAA TAC ATC G
331-332_rev	C GAT GTA TTG GAC GAC TTG GTC GGG GTG GGT GAT GGC TTC GGA TTT GAA GAA CAC GGC G
	C GAT GTA TTG GAC GAC TTG GTC GGG GTG GCC GAT GGC TTC GGA TTT GAA GAA CAC GGC G
396-398_for	AC TAC GTC CGC AGC CAC GAC GAC ATC RGC TGG GTG TTT GCC GAT GAA GAC GCG GCA TAT
396-398_rev	ATA TGC CGC GTC TTC ATC GGC AAA CAC CCA GCY GAT GTC GTC GTG GCT GCG GAC GTA GT
434-448_for	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
434-448_rev	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC

Library 2	Primer sequences
226-229_for	AC CAA TAC GAC CGC ACC CTG AAA RAA TAT GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA TAT GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA GTG GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA GTG GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA TAT TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA TAT TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA GTG TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA GTG TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
226-229_rev	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC ATA TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CG G GTG CTG GTC GGG GKC ATA TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC CAC TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC CAC TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA ATA TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA ATA TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA CAC TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA CAC TTY CAG CAG GGT GCG GTC GTA TTG GT
289, 209, 300_for	C TTG GGC GTT GAC ATC CTG CGT ATG GAT GCG GTT GCG TAT ATT TGG AAA CAA ATG GGG ACA AGC TGC GTG AAC CTG CCG CAG GCG CAC GCC CTC ATC
289, 290, 300_rev	GAT GAG GGC GTG CGC CTG CGG CAG GTT CAC GCA GCT TGT CCC CAT TTG TTT CCA AAT ATA CGC AAC CGC ATC CAT ACG CAG GAT GTC AAC GCC CAA G
331-332_for	C GCC GTG TTC TTC AAA TCC GAA GCC ATC ACC CAC CCC GAC CAA GTC GTC CAA TAC ATC G
	C GCC GTG TTC TTC AAA TCC GAA GCC ATC GGC CAC CCC GAC CAA GTC GTC CAA TAC ATC G
331-332_rev	C GAT GTA TTG GAC GAC TTG GTC GGG GTG GGT GAT GGC TTC GGA TTT GAA GAA CAC GGC G
	C GAT GTA TTG GAC GAC TTG GTC GGG GTG GCC GAT GGC TTC GGA TTT GAA GAA CAC GGC G
396-398_for	AC TAC GTC CGC AGC CAC GAC GAC ATC RGC TGG GTG TTT GCC GAT GAA GAC GCG GCA TAT
396-398_rev	ATA TGC CGC GTC TTC ATC GGC AAA CAC CCA GCY GAT GTC GTC GTG GCT GCG GAC GTA GT
434-448_for	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
434-448_rev	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC

Library 3	Primer sequences
226-229_for	AC CAA TAC GAC CGC ACC CTG AAA RAA TAT GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA TAT GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA GTG GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA GTG GMC CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA TAT TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA TAT TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG AAA RAA GTG TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
	AC CAA TAC GAC CGC ACC CTG CTG RAA GTG TTT CCC GAC CAG CAC CCG GGC GGC TTC TCG CAA
226-229_rev	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC ATA TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CG G GTG CTG GTC GGG GKC ATA TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC CAC TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG GKC CAC TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA ATA TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA ATA TTY CAG CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA CAC TTY TTT CAG GGT GCG GTC GTA TTG GT
	TTG CGA GAA GCC GCC CGG GTG CTG GTC GGG AAA CAC TTY CAG CAG GGT GCG GTC GTA TTG GT
289, 290, 300_for	C TTG GGC GTT GAC ATC CTG CGT ATG GAT GCG GTT ATT TAT ATT TGG AAA CAA ATG GGG ACA AGC TGC ATT AAC CTG CCG CAG GCG CAC GCC CTC ATC
289, 290, 300_rev	GAT GAG GGC GTG CGC CTG CGG CAG GTT AAT GCA GCT TGT CCC CAT TTG TTT CCA AAT ATA AAT AAC CGC ATC CAT ACG CAG GAT GTC AAC GCC CAA G
331-332_for	C GCC GTG TTC TTC AAA TCC GAA GCC ATC ACC CAC CCC GAC CAA GTC GTC CAA TAC ATC G
	C GCC GTG TTC TTC AAA TCC GAA GCC ATC GGC CAC CCC GAC CAA GTC GTC CAA TAC ATC G
331-332_rev	C GAT GTA TTG GAC GAC TTG GTC GGG GTG GGT GAT GGC TTC GGA TTT GAA GAA CAC GGC G
	C GAT GTA TTG GAC GAC TTG GTC GGG GTG GCC GAT GGC TTC GGA TTT GAA GAA CAC GGC G
396-398_for	AC TAC GTC CGC AGC CAC GAC GAC ATC RGC TGG GTG TTT GCC GAT GAA GAC GCG GCA TAT
396-398_rev	ATA TGC CGC GTC TTC ATC GGC AAA CAC CCA GCY GAT GTC GTC GTG GCT GCG GAC GTA GT
434-448_for	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG KCA GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG TYC AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC CGT CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
	GT TTC GAC GGC AGC TTC GCT CGT GGC GTG CCG GAT AGK TAC GAT CCA AGC ACA GGC GAK GCG CGT GTC AGT GGT ACA GCC GCG GCA TTG GT
434-448_rev	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC TGM CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT GRA CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG ACG GCC ACG AGC GAA GCT GCC GTC GAA AC
	AC CAA TGC CGC GGC TGT ACC ACT GAC ACG CGC MTC GCC TGT GCT TGG ATC GTA MCT ATC CGG CAC GCC ACG AGC GAA GCT GCC GTC GAA AC

Table S2. sequencing data of mutants from libraries 1 and 2.

Library 1	Mutant
Amino acid positions	22
226	L
227	E
228	V
229	F
289	A
290	Y
300	V
331	T
332	H
396	G
397	W
398	T
434	R
435	P
436	S
437	R
438	Y
439	D
444	E
445	A
446	S
447	V
448	S

Library 2	Mutants																							
Amino acid positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	87	88	89	90	91
226	K	L	K	L	K	L	L	K	K	L	L	L	L	L	L	L	L	K	K	L	K	L	K	L
227	E	E	E	E	E	E	K	K	E	K	K	E	E	K	E	E	E	K	E	E	E	K	E	E
228	V	V	Y	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	Y	V	H	V	V
229	F	F	F	A	F	F	F	D	F	F	F	F	F	F	F	F	D	F	F	D	F	F	F	F
289	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	S	A	S	A	A
290	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	H	Y	H	F	F
300	V	V	V	V	E	V	V	V	E	V	V	V	V	V	V	V	I	V	V	V	V	V	E	V
331	T	G	G	T	G	T	V	T	T	V	G	T	T	T	G	T	G	T	T	V	G	T	V	T
332	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
396	G	G	S	G	G	S	G	G	S	G	S	G	S	S	S	S	S	G	S	S	G	G	G	G
397	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
398	T	T	V	T	T	V	T	T	V	T	V	V	V	T	V	V	V	T	V	V	T	T	V	T
434	V	V	R	V	V	V	V	V	V	V	V	V	V	V	V	V	R	V	V	V	R	R	V	V
435	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
436	F	F	D	D	F	F	F	F	F	F	D	F	F	F	F	F	D	F	F	F	D	S	F	F
437	Q	Q	Q	S	Q	Q	Q	Q	S	Q	R	Q	R	S	Q	S	R	S	S	S	S	R	S	Q
438	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
439	N	N	D	D	N	N	N	N	D	N	D	N	D	D	N	D	D	D	D	D	D	D	D	N
444	D	D	D	D	D	D	D	D	E	D	D	D	D	E	D	E	E	D	D	E	D	E	D	D
445	C	A	A	R	C	C	C	C	A	C	A	C	A	A	C	A	A	A	A	R	A	A	A	C
446	R	R	S	R	R	R	R	R	R	R	S	R	R	R	R	R	R	R	R	R	R	S	R	R
447	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
448	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S

Library 3	Mutants																												
Amino acid positions	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
226	K	K	K	K	L	L	K	K	L	L	L	K	K	K	K	K	K	K	K	L	K	K	K	L	L	L	L	L	L	L
227	E	E	E	K	K	K	K	E	E	K	K	E	K	K	E	E	E	E	K	E	K	E	E	E	E	E	E	E	E	K
228	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	I	V	V	Y	I	V	V
229	F	F	F	F	F	D	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	A	A	D	C	F	A
289	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
290	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
300	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
331	T	G	G	T	G	T	T	T	T	T	T	V	V	T	T	T	T	T	T	T	V	T	T	T	T	T	T	T	T	G
332	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
396	G	G	G	G	G	G	G	G	S	G	G	G	G	G	S	G	S	G	S	S	G	G	G	S	G	G	S	G	S	S
397	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
398	V	V	T	T	T	V	T	T	V	T	T	T	T	T	V	V	V	T	T	V	T	T	T	V	T	T	V	T	V	V
434	R	V	V	V	V	R	R	V	V	V	V	V	V	V	V	V	V	V	V	V	V	R	V	V	V	V	V	V	V	R
435	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
436	F	F	F	F	D	F	D	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	D	F	F	F	D	F	S
437	S	R	S	Q	R	R	S	S	Q	S	S	Q	Q	Q	R	R	S	R	S	R	Q	R	S	R	Q	Q	R	R	R	R
438	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
439	D	D	D	N	D	D	D	D	N	D	D	N	N	N	D	D	D	D	D	D	N	D	D	D	N	N	D	D	D	D
444	D	D	E	D	D	D	D	D	D	D	D	D	D	D	D	D	D	E	E	D	D	D	E	D	D	D	D	D	E	E
445	A	A	A	C	R	A	R	A	C	A	A	C	C	C	C	A	R	A	A	R	C	A	A	R	C	C	R	R	A	A
446	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	S	R	S	R	R	R	R	S	R	R	R	R	R	R	D
447	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	R
448	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	Q

Computational Modelling and Design. Three-dimensional model of *Np*AS in complex with the desired glucosylated product (ED'A') bound into the enzyme active site was constructed starting from the crystallographic complex with maltoheptaose, a reaction product (PDB ID: 1MW0).[S1] ED'A' was manually docked into the active site of the amylosucrase by superimposing the glucosyl unit (E) of ED'A' onto the glucosyl unit of the crystallographic maltoheptaose located at subsite −1, using the DISCOVER module of InsightII software suite (Accelrys, San Diego, CA, USA). Partial charges were taken from ED'A' calculated using the RESP method [S2] at the HF/6-31G* level of theory using Gaussian03 package (Gaussian, Inc., Wallingford CT, USA). Protein hydrogen atoms were added using the *tleap* module of the Amber 9 software package. [S3] Histidine protonation states were assigned using the *tleap* module. The inactivating Glutamine mutation introduced at position 328 in the crystallographic structure was reverted to the natural catalytic acid/base amino-acid, the Glutamate. This E328 residue was modeled in its protonated form (i.e. as the conjugated acid). Parameters for non-standard atom groups were generated with the *Antechamber* module. The molecular all-atom *ff99SB* [S4], *Glycam06* [S5] and *gaff* force fields [S6] were used for the proteins, carbohydrates and other non-standard groups, respectively. The molecular system was then subjected to 6 cycles of minimizations (each consisting of 500 steps of steepest descent followed by 500 steps of conjugated gradient) using the *Sander* module of Amber 9. Atomic positions of the protein backbone and ED'A' heavy atoms were first restrained using a harmonic potential during the minimization schedule. The force constant was then progressively diminished until a final unrestrained minimization cycle. Distance and dihedral constraints applied between the side chains of the catalytic residues, E328 and D286 and the glycosidic O1 and the anomeric C1 of the glucose from ED'A' were maintained until the end of minimization procedure. Visual inspection of the modelled complex, as well as the crystallographic complex of the amylosucrase with the donor substrate, the sucrose (PDB: 1JGI), was carried out using PyMOL software (Schrödinger, Portland, OR, USA). Based on the analysis of these minimized complexes and our knowledge on amino acid residues involved in catalysis and/or specific recognition of sucrose, 23 positions (226-229, 289-300, 331,332, 396-398, 434-439, 444-448) of the active site were selected to be sampled for mutations. Version 3.2 of the Rosetta software suite (Rosetta Commons, UW techTransfer, Seattle, WA, USA) was used to redesign the *Np*AS active site in order to enhance the interaction with ED'A'. A protocol was used that tests mutations and side-chain conformations at the 23 selected positions and samples variations in the transition and rotation of the docked ED'A'. All amino acid types, except Cysteine, were authorized at the 23 designable positions. 60 residues surrounding the 23 designable residues were allowed to relax toward alternate side-chain rotamers. These 60 residues were selected as follows: if they had their Cα within a distance of 10.0 Å of any ED'A' heavy atoms or when they had their Cα within a distance of 12.0 Å of any ED'A' heavy atoms and their Cβ closer to ligand atoms. Enzyme backbone atoms, non-designed or non-repacked side-chains as well as the internal conformation of ED'A' were held fixed. The initial ED'A' conformation from which perturbations were made was taken from the ED'A'-enzyme complex model minimized with Amber 9. The position of the ED'A' is perturbed up to 1.0 Angstrom by random translations of mean 0.1 Angstroms in each direction. The molfile_to_parmas.py script was used to parameterize ED'A' to be conform to RosettaDesign. Catalytic constraints (i.e. a set of distances, angles and dihedrals) between D286 and E328 catalytic residues and ED'A' were predefined and used in enzyme design calculations to penalize non-productive conformations of the catalytic residues. The enzyme design protocol used comprises three stages: (1) a step of catalytic interactions optimization during which all non-active site residues were mutated to alanine and a gradient-based minimization was performed in order to move the substrate to a position where the catalytic interactions are as ideal as possible; (2) several cycles of Monte Carlo-

based sequence design and minimization by applying catalytic constraints; (3) a final sequence rotamer repacking and minimization step without catalytic constraints. This design protocol was run 20,000 independent times. The output sequences were filtered based on the following criteria: total score < -1330 (lower than the score of the undesigned scaffold), total constraint score of the catalytic residues <1.0 and ligand binding energy < -6.0. The 1,515 unique sequences resulting from this filtering were used to guide library design. The library design strategy developed was based on: (1) the determination from a multiple sequence alignment of 1,515 designed sequences of the position-dependent amino acid residue frequency, a Correlated Mutation (CM) score at position pairs and the residue-pair frequency; (2) the calculation of the Shannon entropy (H_X) [S7] using in house software, called Sequester, to measure the position-dependent amino acid residue variability in multiple sequence alignment of GH13_4 sub-family sequences.

The Correlated mutation (CM) score at (X,Y) positions is computed as shown in Equation 1 :

$$CM_{(X,Y)} = \sum_{i=1}^{20} \sum_{j=1}^{20} \left| F_{(i,j,X,Y)} \right| \cdot N_{(i,X)\,|\,(j,Y)}$$

$$\text{where} \quad F_{(i,j,X,Y)} = \frac{N_{(i,X)\,|\,(j,Y)}}{N_{(j,Y)}} - \frac{N_{(i,X)}}{N_{(Total)}}$$

$N_{(i,X)\,|\,(j,Y)}$ = No of sequences with residue type (i) at position (X) in subset with residue type (j) at position (Y)

$N_{(i,X)}$ = No of sequences with residue type (i) at position (X)

$N_{(j,Y)}$ = No of sequences with residue type (j) at position (Y)

$N_{(Total)}$ = Total no of sequences in MSA

Normalise $CM_{(X,Y)}$ scores wrt to maximum value

The Shannon entropy (H_X) at residue alignment position (X), corrected for the normalized frequency of residue type occurrence, is computed as shown in Equation 2:

$$H_X = \frac{\sum_{i=1}^{20} p_{(i \mid X)} \ln p_{(i \mid X)}}{\sum_{i=1}^{20} p_{(i)} \ln p_{(i)}}$$

where

$p_{(i \mid X)}$ Conditional probability of amino acid residue type (i) at position (X)

$p_{(i)}$ Probability of amino acid residue type (i) at any position (in all proteins)

To minimize sampling bias, normalized residue type probability values were taken as those documented by Ranganathan and co-workers [S8], garnered from sequence data for all natural proteins. Values of H_X lie in the range 0–1; a zero value corresponds to a fully conserved residue position, and a value of unity represents a distribution in which each residue type has an equal chance of occurrence. From these computations, the amino acid types and the combinations of amino acid residues considered in the final designed libraries were selected as follows : (1) at each enzyme designed position, the wild-type amino acid residues with a frequency of occurrence ≥ 5% were maintained; (2) the wild-type amino acid residues with a frequency of occurrence < 5% or absent in the design results and with a Shannon entropy < 0.2% were introduced; (3) the other amino acid types were kept if only their position-dependent frequency was higher than 10% (3) for correlated amino acid pairs, only the combination-pairs with a frequency of occurrence ≥ 5% were preserved while all combinations were considered for non-correlated amino-acid residues.The mutant F3 of *Np*AS was constructed starting from the model of the wild-type enzyme docked with ED'A' in which mutations were introduced using the Biopolymer module of the InsightII software package (Accelrys, San Diego, CA, USA). The conformation of the mutated residue side chain was optimized by manually selecting a low-energy conformation from a side-chain rotamer library. Steric clashes (van der Waals overlap) and nonbonded interaction energies (Coulombic and Lennard-Jones) were evaluated for the different side-chain conformations. The model was subsequently minimized following the same energy minimization procedure described above.

General conditions for the chemical synthesis of disaccharide 1 (D'A'). Anhydrous (anhyd.) solvents – including dichloromethane (DCM)– were delivered on molecular sieves and used as received. Additional solvents cited in the text are abbreviated as MeCN (acetonitrile), cHex (cyclohexane), and EtOAc (ethyl acetate). Reactions requiring anhydrous conditions were run under an argon (Ar) atmosphere, using dried glassware. 4Å molecular sieves (4 Å MS) were activated before use by heating under high vacuum. Analytical thin-layer chromatography (TLC) was performed with silica gel 60 F_{254}, 0.25 mm pre-coated TLC foil plates. Compounds were visualized using UV_{254} and/or orcinol (1 mg·mL^{-1}) in 10% aq. H_2SO_4 with charring. Flash column chromatography was carried out using silica gel (particle size 40-63 µm). NMR spectra were

recorded at 303 K on a Bruker Avance spectrometer equipped with a BBO probe at 400 MHz (^{1}H) and 100 MHz (^{13}C). Spectra were recorded in deuterated chloroform ($CDCl_3$) and deuterated water (D_2O). Chemical shifts are reported in ppm (δ) relative to residual solvent peak $CHCl_3$ in the case of $CDCl_3$, HOD and DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) in the case of D_2O, at 7.26/77.16, and 4.70/0.00 ppm for the ^{1}H and ^{13}C spectra, respectively. Coupling constants are reported in hertz (Hz). Elucidations of chemical structures were based on ^{1}H, COSY, DEPT-135, HSQC, ^{13}C, and HMBC. Signals are reported as s (singlet), d (doublet), t (triplet), dd (doublet of doublet), q (quadruplet), dq (doublet of quadruplet), m (multiplet), quat (quaternary). The signals can also be described as broad (prefix b), or pseudo (prefix p). Of the two magnetically non-equivalent geminal protons at C-6, the one resonating at lower field is denoted H-6a, and the one at higher field is denoted H-6b. Sugar residues are lettered according to the lettering of the repeating unit of the O-Ag from *S. flexneri* adopted in similar studies [S9] (A' = L-rhamnose, D' = *N*-acetyl-D-glucosamine) and identified by a subscript in the listing of signal assignments. High resolution mass spectrometry (HRMS) spectra were recorded on a WATERS QTOF Micromass instrument in the positive-ion electrospray ionisation (ESI^+) mode. Solutions were prepared using 1:1 $MeCN/H_2O$ containing 0.1% formic acid. Optical rotations were obtained using sodium D line at room temperature (24 °C) on a Bellingham + Stanley Ltd. ADP220 polarimeter. $[\alpha]_D$ values are given in 10^{-1} deg cm^2 g^{-1}.

Bacterial strains, plasmids & chemicals. Plasmid pGST-amylosucrase (pGST-AS) derived from the pGEX-6P-3 (GE Healthcare Biosciences, Piscataway, U.S.A) encoding GST fused to *NpAS* gene, was used for the construction of semi-rational libraries. *E.coli* TOP 10 electrocompetent cells (Invitrogen, Carlsbad, U.S.A.) were used as the host for the plasmid transformation, gene expression, large screening and production of recombinant amylosucrase variants. *Dpn*I, EcoRI HF and NotI HF restriction enzymes, T4 DNA ligase and Antartic Phosphatase were purchased from New England Biolabs (Beverly, MA, U.S.A.). Phusion DNA-polymerase was purchased from Finnzymes (Espoo, Finland). Ampicillin, lysozyme and isopropyl-β-D-thiogalactopyranoside (IPTG) were purchased from Eurodemex (Souffelweyersheim, France); Bromothymol blue was purchased from Sigma – Aldrich (St Louis, Mo, U.S.A.). Oligonucleotides were synthesized by Eurogenetec (Liege, Belgium). DNA plasmid extraction (GenElute HP plasmid Miniprep kit, Qiagen, Chatsworth, CA), DNA purification and gel extraction (GenElute Gel Extraction kit) columns were purchased from Sigma-Aldrich (St Louis, Mo, U.SA.).

Library screening on sucrose in solid media. A pH-based high-throughput screening assay earlier described by Champion *et al.* [S10] was used to isolate sucrose active clones. The library was transformed into electrocompetent cells and plated on membranes placed onto 22 cm square plates containing solid LB agar (200 mL), glycerol (1% w/v), and ampicillin (100 μg/mL). After overnight growth at 37°C, each membrane was transferred to a 22 cm square plate containing inducing medium [solid LB agar (200 mL), IPTG (1mM), ampicillin (100 μg/mL)] supplemented with sucrose (50 g/L) and stained in blue by addition of 50mM Tris-HCl at pH 7.5 and pH indicator bromothymol blue (BTB, 0.1 g/L in 1% ethanol). The plates were incubated for 24 h at 30 °C. Active clones (green and yellow) were isolated in microplates containing LB medium (200 μL), glycerol (12% w/v) and ampicillin (100 μg/mL).

References

(S1) Skov, L. K.; Mirza, O.; Sprogøe, D.; Dar, I.; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2002**, *277*, 47741-47747.

(S2) Bayly, C.; Kollmann, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620-9631.

(S3) Case, D. A.; Darde, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollmann, P. A. **2006**, AMBER 9, *Univ. California, San Francisco, CA, USA*.

(S4) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins Struct. Funct. Genet.* **2006**, *65*, 712-725.

(S5) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622-655.

(S6) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(S7) Daudé, D.; Topham, C. M.; Remaud-Siméon, M.; André, I. *Protein Sci.* **2013**, *22*, 1754-1765.

(S8) Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. *Nat. Struct. Biol.* **2003**, *10*, 59-69.

(S9) Champion, E.; André, I.; Moulis, C.; Boutet, J.; Descroix, K.; Morel, S.; Monsan, P.; Mulard, L. A.; Remaud-Siméon, M. *J. Am. Chem. Soc.* **2009**, *131*, 7379-7389.

(S10) Champion, E.; Moulis, C.; Morel, S.; Mulard, L. A.; Monsan, P.; Remaud-Siméon, M.; André, I. *ChemCatChem* **2010**, *2*, 969-975.

(S11) Perepelov, A. V.; Shekht, M. E.; Liu, B.; Shevelev, S. D.; Ledov, V. A.; Senchenkova, S. N.; L'vov, V. L.; Shashkov, A. S.; Feng, L.; Aparin, P. G.; Wang, L.; Knirel, Y. A. *FEMS Immunol. Med. Microbiol.* **2012**, *66*, 201-210

Chapter III:

Novel product specificity toward erlose and panose exhibited by multi-site engineered mutants of amylosucrase.

In the first chapter, we showed how the computer-aided engineering approach enabled to profoundly reshape the active site of amylosucrase enabling recognition of a non-natural disaccharide. During screening campaign of this library on sucrose, several mutants were found to synthesize novel molecules from sole sucrose, which were not synthesized by parental wild-type enzyme. This chapter will be devoted to the detailed characterization of three of these mutants, named 39A8, 47A10 and 37G4, as well as of the new products synthesized. These mutants contained between 7 and 11 mutations in the active site and the new molecules were identified as being erlose and panose. Optimization of the production of these trisaccharides was carried out using different sucrose concentrations. Molecular modelling studies were also carried out to shed some light on the molecular factors involved in the novel product specificity of amylosucrase mutants.

INTRODUCTION

Amylosucrase from *Neisseria polysaccharea* (*Np*AS) is an α-transglucosylase of Glycoside-Hydrolase family 13[1,2]. This sucrose-utilizing enzyme follows a double-displacement mechanism that proceeds via the formation of a covalent α-D-glucosyl enzyme intermediate with a concomitant release of fructose[3]. This intermediate is subsequently attacked either by water molecules (hydrolysis reaction), fructose, glucose or maltooligosaccharides released during the reaction course (Figure 1). Naturally, the main reaction catalyzed is the formation of soluble maltooligosaccharides and insoluble amylose-like polymer. Transfer onto fructose results in the formation of two sucrose isomers ie: turanose (3-O-α -D-glucopyranosyl-D-fructose) and trehalulose (1-O-α-D-Glucopyranosyl-D-fructose).
Three-dimensional structures of *Np*AS in apo form or in complex with sucrose or maltoheptaose allowed to map the buried catalytic pocket of the enzyme and delineate -1 to +6 subsites[4,5]. Wild-type *Np*AS catalyzes glucosylation of a large repertoire of exogenous acceptors and naturally reveals a broad promiscuity towards carbohydrates, polyols or flavonoids[6,7,8]. In addition, enzyme engineering allowed to evolve *Np*AS and generate mutants with controlled and extended capacity of glucosylation[9,10]. Site-directed mutagenesis targeting *Np*AS active site enabled the distribution of natural products synthesized from sucrose to be finely tuned[11,12]. Neutral drift approach and semi-rational design allowed to sort out mutants with drastically improved ability to glucosylate acceptor molecules compared to parental *Np*AS[8,10,13,14] .

Very recently, *Np*AS active site was profoundly reshaped to adapt the enzyme to the glucosylation of unnatural and slightly protected disaccharides entering in a programmed chemo-enzymatic pathway designed for bacterial antigen synthesis[15]. Using a computer-aided approach, a semi-rational library of ca. $2.7x10^4$ variants, targeting 23 amino acid positions of subsites +1 and +2 in the active site first shell, was designed. Upon library construction, the mutants were screened, first, on sucrose activity and, next, on their ability to catalyze the glucosylation of the target acceptor, a partially protected β-linked disaccharide, the allyl (2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranoside. Whereas only one mutant containing 7 mutations was able to glucosylate the target disaccharide, 55 mutants containing between 5 and 17 mutations were still active on sucrose in spite of their high number of mutations.

The objective of the work disclosed herein was the characterization of this small library of 55 active mutants and comparison of parental enzyme and mutant performances to investigate in more details the effect of the mutations introduced at the +1 and +2 subsites of *Np*AS onto the products formed from sucrose substrate. Mutants were classified on the basis of their product profiles and representative members of each class were deeply studied at both biochemical and structural levels. As a result, several mutants producing novel trisaccharides, not synthesized by wild-type *Np*AS, and identified as being erlose and panose, were isolated. These biocatalysts could offer alternative routes for the direct production from sucrose of sweeteners or prebiotic molecules.

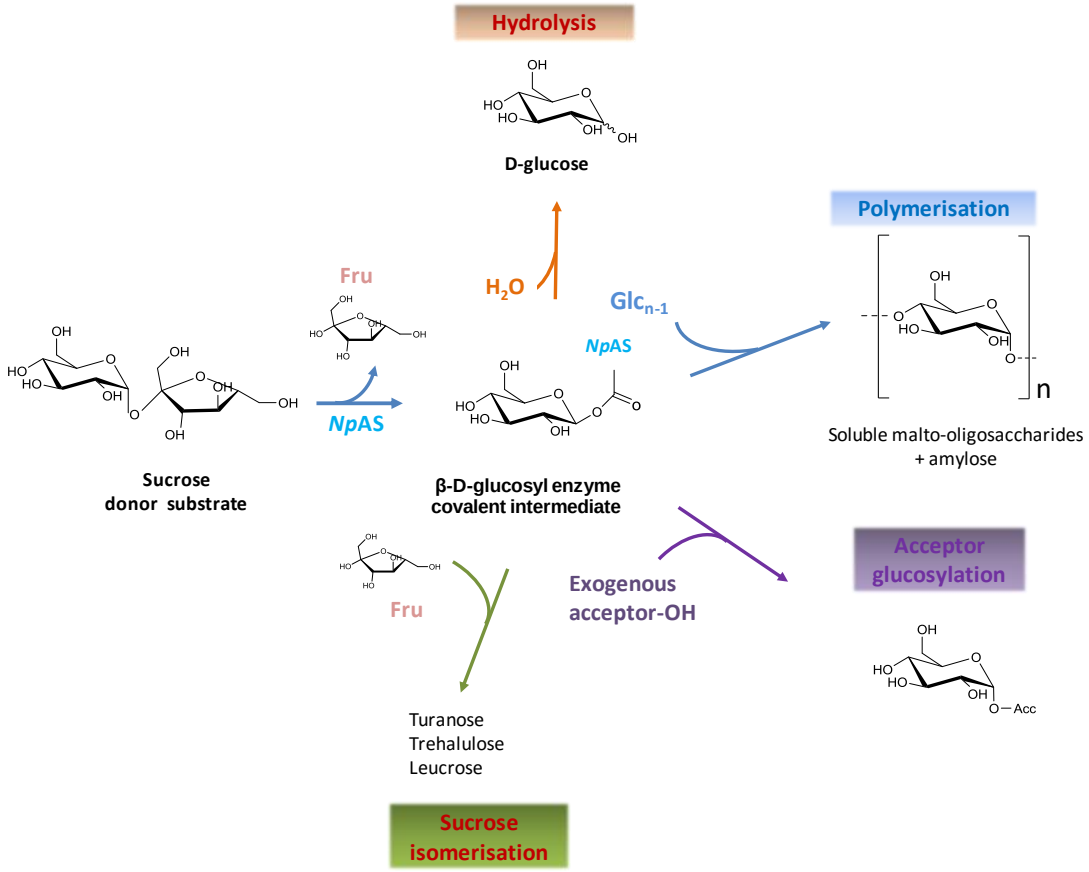


Figure 1. Products synthesized by *Np*AS via the formation of a covalent β-D-glucosyl-enzyme intermediate. *The main reaction products are insoluble high molecular weight α-1,4-linked-glucans and soluble malto-oligosaccharides synthesized by polymerization reaction. Glucose is released by hydrolysis. Glucosylation of fructose released during the first reaction step results in the formation of turanose, trehalulose, and leucrose by sucrose isomerisation. Glucosylation of exogenous non-natural hydroxylated acceptors can also occur if acceptors are recognized by the enzyme. Fru, fructose; Glc, glucose; Acc, acceptor molecule.*

RESULTS

Analysis of library screening on sucrose.

From semi-rational libraries of *Np*AS mutants designed to tailor novel acceptor specificity toward a partially protected β-linked disaccharide, 55 sucrose-active mutants were isolated[15]. The products synthesized from sucrose with these enzymes were analyzed by HPLC to compare their product profiles and exclude pure sucrose-hydrolases. As a result, 38 variants (out of 55) were identified to be transglucosylases. They contain between 5 and 17 mutations. Despite their high number of mutations, all of them turned out to be more active on sucrose than wild-type *Np*AS. Analysis of their product profiles further showed that 21 mutants synthesize the same products as the wild-type enzyme, although in varying amounts (Figure 2). On the basis of their HPLC chromatograms, the other mutants were classified in three groups. The first class, contains 11 mutants that synthesize, in addition to all natural products, a new product called P1, not formed by *Np*AS, with a retention time comprised between those of maltotriose and maltotetraose (Figure 2B). Another group of 6 mutants revealed their ability to produce a second type of molecule P2 (7.2 min) showing a retention time different from that of P1 (7.0 min) and which was not synthesized by parental *Np*AS (Figure 2C). The last group is composed of only one mutant that

is able to synthesize both molecules (P1 and P2) (Figure 2D).

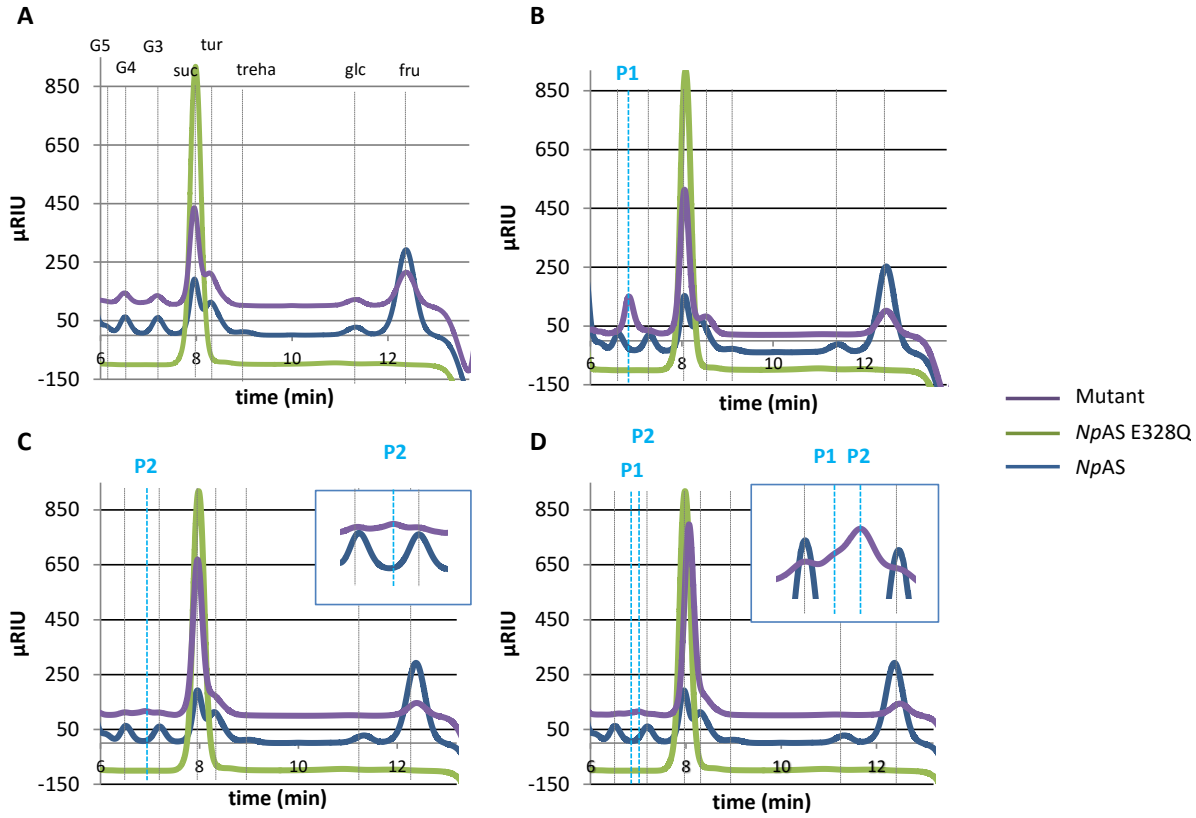


Figure 2. HPLC profiles of products synthesized from 146mM sucrose by wild-type *Np*AS and one representative mutant of each defined class. *In purple: product profile for each category of mutants; In green: profile obtained with inactive* Np*AS (*Np*AS E328Q); in blue: profile obtained with wild-type* Np*AS. Profile A corresponds to mutants that synthesize similar products to the wild-type enzyme (Class A). Profile B corresponds to mutants that synthesize a new product P1 (7.2 min), not observed for the wild-type enzyme (Class B). Profile C corresponds to mutants that synthesize novel product P2 (7.0 min), not formed by the wild-type enzyme (Class C). Profile D corresponds to a mutant that synthesizes both P1 and P2 molecules (Class D). Glc, Glucose; Fru, Fructose; Treha, Trehalulose; Tur, Turanose; Suc, Sucrose; G3, maltotriose; G4, maltotetraose; G5, maltopentaose; P1, product 1; P2, product 2.*

Characterization and quantification of products synthesized from sucrose.

Among mutants able to synthesize P1, P2 or both, the best producer of new compounds (based upon peak area comparison at 24h) was selected as the representative enzyme of each class. This led to the selection of mutant 47A10 for class B, 39A8 for class C and 37G4 for class D. Their amino acid sequences at targeted positions are reported in Table 1.

Table 1. Sequences of the selected mutants in comparison to the wild-type enzyme. *The 23 targeted positions are shown. In orange, are highlighted mutated positions compared to wild-type.*

			Mutants		
Amino acid positions		wild-type	47A10	37G4	39A8
226	Loop 3 (domain B)	R	L	K	K
227		E	E	E	E
228		I	V	V	V
229		F	A	F	F
289	Loop 4 (domain A)	A	I	I	I
290		F	Y	Y	Y
300		E	I	I	I
331	Loop 5 (domain A)	V	T	T	T
332		H	H	H	H
396	Loop 7 (Domain B')	G	G	S	S
397		W	W	W	W
398		T	T	V	V
434		V	V	V	V
435		P	P	P	P
436		F	F	F	F
437		Q	Q	R	S
438		Y	Y	Y	Y
439		N	N	D	D
444		D	D	D	D
445		C	C	C	R
446		R	R	R	R
447		V	V	V	V
448		S	S	S	S
Number of mutations			7	10	11

All three mutants were then produced, purified to homogeneity and incubated with 146mM sucrose. Unlike wild-type *Np*AS, the three mutants only produced soluble compounds showing that the polymerization activity was fully abolished for all mutants. The HPAEC-PAD profiles of the soluble products formed are provided in Figure 3.

Maltooligosaccharides, turanose and trehalulose are present in all chromatograms. The most striking differences reside in the synthesis of P1 and P2 and the traces of P3 products observed for the mutants while the wild-type *Np*AS was unable to do so. Whereas HPLC chromatograms of our preliminary screening assays indicated only P2 or P1 production for mutants 39A8 and 47A10, respectively, HPAEC-PAD analyses which are more sensitive revealed that both P1 and P2 compounds are produced by these mutants. Furthermore, HPAEC-PAD analysis coupled to MS detection indicated a molar mass of m/z = 505g/mol [ES + H^+]) for P1, P2 and P3 showing that all compounds are trisaccharides. The reaction mixes were submitted to the action of an invertase and an α-amyloglucosidase and digestions were further analyzed by HPAEC-PAD. Product P1 was cleaved by the invertase to yield maltose and fructose while P2 was not. Furthermore, amyloglucosidase hydrolysis yielded glucose and sucrose from P1 and glucose and maltose from P2. These results suggested that P1 and P2 correspond to erlose, α-D-Glucopyranosyl-(1→4)-α-D-Glucopyranosyl-(1→2)-β-D-Fructose and panose, α-D-Glucopyranosyl-(1→6)-α-D-Glucopyranosyl-(1→4)-α-D-glucose, respectively. This was confirmed by the retention times of pure commercial erlose and panose. Due to the very low level of P3 production, its structure was not elucidated.

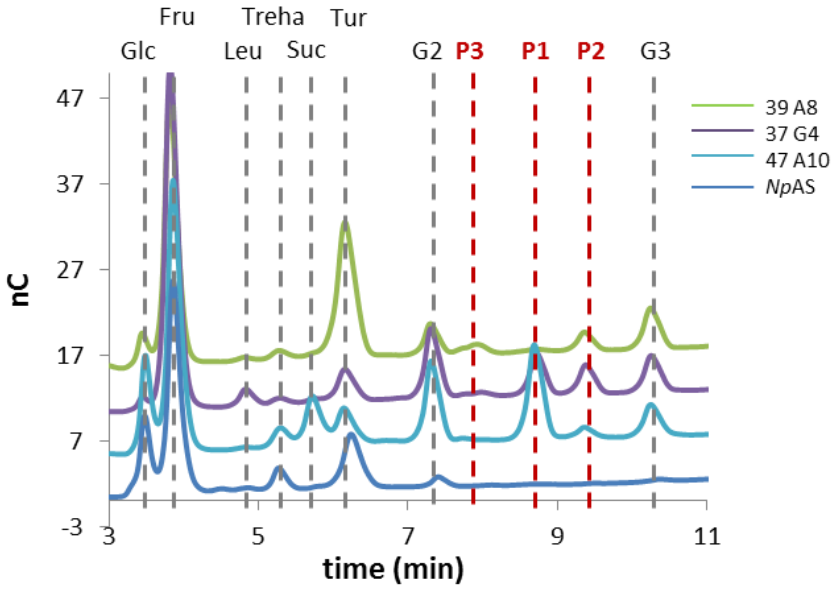


Figure 3. HPAEC-PAD product profiles obtained after 24h reactions in the presence of 146mM sucrose and 1U/mL purified enzyme. *The profile of mutant 37G4 is shown in purple, mutant 39A8 in green, mutant 47A10 in cyan and the wild-type* Np*AS is in blue. Glc, Glucose; Fru, Fructose; Treha, Trehalulose; Tur, Turanose; Suc, Sucrose; Leu, Leucrose; G_2, Maltose; P1, product 1; P2, product 2; P3, product ; G_3, maltotriose.*

The percentage of glucosyl units incorporated from sucrose, in each type of product is reported in Figure 4. Compared to *Np*AS, and in agreement with their loss of polymerase activity, all three mutants incorporated higher amounts of glucosyl units in maltose (27.3% (mutant 47A10); 24.8% (mutant 37G4); 15.3% (mutant 39A8); *Np*AS 2.9%) and maltotriose 20.5% (mutant 47A10); 18.6% (mutant 37G4), 23% (mutant 39A8); *Np*AS 2.9%, respectively). Compared to other mutants, mutant 39A8 is specialized in turanose production and incorporates 46% of the glucosyl residues in turanose, versus only 20% for wild-type *Np*AS. With mutants 47A10 and 37G4, 23% and 30% glucosyl units were incorporated into erlose and panose, respectively. Much lower values were observed with mutant 39A8 (only 1.3%). In comparison, panose was mainly produced by mutants 37G4 and 39A8, 17.7% and 8.7 % of the glucosyl units, respectively, being incorporated into this trisaccharide. In comparison, the value goes down to 1.7% with mutant 47A10.

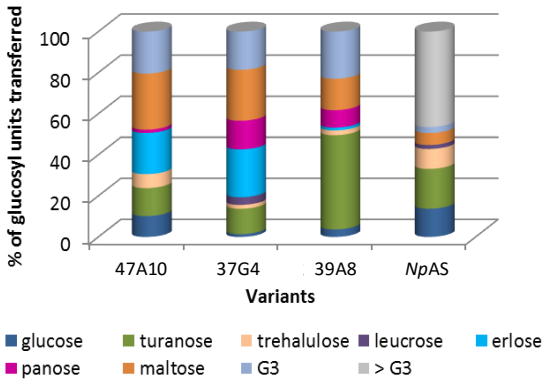


Figure 4. Percentage of glucosyl units transferred from sucrose to the various products with mutants 47A10, 39A8, 37G4 and wild-type *Np*AS. *Reactions were performed with 146mM sucrose and 1U/ml of each enzyme. G: Glucose. The percentages were determined from HPAEC analysis.*

Determination of kinetic parameters and stability.

The steady-state kinetic parameters (k_{cat}, K_m, k_{cat}/K_m) were determined for all purified enzymes upon varying sucrose concentrations from 20 to 600mM. Only

mutant 39A8 showed a standard saturation kinetic behavior comparable to that observed for the wild-type enzyme and its k_{cat}/K_m value was found increased by 2.2-fold. Regarding mutants 47A10 and 37G4, no saturation behavior could be observed in these conditions suggesting that K_m values might be very high and the affinity for sucrose quite low. Linear plots correctly fitted initial rate of sucrose consumption versus sucrose concentration. Consequently, catalytic efficiencies were determined from the slope of the linear regression of initial rates of transglucosylation measured according to variable substrate concentrations. The k_{cat}/K_m values towards sucrose of both mutants decreased by 1.7-fold (47A10) and 2.1 fold (37G4) compared to that of *Np*AS.

Table 2. Determination of apparent kinetic parameters and thermostability values of mutants 47A10, 39A8 and 37G4 and wild-type *Np*AS.

Enzymes	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	Behavior	Tm (°C)
Np*AS	1.3	50.2	0.026	Saturation	49.0
47A10	n.d.	n.d.	0.015	Linear	46.0
37G4	n.d.	n.d.	0.012	Linear	46.0
39A8	4.5	78.43	0.058	Saturation	47.1

*Data from Potocki de Montalk *et al*., 1999, for sucrose > 20mM.

Thermal stability of all three mutants was assessed using circular dichroïsm. Tm values are listed in Table 2. Overall, introduced mutations had a negative impact on thermal stability as Tm values were found lowered compared to those of wild-type *Np*AS (-2.8°C for 39A8; -3.9°C for 47A10 and 37G4, respectively). However, with regard to the considerable number of buried mutations (Table 1), between 7 and 11 mutations in subsites +1 and +2 of their active sites, their stability was not drastically affected.

Optimization of erlose and panose production.

Reactions were performed at various sucrose concentrations (146, 400 and 600mM). At 400mM and 600mM, sucrose was not fully consumed after 24h reaction with mutant 47A10, indicating a possible inhibition by the substrate, which was also observed with mutant 37G4 at 600mM sucrose (Table 3).

Table 3. Erlose and panose production data using mutants of *Np*AS.

146mM sucrose = 50g/L	47A10	37G4	39A8
Sucrose consumption, %	87.2	98.7	98.8
Erlose production, g/L	6.4	8.5	0.5
Erlose production yield[a]	14.7	17.2	1.0
Panose production, g/L	0.4	3.4	2.1
Panose production yield[b]	0.8	6.8	4.2
400mM sucrose = 137 g/L			
Sucrose consumption, %	65.5	95.0	96.9
Erlose production, g/L	36.8	49.4	5.5
Erlose production yield	41.1	38.0	4.1
Panose production, g/L	1.5	13.1	1.0
Panose production yield	1.7	10.1	0.8
600mM sucrose = 205 g/L			
Sucrose consumption, %	81.2	61.0	94.7
Erlose production, g/L	70.6	57.5	9.6
Erlose production yield	42.4	45.9	4.9
Panose production, g/L	3.8	8.0	1.5
Panose production yield	2.3	6.4	0.8

[a] Erlose production yield= [erlose (g/L)]/ [sucrose consumed (g/L)]
[b] Panose production yield = [panose (g/L)]/ [Sucrose consumed (g/L)]

Nevertheless, for all mutants, erlose production was significantly improved at high sucrose concentration, the best values being obtained at 600mM. As seen in Figure 5, increasing sucrose concentration greatly favored acceptor reaction on sucrose. As a result, the percentage of glucosyl units incorporated into erlose increased by 2.5-fold (47A10), 2.1-fold (37G4) and 5.8-fold (39A8) at 600mM in comparison to the values observed at 146mM sucrose. In these conditions,

concentrations of erlose attained 70.6 g/L, 57.5 g/L and 9.6 g/L with corresponding yields of 42%, 46% and 5.2% in the reactions catalyzed by mutants 47A10, 37G4 and 39A8, respectively. In contrast, increasing sucrose concentration had a limited impact on panose yield (Figure 5).

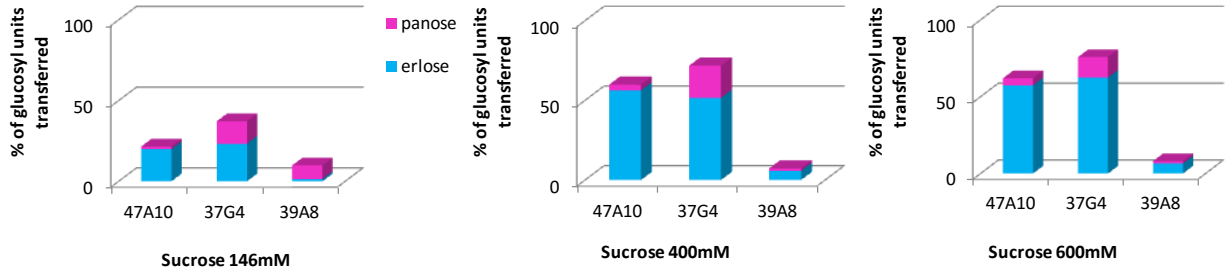


Figure 5. Incorporation of glucosyl units from sucrose into erlose (in cyan) and panose (in magenta) at 146mM, 400mM and 600mM sucrose with mutants 47A10, 37G4 and 39A8.

Structural insight on the effect of mutations onto novel product specificity toward erlose and panose.

To investigate further the molecular determinants possibly involved in the product specificity, three-dimensional models of the enzymes in complex with the various products of interest (maltotriose, turanose, erlose, panose) were built and analyzed in details (Figure 6). Six amino acid positions (226, 228, 289, 290, 300 and 331) are found commonly mutated in the three mutants. Among them, only position 226 differs in the type of mutation as the arginine is found respectively mutated by a lysine in mutants 37G4 and 39A8 and by a leucine in mutant 47A10. These six positions belong respectively to loop 3 (residues 224-236), loop 4 (residues 285-303) and loop 5 (residues 327-332) of the catalytic barrel. They have been shown to be highly flexible by molecular dynamics simulations and assumed to play a determinant role in the formation of the active site topology and catalytic mechanism of the enzyme[16,17,18]. In particular, Arg226 residue in native *Np*AS has been reported to be engaged in a salt bridge with Glu300, controlling the shape of +2/+3 subsites and interfering with the binding of long acceptors. Mutation of residues 226 and 300 leads thus to a disruption of the salt bridge and a reorganization of the interactions at +1 and +2 subsites with mutated residues V331T, I228V, A289I and F290Y. In particular, mutation of Arg226 was found to open up the active site around +3 subsite, likely interfering with the binding of oligosaccharides longer than DP3. In mutant 47A10, only one additional mutation, F229A, was found in addition to the 6 commonly mutated positions. Docking of erlose molecule, the main product formed by this mutant, suggested that fructofuranosyl moiety is stabilized at +2 subsite by hydrogen bonding interactions of fructose (O_3·H) with R446 side chain and I330 carbonyl group. In comparison with maltotriose binding mode, docking of panose in mutants 39A8 and 37G4 indicates a shift of glucopyranosyl units at +1 and +2 subsites

due to the α-1,6 linkage present in the molecule. As a result, all hydrogen bonding network stabilizing the glupyranosyl moieties at +1 and +2 subsites is altered. Mutations G396S, T398V and N439D are present in both mutants 39A8 and 37G4. In particular, mutations G396S and T398V appear to sterically hinder binding at subsites +2/+3 of oligosaccharides longer than DP3. Mutants 39A8 and 37G4 only differ by the amino acid type present at positions 437 and 445. Interestingly, all these mutations are in the vicinity of the +2 subsite and carried by the B' domain (residues 396-460) which is also highly flexible and assumed to be involved in product specificity. Although molecular determinants responsible for the distinct product profiles observed for the two mutants are still unclear, mutant 39A8 was found to produce majorly turanose and 37G4 synthesized mainly both erlose and panose.

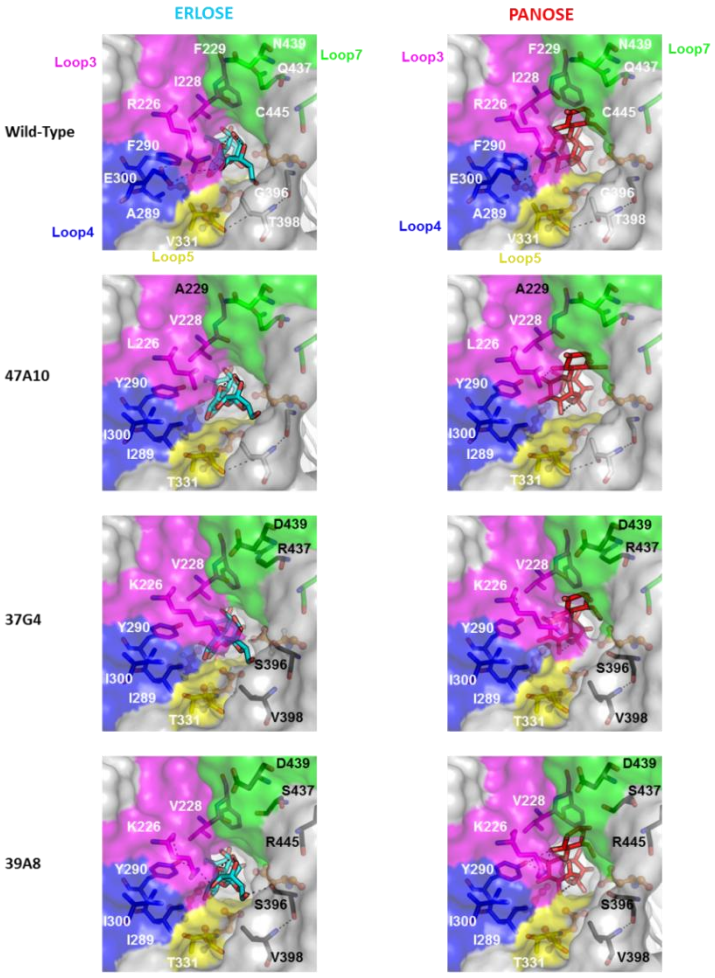


Figure 6. Molecular docking of erlose and panose in wild-type *Np*AS and its mutants 47A10, 37G4 and 39A8. *Mutated positions are shown and colored depending on the loop they belong to. Labelled in white are highlighted commonly mutated positions and in black are shown mutated positions that are specific to the mutant.*

DISCUSSION

Semi-rational design is widely viewed as a very efficient approach for tailoring substrate or product specificity of enzymes. Applied to *Np*AS enzyme, a computer aided-approach to engineer subsites +1 and +2 allowed to isolate one mutant with a totally new ability to recognize an unnatural and chemically protected disaccharide of interest for novel vaccine elaboration[15]. The re-design of the catalytic pocket generated 55 sucrose-active clones out of 63,000 clones. In the present work, we examine in more details the properties of these variants in terms of product specificity, efficiency and stability. Indeed, we anticipated that reshaping of subsites +1 and +2 might have impacted these properties in comparison to wild-type *Np*AS. Mutant discrimination, based on product profile, revealed that 62 % of active enzymes differed from wild-type *Np*AS and showed a novel product specificity, highlighted by the production of erlose (α-D-Glucopyranosyl-(1→4)-α-D-Glucopyranosyl (1→2)-β-D-Fructose) and panose (α-D-Glucopyranosyl-(1→6)-α-D-Glucopyranosyl-(1→4)-α-D-glucose) from sucrose substrate. Noteworthy, these mutants were specialized in short oligosaccharides production and lost their ability to transfer glucosyl units onto oligosaccharides of DP higher than 3. The high percentage of hits showing a new product specificity further demonstrates how powerful computer-aided design can be to address the challenge of changing substrate specificity. In addition, molecular modelling of the mutants revealed that compared to wild-type enzyme, abolition of polymerase activity could be attributed to subsites +1 and +2 reshaping by suppression of salt bridge interaction between Arg226 and Glu300 present in wild-type enzyme and mutation of residues G396 and T398 into bulkier residues (substitution by serine or valine residues, respectively) that hamper binding of oligosaccharides longer than DP3 in subsites +1, +2 and +3 and thus favoring the accumulation of lower size products. Sucrose and natural acceptor binding in mutants 47A10 and 37G4 has also probably been impacted as revealed by the decrease of mutant catalytic efficiency (by 2-fold) compared to the wild-type enzyme. Noteworthy, with regard to the number of mutations found in the active site, such a decrease remains acceptable. In contrast mutant 39A8 was found more efficient on sucrose substrate than *Np*AS. Increase of catalytic efficient relative to wild-type *Np*AS is mainly due to a k_{cat} increase. As the mutant produces incredibly high amounts of turanose (121 g/L with 59 % yield at 600mM sucrose and 23 g/L with 45.9% yield at 146mM) in comparison with wild-type *Np*AS, we can suggest that fructose is much better accommodated by the mutant to intercept the covalent β-glucosyl-enzyme intermediate. Interestingly, the introduction of 7 to 11 mutations in subsites +1 and +2 of *Np*AS mutants allowed the emergence of a new α-1,6 glucosidic linkage specificity revealed by the presence of panose in addition to maltotriose. This indicates a dual accommodation of maltose acceptor at subsite +1, mainly observed for mutant 37G4, which incorporated 13.9% and 18.6% of glucosyl units in panose and maltotriose, respectively. The most striking fact is that all mutants 47A10, 37G4 and 39A8 acquired the ability to use sucrose as acceptor to produce erlose. Both mutants 47A10 and 37G4 efficiently produced erlose and increasing sucrose concentration favored erlose production over other products, allowing to reach 42% and 46 % yields, respectively. Of the two mutants, mutant 47A10 was the most specific for erlose production. Reshaping of the active site thus promoted sucrose binding at subsites +1 and +2 to the detriment of glucose and maltose in the mutant in comparison with the wild-type *Np*AS. Erlose and panose are interesting bioactive

oligosaccharides due to their acariogenic and low-calory value properties[19,20,21]. Erlose is half as sweet as sucrose while providing similar taste[22]. Furthermore, it could also be used to prevent crystal formation and browning reaction during heat processed food[23,24]. Erlose may not be resistant to digestive enzymes and its prebiotic properties remains to be investigated. However, erlose naturally occurs in honey and other natural sources but is naturally present in small amounts. Thus, several enzymatic routes leading to this trisaccharide were previously investigated. One α-glucosidase from *Archachatina ventricosa* of yet unknown GH family was shown to produce erlose from single sucrose[25]. A mix of gluco-oligosylfructosides of DP ranging from 3 to 6 was synthesized in which erlose was identified but not quantified. Cyclodextrin glucanotransferase from *Thermoanaerobacter* sp. was also shown to produce erlose from cyclodextrin donor and sucrose acceptor. From 16mM cyclodextrin and 155mM sucrose, 35mM erlose was synthesized, which represents 17.6g/L[48]. Transfructosylation reaction also allowed to produce erlose. However, in all cases, maltose acceptor was required in addition to sucrose donor. Production levels around 138g/L, 163g/L or 61g/L were obtained with levansucrase from *B. amyloliquefaciens* (using 0.8M sucrose/0.4M maltose[26], *Bacillus subtilis* (0.7M sucrose/0.7M maltose[27] or inulosucrase from *Lactobacillus gasseri*[28] 0.6M Sucrose/1.17M Maltose, respectively). Our process of production compares favorably with these alternative enzymatic processes as the mutants of amylosucrase produce erlose in fairly high amounts from sole sucrose. In addition, one of the mutants reported here was also a substantial producer of panose, a trisaccharide of interest also for food industry. It displays similar sweetening properties as erlose, but it also displays anti-fading and anti-oxidant properties of interest for food industry[29]. Several studies have proven its non-digestibility in the upper gastrointestinal tract by *in vitro* assays[20,30–34]. Numerous synthetic routes to panose exist and the molecule can be produced in high yield via enzymatic processes including hydrolysis of pullulan or starch with pullulanase or α-amylase[35,36,37,38], or by transglucosylation reaction with dextransucrase from *L. mesenteroides* from sucrose donor and maltose acceptor[39–42] and by glycoside-hydrolases from the marine anaspidean mollusk *Aplysia fasciata*[43]. However, no study reported the direct production of panose from sucrose alone.

CONCLUSIONS

In conclusion, this is the first report of novel products obtained from sole sucrose using engineered amylosucrases. Indeed, erlose and panose trisaccharides are not produced by any of the wild-type amylosucrases characterized to date. Yields reported here after optimization of substrate concentrations are quite encouraging as they can surely be further optimized in batch or fed-batch reactors to propose novel synthetic routes, especially for the production of erlose, directly from an agroresource, sucrose.

THEORETICAL AND EXPERIMENTAL PROCEDURES

Bacterial strains, plasmids & chemicals.

Plasmid pGST-amylosucrase (pGST-AS) derived from the pGEX-6P-3 (GE Healthcare Biosciences, Piscataway, U.S.A) encoding GST fused to *N. polysaccharea* amylosucrase, was used for the construction of semi-rational libraries. *E.coli* TOP 10 electrocompetent cells (Invitrogen, Carlsbad, U.S.A.) were utilized as gene expression and production of recombinant amylosucrase variants. DpnI, EcoRI HF and NotI HF restriction enzymes, T4 DNA ligase and Antartic

Phosphatase were purchased from New England Biolabs (Beverly, MA, U.S.A.). Phusion DNA-polymerase was purchased from Finnzymes (Espoo, Finland). Ampicillin, lysozyme and isopropyl-β-D-thiogalactopyranoside (IPTG) were purchased from Eurodemex (Souffelweyersheim, France); Bromothymol blue was ordered from Sigma – Aldrich (St Louis, Mo, U.S.A.). Oligonucleotides were synthesized by Eurogenetec (Liege, Belgium). DNA plasmid extraction (GenElute HP plasmid Miniprep kit, Qiagen, Chatsworth, CA, U.S.A.), DNA purification and gel extraction (GenElute Gel Extraction kit) columns were purchased from Sigma-Aldrich (St Louis, MO, U.SA.). DNA sequencing was performed by GATC Biotech AG. (Constance, Germany). Erlose and panose standards were purchased from Carbosynth (Compton, Berkshire, U.K.).

Library construction.

The construction of the library from which were isolated the mutants characterized herein was previously described[15]. Briefly, the libraries of *Np*AS were initially designed to recognize and glucosylate a partially protected β-linked disaccharide acceptor called allyl 2-deoxy-2-*N*-trichloroacetyl-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside. 63,000 variants were screened using a pH-based assay on solid medium as described earlier[15]. The 58 active clones were stored in 96-well microplates containing LB medium (200µl), ampicillin (100µg.mL^{-1}) and glycerol (12% w/v).

Library screening in microplate format.

The 58 mutants were grown at 30°C for 24h in microtiter plates that serve to inoculate 96-Deep Well plates containing autoinducing media ZYM-5052[44] supplemented with ampicillin (100 µg/mL). After 24 hour growth at 30°C, cell lysis was performed by addition of lyzozyme (0.5mg/mL) and DNAse and freezing at -80°C overnight. After thawing at room temperature, microplates were centrifuged (20min, 3,000g, 4°C) to harvest the enzyme extract. The enzyme extract was divided in two fractions. On one side, 50µL of the soluble fractions were incubated with 50µL of 292mM sucrose solution for 1h at 30°C. Reactions were stopped by addition of 100µL of a dinitrosalycilic acid (DNS) solution, heated at 95°C during 10 min. Reactions mixes are then diluted twice or 5-times (for the highest activities) before optical density measurements at 540 nm. The content of reducing sugars was determined from a calibration curve established with fructose standards. 55 clones were confirmed and sequenced on the entire gene by GATC Biotech GA (Constance, Germany). The second part of the enzyme extract (100µl) was also incubated with sucrose (100µl at 292mM) for 24h at 30°C. The reaction was stopped by heating at 95°C for 5min. The mix was centrifugated and the supernatant was analyzed by HPLC using an Aminex HP87K column (300 x 7.8 mm, Biorad) with ultra-pure water at 0.6 mL/min as eluant at 65°C. The detection was performed with a Shodex RI 101 series refractometer. The mutants differing from the wild type *Np*AS on the basis of their product profiles were then classified into three classes according to their product profiles. One representant of each class (the best one for production of P1 or P2 or P1/P2 according to peak area at t24h by RI detection in HPLC) was selected *ie* mutants 47A10, 39A8, 37G4 and produced at higher scale to allow enzyme purification and characterization.

Characterization of the selected mutants.

Production and Purification.

A pre-culture (50mL of LB inoculated with 50µL of stock culture *of* the selected clones) was performed overnight at 30°C and used to inoculate 1L of LB+ampicillin

at $OD600_{nm}$ = 0.2 and grown at 23°C. Cultures were induced at $OD600_{nm}$ = 0.5 with IPTG and incubated for 17h. Cells were centrifugated at 5,000 Rpm at 4°C for 10 min. The pellet was resuspended in Tris 50mM, NaCl 150mM at pH 7 and $OD600_{nm}$=80. After sonication, the crude extract was centrifugated at 10,000 Rpm, 20min, 4°C. Before purification by affinity chromatography on the AKTA Express systems, these supernatants were filtered on a syringe with a Minisart NML filter (Sartorius, Goettingen, Germany) to remove additional cell wastes. Then, they were loaded onto Glutathion Sepharose 4B column (Dutscher, Brumath, France) in the presence of Tris buffer supplemented in EDTA and DTT (50mM Tris, 150mM NaCl, 1mM EDTA, 1mM DTT, pH7.0). In order to maximize protein fixation on gel, the flow through was recycled to be loaded a second time on columns. Then, PreScission protease (Dutscher, Brumath, France) (500U in 5mL buffer (50mM Tris, 150mM NaCl, 1mM EDTA, 1mM DTT, pH7.0)) was injected and cleavage of GST Tag occurred during 1 night always at 4°C. The next day, protein and GST Tag were eluted in Deep-Well by fraction of 2mL in a Deep-Well by UV detection at 280nm. Protein concentration was estimated using Nanodrop ND-1000 spectrophotometer. Specific activity was determined at 30°C, in 50mM Tris-HCl buffer pH7.0, using 50g/L sucrose. Reducing sugar production was determined using dinitrosalicylic acid assay. One unit of amylosucrase activity corresponds to the amount of enzyme that catalyzes the release of 1µmol of reducing sugars per minute in the assay conditions.

Analysis of product formed from sucrose.

Reactions were performed at 30°C during 24h with 146mM sucrose in the presence of wild-type enzyme or mutants at 1U/mL. Reaction mixs were stopped by heating at 95°C during 5min. The reaction medium was centrifugated. In contrast to wild-type *Np*AS, the three mutants did not synthesize insoluble products. Supernatants were all diluted at $10mg.L^{-1}$ and analyzed by HPAEC using the same column as above coupled to mass spectrometry detection, ¼ of the flow containing the sample was injected into the MSQ mixed with TFA at 150mM and formic acid at 0.1% (v/v). Samples were eluted with a gradient of sodium acetate (from 6 to 300mM in 28min) in 150mM NaOH. Detection was performed using a Dionex ED40 module with a gold working electrode and Ag/AgCl reference electrode.

Optimization of erlose and panose production.

Reactions were performed with 1U/mL of each purified mutants or wild-type enzyme in the presence of 146mM, 400mM and 600mM sucrose concentrations at 30°C during 24hours. Following 5 min heating at 95°C, reaction mixes were centrifugated. The soluble fractions were analyzed by HPAEC-PAD as described above.

Kinetic parameters determination.

Initial velocity of sucrose consumption determined from reactions were carried out with pure enzymes (0.07 and 0.1U/ml) at sucrose concentrations ranging from 20 to 600mM and 30°C under agitation. After heating at 95°C and centrifugation, aliquots were analyzed by HPLC RI with a Biorad HPX-87K column to determine sucrose consumed and fructose release.

Stability studies.

The melting point (Tm) of amylosucrase mutants was determined by Circular Dichroism analysis on a JASCO J815 spectropolarimeter, using a Peltier cell temperature controller with a thermal transition monitored from 20 to 80°C. Purified enzymes were diluted at 2µM in Tris buffer (50mM Tris, 150 mM NaCl, 1mM DTT, 1 mM EDTA, pH7.0).

Computational procedures.

Starting coordinates were extracted from the crystallographic structure of *Np*AS in complex with sucrose (PDB ID: 1JGI)[45] to generate 3D models of the different enzymes of interest and mutations were introduced using the Biopolymer module of the InsightII software package (Accelrys). The conformations of the mutated residue side chains were optimized by manually selecting a low-energy conformation from a side-chain rotamer library. Steric clashes (van der Waals overlap) and non-bonded interaction energies (Coulombic and Lennard-Jones) were evaluated for the various side-chain conformations. The 3D models were then refined using the CFF91 force field implemented within the DISCOVER module of InsightII software suite (Accelrys, San Diego, CA, USA). For the minimization, the CFF91cross terms, a harmonic bond potential, and a dielectric of 1.0 were used. An initial minimization with a restraint on the protein backbone was performed using a steepest descent algorithm followed by conjugated gradient minimization steps until the maximum RMS was less than 0.1. In a subsequent step, the system was fully relaxed. Using as templates the three dimensional structures of *Np*AS co-crystallized with sucrose (PDB ID: 1JGI)[45], maltoheptaose (PDB ID: 1MW0)[46] and turanose (PDB ID: 3UEQ)[47], we have built models for the wild-type *Np*AS and the three selected mutants in complex with molecules of interest, maltotriose, panose, erlose and turanose. Models were then subjected to the same energy minimization procedure described above from sucrose. Graphics were done using Pymol (Schrodinger, Portland, OR, USA).

ACKNOWLEDGMENT

The authors are grateful to G. Cioci and S. Tranier in using the PICT facility for protein biophysical characterization, N. Monties for her help in the product analysis by HPAEC-PAD and HPAEC-PAD-MS of the ICEO high-throughput facility, devoted to the engineering and screening of new and original enzymes, of the Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (Toulouse, France). This work was supported by the French National Research Agency (ANR Project GLUCODESIGN, ANR-08-PCVI-002-02). This work was granted access to the HPC resources of the Computing Center of Region Midi-Pyrénées (CALMIP, Toulouse, France).

ABBREVIATIONS

*Np*AS, *Neisseria polysaccharea* amylosucrase; DNA; deoxyribonucleic acid; IPTG, isopropyl α-D-thiogalactopyranoside; PCR, Polymerase Chain Reaction; LB, Lysogeny Broth; DNS, DiNitroSalycilic acid; HPLC, High Performance Liquid Chromatography; HPAEC-PAD, High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection; GST, Gluthation – S- Transferase; EDTA, EthyleneDiamineTetraAcetic acid; DTT, DiThioThreitol; RI, Refractive Index; UV,Ultra Violet; DP, Degree of Polymerization; GH, Glycoside-Hydrolase; Tm, Melting Point; OD; Optical Density; K_m; Michaelis dissociation constant; k_{cat}, catalytic constant; MSQ, Mass Spectrometry Quadrupole; RMS, Root Mean Square; TFA, TriFluoroacetic Acid.

REFERENCES

(1) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233.

(2) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Res.* **2014**, *42*, D490.

(3) McCarter, J. D.; Withers, S. G. *Curret Opin. Struct. Biol.* **1994**, *4*, 885.

(4) Potocki de Montalk, G.; Remaud-Siméon, M.; Willemot, R. M.; Planchot, V.; Monsan, P. *J. Bacteriol.* **1999**, *181*, 375.

(5) Potocki de Montalk, G.; Remaud-Siméon, M.; Willemot, R. M.; Sarçabal, P.; Planchot, V.; Monsan, P. *FEBS Lett.* **2000**, *471*, 219.

(6) Daudé, D.; Champion, E.; Morel, S.; Guieysse, D.; Remaud-Siméon, M.; André, I. *ChemCatChem* **2013**, *5*, 2288.

(7) Andre, I.; Potocki-Veronese, G.; Morel, S.; Monsan, P.; Remaud-Siméon, M. *Top. Curr. Chem.* **2010**, *294*, 25.

(8) Malbert, Y.; Pizzut-Serin, S.; Massou, S.; Cambon, E.; Laguerre, S.; Monsan, P.; Lefoulon, F.; Morel, S.; André, I.; Remaud-Siméon, M. *ChemCatChem* **2014**, *6*, 2282.

(9) Van der Veen, B.; Potocki-Véronèse, G.; Albenne, C.; Joucla, G.; Monsan, P.; Remaud-Siméon, M. *FEBS Lett.* **2004**, *560*, 91.

(10) Champion, E.; Andre, I.; Moulis, C.; Boutet, J.; Descroix, K.; Morel, S.; Monsan, P.; Mulard, L. A.; Remaud-Siméon, M. *J. Am. Chem. Soc.* **2009**, *131*, 7379.

(11) Albenne, C.; Skov, L. K.; Mirza, O.; Gajhede, M.; Potocki-Veronese, G.; Monsan, P.; Remaud-Siméon, M. *FEBS Lett.* **2002**, *527*, 67.

(12) Cambon, E.; Barbe, S.; Pizzut-Serin, S.; Remaud-Siméon, M.; André, I. *Biotechnol. Bioeng.* **2014**, *111*, 1719.

(13) Thèse de doctorat, INSA de Toulouse, Daudé, D.; Remaud-Siméon, M.; André, I. Molecular Evolutionary Perspectives of Amylosucrases: From Substrate Promiscuity to Tailored Catalysis, **2013**.

(14) Champion, E.; Guérin, F.; Moulis, C.; Barbe, S.; Tran, T. H.; Morel, S.; Descroix, K.; Monsan, P.; Mourey, L.; Mulard, L. A.; Tranier, S.; Remaud-Siméon, M.; André, I. *J. Am. Chem. Soc.* **2012**, *134*, 18677.

(15) Vergès, A.; Cambon, E.; Barbe, S.; Salamone, S.; Le Guen, Y.; Moulis, C.; Mulard, L. a.; Remaud-Siméon, M.; André, I. *ACS Catal.* **2015**, *5*, 1186.

(16) Skov, L. K.; Mirza, O.; Henriksen, a; De Montalk, G. P.; Remaud-Siméon, M.; Sarçabal, P.; Willemot, R. M.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2001**, *276*, 25273.

(17) Albenne, C.; Skov, L. K.; Mirza, O.; Gajhede, M.; Feller, G.; D'Amico, S.; André, G.; Potocki-Véronèse, G.; Van der Veen, B.; Monsan, P.; Remaud-Siméon, M. *J. Biol. Chem.* **2004**, *279*, 726.

(18) Skov, L. K.; Mirza, O.; Henriksen, A.; De Montalk, G. P.; Remaud-Siméon, M.; Sarcabal, P.; Willemot,

R. M.; Monsan, P.; Gajhede, M. *Acta Crystallogr. Sect.* D **2000**, *56*, 203.

(19) Yamasaki, Y.; Suzuki, Y. *J. Japanese Soc. Starch* **1980**, *27*, 74.

(20) Kohmoto, T.; Fukui, F.; Takaku, H.; Machida, Y.; Arai, M.; Mitsuoka, T. *Bifidobact. Microflora* **1988**, *7*, 61.

(21) Patent, US 4518581 A, Miyake, T.; Yoshida, M.; Takeuchi, K. Imparting low- or anti-cariogenic property to orally usable products, **1985**.

(22) Côté, G. L. *Novel Enzyme Technology for Food Applications*; Elsevier, **2007**; pp. 243–269.

(23) Prapulla, S. G.; Subhaprada, V.; Karanth, N. G. *Adv. Appl. Microbiol.* **2000**, *47*, 299.

(24) Yamada, T.; Kimura, S.; Igarashi, K. *Caries Res.* **1980**, *14*, 239.

(25) Soro, R. Y.; Diopoh, J. K.; Willemot, R.-M.; Combes, D. *Enzyme Microb. Technol.* **2007**, *42*, 44.

(26) Tian, F.; Karboune, S. *J. Mol. Catal. B Enzym.* **2012**, *82*, 71.

(27) Canedo, M.; Jimenez-Estrada, M.; Cassani, J.; Lopez-Munguia, A. *Biocatal. Biotransformation* **1999**, *16*, 475.

(28) Díez-Municio, M.; de las Rivas, B.; Jimeno, M. L.; Muñoz, R.; Moreno, F. J.; Herrero, M. *Appl. Environ. Microbiol.* **2013**, *79*, 4129.

(29) Patent, CA 2378464 A1, Higashimura, Y.; Emura, K.; Kuze, N.; Shirai, J.; Koda, T. Fading Inhibitors, **2001**.

(30) Chung, C.-H.; Day, D. F. *J. Ind. Microbiol. Biotechnol.* **2002**, *29*, 196.

(31) Chung, C. H.; Day, D. F. *Poult. Sci.* **2004**, *83*, 1302.

(32) Patent, US 4782045 A, Machida, Y.; Fukui, F.; Komoyo, T. Promoting the proliferation of intestinal bifidobacteria, **1986**.

(33) Mäkeläinen, H.; Hasselwander, O.; Rautonen, N.; Ouwehand, A. C. *Lett. Appl. Microbiol.* **2009**, *49*, 666.

(34) Charalampopoulos, D.; Rastall, R. A. *Curr. Opin. Biotechnol.* **2012**, *23*, 187.

(35) Thompson, A.; Wolfrom, M. L. *J. Am. Chem. Soc.* **1951**, *73*, 5849.

(36) Sakano, Y.; Kogure, M.; Kobayashi, T.; Tamura, M.; Suekane, M. *Carbohydr. Res.* **1978**, *61*, 175.

(37) Gibson, G. R.; Probert, H. M.; Loo, J. Van; Rastall, R. A.; Roberfroid, M. B. *Nutr. Res. Rev.* **2004**, *17*, 259.

(38) Stanley, P.; Whelan, W. J.; Edwards, T. E. In *Journal of Chemical Society*; **1955**; pp. 355–359.

(39) Suzuki, J. *J. Japanese Soc. Starch Sci.* **1988**, *35*, 93.

(40) Kitahata, S. *J. Japanese Soc. Starch Sci.* **1990**, *37*, 49.

(41) Robyt, J. F.; Eklund, S. H. *Carbohydr. Res.* **1983**, *121*.

(42) Robyt, J. F.; Walseth, T. F. *Carbohydr. Res.* **1978**, *61*, 433.

(43) Andreotti, G.; Giordano, A.; Tramice, A.; Mollo, E.; Trincone, A. *J. Biotechnol.* **2006**, *122*, 274.

(44) Studier, F. W. *Protein Expr. Purif.* **2005**, *41*, 207.

(45) Mirza, O.; Skov, L. K.; Remaud-Siméon, M.; Potocki De Montalk, G.; Albenne, C.; Monsan, P.; Gajhede, M. *Biochemistry* **2001**, *328*, 9032.

(46) Skov, L. K.; Mirza, O.; Sprogøe, D.; Dar, I.; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2002**, *277*, 47741.

(47) Guérin, F.; Barbe, S.; Pizzut-Serin, S.; Potocki-Véronèse, G.; Guieysse, D.; Guillet, V.; Monsan, P.; Mourey, L.; Remaud-Siméon, M.; André, I.; Tranier, S. *J. Biol. Chem.* **2012**, *287*, 6642.

(48) Monthieu, C.; Guibert, A.; Taravel, F. R.; Nardin, R.; Combes, D. *Biocat. Biotrans.* ***2003***, *21*, *7*.

Chapter IV:

Isolation and characterization of an efficient mutant of *Neisseria polysaccharea* amylosucrase for the synthesis of controlled size maltooligosaccharides.

This chapter will be focused on one mutant named 30H3, identified during screening of the semi-rational library as being highly active on sucrose as it showed more than 6-fold enhanced activity compared to wild-type enzyme. This mutant was further characterized at both biochemical and structural levels. The product profile from sole sucrose revealed a different behaviour from wild-type amylosucrase. Indeed, only soluble maltooligosaccharides of controlled size chains (DP between 2 and 21) with a narrow polydispersity were observed, without any production of insoluble amylose. Sequencing of this mutant revealed the presence of 9 mutations in the active site. Its X-ray structure was determined and further investigated by molecular dynamics to understand the molecular origins of its particular behaviour toward sucrose and absence of polymerization activity.

INTRODUCTION

Amylosucrases (AS, E.C. number 2.4.1.4) are α-transglycosylases that belong to Glycoside-Hydrolase (GH) family 13, a class of catalysts that mainly contains enzymes involved in starch transformation. However, instead of degrading starch, amylosucrases use sole sucrose to synthesize an insoluble amylose-like polymer (with an average molecular mass of 8,941g/mol). They also produce soluble maltooligosaccharides of varying distributions and sizes (going up to a degree of polymerization of 25) depending on the enzyme origin[1–3]. In addition, side reactions may also occur by transglucosylation of fructose released during the course of the reaction, to form sucrose isomers, such as turanose (α-D-glucosylpyranosyl-1,3-D-fructofuranose) and trehalulose (α-D-glucosylpyranosyl-1,1-D-fructofuranose).

To date, seven amylosucrases have been characterized from *Neisseria polysaccharea*[4], *Deinococcus radiodurans*,[2] *Deinococcus geothermalis*,[1,5] *Alteromonas macleodii*[6], *Arthrobacter chlorophenolicus*[7], *Methylobacillus flagellatus*[8] and *Synechococcus sp*[9]. Only three of these amylosucrases from *N. polysaccharea* (*Np*AS), D. *geothermalis* (*Dg*AS) and *D. radiodurans* (*Dr*AS) had their three-dimensional structure determined by X-ray crystallography either in apo form or in the presence of various ligands[10,11–15,16,5]. Their structure is organized in five domains, namely A, B, C, N and B'. Domains N and B' are only found in amylosucrases, whereas the three other domains are conserved within GH13 enzymes. The 3D-structures have shown that catalytic residues are located at the bottom of a very narrow and constrained pocket whose topology is essentially due to the presence of a salt bridge at the bottom that prevents access to negative subsites and the B' domain that covers up the active site and shields it from external environment. Thanks to the co-crystallization of *Np*AS with sucrose and maltoheptaose, a detailed amino acid mapping of the various binding sites was rendered possible. In particular, three Oligosaccharide Binding sites (OB) were identified, namely OB1, corresponding to the catalytic site, OB2 at the surface of the B' domain and OB3 at the surface of the C domain. For each oligosaccharide site, the multiple binding subsites were characterized in details and the amino acid residues involved in interactions with the oligosaccharides were identified. Importantly, it was suggested that OB1 and OB2 could be connected in order to favor polymer elongation and glycogen activation with the assistance of a movement of the amylosucrase-specific B' domain[17].

Structural knowledge combined with molecular modelling has helped to better understand the mode of action of these enzymes and more particularly, elucidate the molecular determinants involved in substrate and product specificity of the various amylosucrases[18]. This information used in combination with mutagenesis already led to the re-engineering of amylosucrase properties, in particular product specificity. *Np*AS is by far the most extensively investigated amylosucrase. Introduction of single amino acid mutations at the various binding subsites of OB1 and OB2 has been shown to affect drastically the hydrolysis/tranglycosylation capability of the enzyme. Some of these single mutants turned out to be either pure sucrose-hydrolases, having lost their transferase ability (such as mutant D394A), or specific of the synthesis of very short oligosaccharides not longer than DP3 without any polymerase activity (such as mutants R415A or R446A), or producing preferentially high molecular weight amylose (such as mutants R226A and R226N)[18]. Whenever maltooligosaccharide production was observed, the size distribution was generally found quite

polydisperse. In the current study, we analyzed further a very efficient sucrose-utilizing mutant of *Np*AS that was isolated from a semi-rational library of ~2.7x10^4 variants of *Np*AS targeting 23 positions of subsites +1 and +2. This mutant was found to contain 9 mutations and it revealed not only a considerably higher activity on donor sucrose compared to the parental wild-type enzyme but also an unique ability to efficiently produce short chain soluble maltooligosaccharides (DP $\leq$ 21) of very narrow polydispersity, never observed before for any amylosucrase, and without any synthesis of insoluble amylose. Biochemical and structural characterization of this unique mutant combined with molecular modelling studies provided insightful information to understand molecular determinants possibly at the origin of this novel product specificity.

RESULTS AND DISCUSSION

Isolation of sucrose-active mutants.

A semi-rational library of *Np*AS mutants was recently constructed to search for enzymes having a novel acceptor specificity toward a partially protected β-linked disaccharide[19]. During the evaluation of the library, a primary screening was carried out to identify the enzymes that had kept their ability to utilize sucrose as substrate. Altogether, 55 active mutants on sucrose were isolated by DNS assays performed in microplates and analyzed with respect to their product profiles. This led to the identification of 38 transglucosylases (out of 55) that contain between 5 and 17 mutations. Among others, this analysis revealed an interesting mutant, named 30H3, which displayed a 6.6-fold enhanced activity on sucrose compared to that of wild-type *Np*AS based on the DNS assay. Next, it was produced and purified to homogeneity in order to proceed to its characterization. Remarkably, the sequence of mutant 30H3 revealed the presence of 9 mutations at subsites +1 and +2 of the active site, targeting amino acid positions known from X-ray studies to be tightly involved in oligosaccharide recognition by wild-type enzyme (Table 1).

Table 1. Sequence of the selected mutant in comparison to the wild-type enzyme. *The 23 targeted positions are shown. In red, are highlighted mutated positions.*

Amino acid positions		wild-type	30H3
226	Loop 3 (domain B)	R	K
227		E	E
228		I	V
229		F	F
289		A	I
290	Loop 4 (domain A)	F	Y
300		E	I
331		V	T
332	Loop 5 (domain A)	H	H
396		G	G
397	Loop 7 (Domain B')	W	W
398		T	T
434		V	V
435		P	P
436		F	F
437		Q	S
438		Y	Y
439		N	D
444		D	D
445		C	A
446		R	R
447		V	V
448		S	S
Number of mutations		-	9

To compare the comportment of this mutant with that of the wild-type enzyme, their sucrose consumption was followed during the reaction course. In the presence of 0.3mg/mL purified enzyme (30H3 or *Np*AS), sucrose consumption (146mM) was monitored by HPLC over 24h. As shown in Figure 1, mutant 30H3 consumed 96% sucrose in less than 2h while the wild-type enzyme needed nearly 7h. The sucrose consumption enhancement was estimated by linear regression between 0 and 40min, revealing a 5.6-fold increase compared to wild-type enzyme.

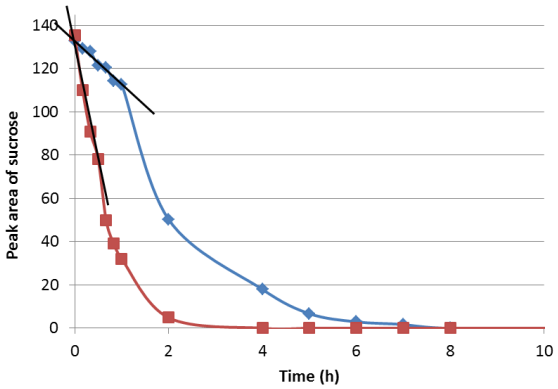


Figure 1. Sucrose consumption kinetics on sucrose catalyzed by mutant 30H3 (red) and wild-type *Np*AS (blue) followed by HPLC *(using 0.3mg/mL purified enzyme and 146mM sucrose).*

Product profiles.

To analyze further the transglycosylating aptitude of mutant 30H3, reactions were carried out in the presence of 146mM sucrose at 30°C (standard conditions). The reaction mixture taken after total sucrose depletion was analyzed by HPAEC-PAD. The chromatograms revealed the presence of short maltooligosaccharides (MOS) up to a maximum DP of 21 and traces of other oligosaccharides (DP 18 on the figure becase of zoom accuracy). By comparison, in similar conditions, wild-type *Np*AS produces MOS of DP comprised between 2 to 90 which are distributed in the soluble or insoluble fraction for the smallest and the highest DP, respectively (Figure 2A). Surprisingly, the mutant was unable to produce insoluble high molecular weight amylose-like polymer. No insoluble glucans were detected for mutant 30H3 while wild-type *Np*AS incorporates 22% of the glucosyl from sucrose in highly crystalline amylose (DP > 17). The influence of sucrose concentration (146, 300, 400 and 700mM) on MOS production was then investigated. Product profiles obtained at these concentrations are shown in Figure 2B. In all cases, the maximum DP never exceeded 21 indicating no impact of the sucrose concentration on product size. However, it is noteworthy that with increasing sucrose concentrations, the amounts of oligosaccharides showing retention times between those of MOS (indicated by blue arrows in Figure 2) increased noticeably. Although these products have not been characterized, one could hypothesize that they could correspond to polyglucosylated forms of sucrose or sucrose isomers but this remains to be confirmed.

The number-average DP (DPn), weight-average DP (DPw) and the polydispersity of the products synthesized at the different initial sucrose concentration were determined. As shown in Table 2, the DPn (between 4.8 and 7.0) and DPw values (between 6.8 and 8.9) of the products synthesized with mutant 30H3 are considerably lower than those of the products synthesized with wild-type *Np*AS: DPn = 24, 19, 17 and DPw = 73, 50, 39 at 100, 300 and 600mM sucrose,[20] respectively. In addition, the polydispersity of the products synthesized by mutant 30H3, which is around 1.3, is twice lower than that previously reported for the products of the wild-type enzyme [20] (Table 2). Mutant 30H3 is thus specialized for the production of a soluble short chain MOS with an average DP around 6.8 and 8.9 (depending on the sucrose concentration) and a remarkably low polydispersity. Such comportment is totally unusual and was never observed for wild-type *Np*AS.

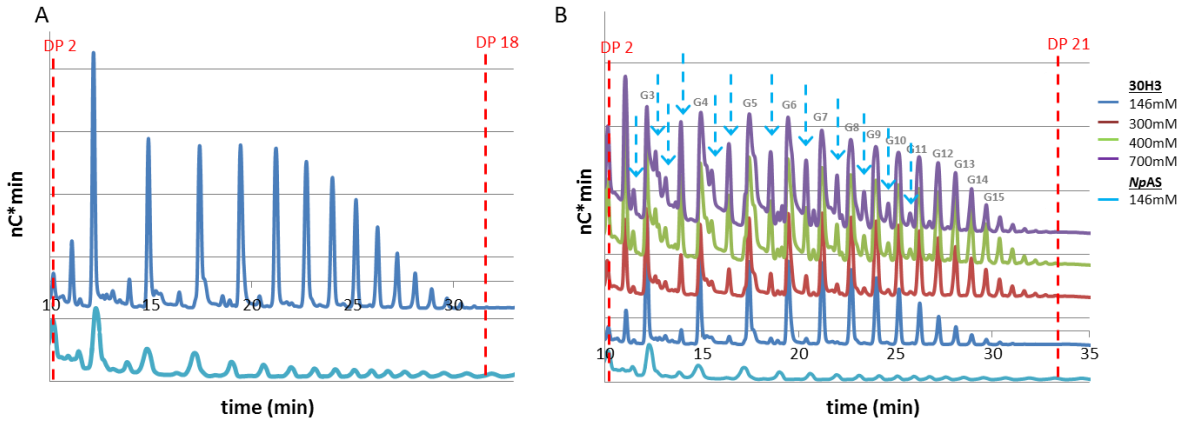


Figure 2. Comparison of crude extract profiles analyzed by HPAEC-PAD ; *(A) mutant 30H3 (blue) and* Np*AS (cyan) diluted 5- fold in the presence of 146mM sucrose at 30°C ; (B) mutant 30H3 in the presence of 146mM (blue), 300mM sucrose (red), 400mM sucrose (green) and 700mM sucrose (purple) diluted 5, 21, 14 and 16- fold, respectively.* Np*AS is represented in cyan and it was diluted 5-fold. Blue arrows indicate new products in comparison to the* Np*AS. Gn, maltooligosaccharides of DP n.*

Table 2. Average DP and Polydispersity of products formed by mutant 30H3 using 146mM, 300mM, 400mM and 700mM sucrose. *In brackets is given data for* Np*AS for 100mM, 300mM and 600mM sucrose.*

	$\overline{DP_n}$	$\overline{DP_w}$	Polydispersity
146mM	4.8 (24-100mM)	6.8 (73-100mM)	1.4 (3.0-100mM)
300mM	7.0 (19-300mM)	8.9 (50-300mM)	1.3 (2.6-300mM)
400mM	6.2	8.1	1.3
700mM	6.2 (17-600mM)	7.9 (39-600mM)	1.3 (2.3-600mM)

The products obtained after 24h reaction at 30°C using different initial sucrose concentrations were quantified. The percentage of glucosyl units incorporated, from sucrose, in each type of products is reported in Figure 3. From 146mM sucrose, mutant 30H3 synthesizes higher amounts of turanose than wild type *Np*AS. Around 33% of the glucosyl units are converted into MOS of DP ranging from 3 to 8 for mutant 30H3 versus 14% for *Np*AS and 28.9% in MOS of DP 9 to 17 for mutant 30H3 versus 12.4% for *Np*AS. Of note, oligosaccharides that do not belong to the MOS series are produced in trace amounts, at this sucrose concentration. At 300mM sucrose, turanose formation is reduced, MOS of higher DPs (8 to 21) are produced in higher amounts, what is reflected by the weight average DP estimated to be of 8.9 (Table 2). This effect is no longer observed at higher sucrose concentration (400 and 700mM). These initial sucrose concentrations favor the production of turanose as well as that of the oligosaccharides not belonging to the MOS, supporting our hypothesis that this new series of oligosaccharides may result from a transfer of the glucosyl units from sucrose onto either sucrose itself or on sucrose isomers. Consequently, the formation of MOS is reduced as well as their weight-average DP. In addition, the ratio between turanose and trehalulose was close to 5 for all concentrations, except for 146mM (equal to 17) where trehalulose production was very low. Furthermore, the increase of sucrose concentration leads to a higher amount of sucrose isomers (Figure 3).

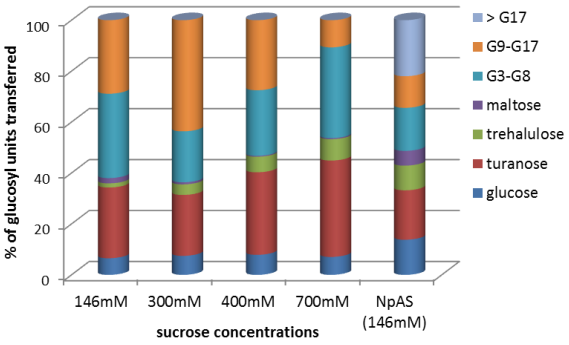


Figure 3. Yields of glucosyl units incorporated into various products synthesized in the presence of 146, 300, 400 and 700mM sucrose by mutant 30H3 and wild-type *Np*AS enzyme.

Determination of kinetic parameters and stability.

Steady-state kinetic parameters (k_{cat}, K_m, V_m, k_{cat}/K_m) were determined for the purified mutant 30H3 in the presence of varying sucrose concentrations ranging from 25mM to 200mM. Kinetic data obtained for the mutant in the presence of sole sucrose did not reveal any enhancement of catalytic efficiency, k_{cat} or K_m (Table 3). Obviously, at the initial reaction step, mutant 30H3 is less efficient than the wild-type *Np*AS, although over the entire duration of the reaction, it consumed sucrose much faster (Figure 1). This behavior may be the result of a possible auto-activation by the released products, which can clearly be hypothesized to explain the 5.6-fold enhancement of sucrose consumption observed in the first 40 min of the reaction. Such a kinetic comportment was previously reported for the mutant R226N of *Np*AS[21] (Cambon *et al)*. However this mutant synthesized a higher amount of insoluble MOS compared to *Np*AS and mutant 30H3. This suggests that the mutations introduced in mutant 30H3 greatly favors elongation of short MOS but prevents to some extent an extensive elongation, changing the polymerase into an oligosaccharide synthase.

Table 3. Determination of apparent kinetic parameters and thermostability values of mutant 30H3 and wild-type *Np*AS.

	k_{cat}/Km ($mM^{-1}s^{-1}$)	Vm_{app} (µmol/min/mg)	k_{cat} (s^{-1})	Km_{app} (mM)	Tm (°C)
30H3	0.013	1.35	1.61	128.3	46.1
Np*AS	0.026	1.1	1.30	50.2	49.0

*Data from Potocki de Montalk *et al*., 1999, for sucrose > 20mM.

Structural insight by X-ray crystallography and molecular modelling studies.

Mutant 30H3 was crystallized in conditions similar to those previously described for the crystallization of the wild-type enzyme (PEG 6000 30%, HEPES 0.1M pH 7.3)[10]. The 3D structure of the mutant 30H3 solved at 2.0 Å (Table 4) is highly similar to that of the wild-type *Np*AS (Figure 4A, B) with a rmsd calculated on 628 Cα atoms of 0.3Å. As can be seen in Figure 4 A, a molecule of Tris (a component of the elution buffer and of the prescission protease buffer) was found into the active site. This was already observed in X-ray structures of wild-type *Np*AS. It was suggested that Tris molecule (and other ligands), when bound in the active site, could stabilize loops forming the pocket topology and facilitates crystallization of amylosucrases[22]. Crystallographic data also revealed the presence of different ligands in the structure (Figure 4A). Three sucrose molecules were found respectively bound at SB2 and SB3[14] binding sites previously described by Skov *et al*. and a third one never reported before that we will call SB4. Both SB3 and SB4 sites are found at the surface of the enzyme. SB4 is also positioned at the interface between two-symmetry-related molecules. A maltotriose molecule was also found bound at OB2 site[14].

Table 4. Data collection and refinement statistics for mutant 30H3. *Values in parentheses are for the outer resolution shell.*

Data collection		Refinement	
a. b. c (Å)	60.0, 97.6, 115.2	**Resolution (Å)**	74.4 – 2.0
Space group	$P22_12_1$	**Nb. of unique reflections**	45528 (2780)
α. β. γ (°)	90.0, 90.0, 90.0	R_{work}/R_{free} (%)	14.80/21.30
Resolution (Å)	74. 4 - 2.0, (2.05 – 2.0)	**Total of number of atoms**	6063
R_{sym}	0.090 (0.451)	**Nb of residues in the protein**	628
I/σI	15.5 (2.85)	**Nb of ligand molecules**	1 Tris, 3 PEG, 3 sucrose, 1 G3
Redundancy	4.0 (3.4)	**Nb of water molecules**	810
Completeness (%)	98.2 (82.4)	**Rmsd bond lengths (Å)**	0.0193
Nb of molecules /AU	1	**Rmsd bond angles (°)**	1.904
Matthews coefficient ($Å^3$/Da)	2.3	**B-factors average ($Å^2$)**	19
Observations	1 Tris bound into the active site 3 binding sites of sucrose on surface 1 binding site of maltotriose on surface		

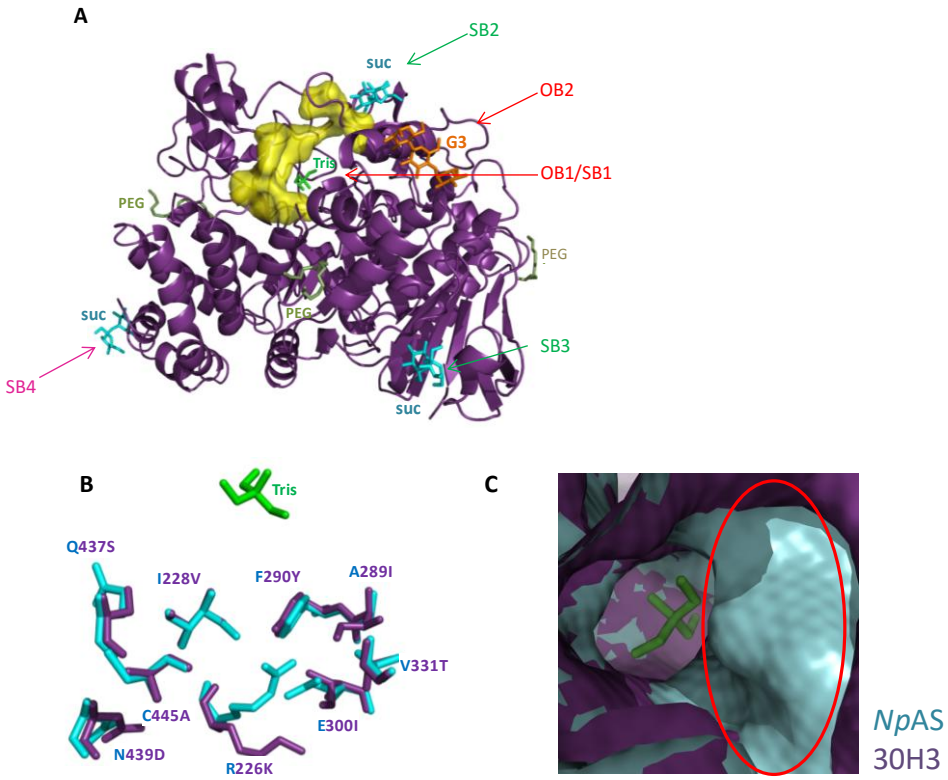


Figure 4. View of the three-dimensional structure of mutant 30H3 (purple) superimposed onto structure of wild-type *Np*AS (in cyan, PDB code: 1G5A)[13]. ***A****: Overall organization of the mutant 30H3 with a tris molecule (green stick) bound into the active site, PEG (dark green sticks), three sucrose (cyan sticks), one maltotriose (G3) (orange sticks) and the 9 mutations around the catalytic pocket (yellow molecular surface).* ***B:*** *Superimposition of active site residues at mutated positions 226, 228, 289, 290, 300, 331, 437, 439 and 445. Wild-type structure is shown as cyan sticks and mutant 30H3 in purple. A tris molecule is present at the same position in the active site of both enzymes (shown as green sticks).* ***C:*** *Molecular surface representation of mutant 30H3 colored in purple superimposed onto wild type colored in cyan and, bound to sucrose (in green) derived from structure PDB: 1G5A.*

Amongst the 9 mutations identified by sequencing, R226K, I228V are carried by loop 3; A289I, F290Y, E300I by loop 4; V331T by loop 5 and Q437S, N439D and C445A are carried by loop 7 (B' domain) (Table 1). These mutations introduced at subsites +1 and +2 of the active site (Figure 4A) led to a slight opening of the catalytic pocket as seen in Figure 4C, mainly due to mutations at positions 226 and 300 that eliminated the R226-E300 salt bridge interaction present in wild-type *Np*AS and as consequence, uncluttered the active site. Unfortunately, all attempts to co-crystallize mutant 30H3 with sucrose and maltooligosaccharides were unsuccessful. Therefore, in order to shed some light on the impact of the mutations onto the mutant specificity for the production of short chain MOS, we superimposed the structure of mutant 30H3 onto that of wild-type enzyme in complex with maltoheptaose (PDB code: 1MW0). Interestingly, another mutant named 39A8 was recently characterized (Vergès *et al.*, in preparation). Like mutant 30H3, this mutant contains also a high number of mutations at subsites +1 and +2 (11 mutations). However, these mutants only differ by the type of amino acid residues at positions 396, 398 and 445 from the B' domain. Whereas mutant 39A8 contains G396S, T398V, and C445R mutations, the mutant 30H3 kept the G396 and T398 amino acid residues found in wild-type *Np*AS. Therefore, only C445A mutation is

specific to mutant 30H3 which was changed into C445R in mutant 39A8. Although amino acid differences were minor between the two mutants, drastic differences were observed in the products synthesized compared to the wild-type *Np*AS. One of the most striking differences is the absence of high molecular weight polymer for both mutants. Inspection of Cys445 structural environment suggested that residue mutation could lead to changes in the dynamics of hairpin loop 7 (G433-G449) encompassed in domain B' and affect binding of oligosaccharides at OB2 site (Figure 5A). Interestingly, the sole X-ray structure of wild-type *Np*AS in complex with a complete maltoheptaose molecule bound at OB2 site revealed an oxidized C445, covalently bound to a DTT molecule which was itself engaged in an ionic interaction with R425. Such modification of C445 likely restricted movement of the loop carrying R425, which acts as a binding platform in OB2 site, and as a result helped crystallization of the wild-type enzyme in complex with a maltoheptaose molecule bound at OB2 site. The role of loop 7 dynamics in polysaccharide elongation by connecting OB1 and OB2 sites was previously proposed by molecular modelling[23]. Amino acid substitution at C445 position, like in mutants 30H3 and 39A8, could thus be assumed to alter flexibility of domain B' and therefore affect the connection between OB1 and OB2 site formation, what could be responsible for the suppression of polymer formation for both mutants. To investigate further the impact of mutations on dynamics of the enzymes, we carried out MD simulations in same conditions (50ns at 298K) on *Np*AS and mutant 30H3 in complex with sucrose. In Figure 5B are shown snapshots taken at the end of the simulation. They indicate larger conformational changes of parts of the B' domain, in particular of region G396-G428, forming a short α-helix turn. Interestingly, analysis of simulated B-factors during the course of MD simulations showed a much higher flexibility of this region (G396-G428) in wild-type *Np*AS than in mutant 30H3 (Figure 5C). More detailed analysis revealed that this region carries essential residues for catalysis, Asp393 which stabilizes the transition state formed with sucrose during the course of the reaction and Asp394 which has been found to stabilize fructosyl (O_6H) from sucrose in crystallographic structure of *Np*AS in complex with sucrose. By lowering flexibility of this region, amino acid substitutions introduced in mutant 30H3 might have lowered energy barrier corresponding to transition state formation with sucrose, what is in agreement with slightly improved k_{cat} values for V_m values observed for mutant 30H3 in comparison with wild-type *Np*AS (Table 3). Another main difference in products synthesized by mutants 30H3 and 39A8 resides in the size of the soluble MOS synthesized by the two enzymes. Whereas mutant 39A8 was found unable to synthesize oligosaccharides longer than DP3, mutant 30H3 produces a series of MOS of DP 3 to 21. Looking closely at subsites +2 and +3, we observed that introduction of G396S and T398V mutations in mutant 39A8 alters the hydrogen bonding interaction with Glc(O_6H) at subsite +3 reported for wild-type *Np*AS. They also sterically hinder accommodation of glucopyranosyl moiety at +3 subsite in longer maltooligosaccharide chains. In contrast, mutant 30H3 kept the wild-type amino acid residues at these two positions, that is to say G396 and T398, what could explain why this mutant conserved its ability to elongate MOS compared to 39A8 (Figure 6). Additional mutations introduced at positions 226, 228, 289, 290, 300 and 331 from subsites +1 and +2 are surely responsible for differences observed in the polydispersity of the oligosaccharide profiles obtained for mutant 30H3 and wild-type *Np*AS, although molecular explanations remain still obscure. Interestingly, position 226 is mutated into

a lysine residue in mutant 30H3. The effect of mutations at this position was recently thoroughly investigated[21]. No matter which amino acid is introduced at this position, activity on sucrose was always found enhanced compared to wild-type *Np*AS and their ability to produce insoluble amylose was conserved[17,18]. That was also the case for R226K mutant showing that this single mutation alone did not affect ability of the enzyme to synthesize amylose. On the contrary, no polymer formation was detected for mutant 30H3, indicating that the combination with the other 8 mutations suppressed the ability to form high molar mass amylose usually observed for single mutants at position 226. In particular, we suggest that polymer elongation could be impeded due to mutation C445A that could alter conformational dynamics of domain B' (G396-D460) and as a result, could affect polysaccharide binding onto OB2 site and prevent connecting OB1 to OB2. Conversely, oligosaccharide elongation in OB1 was still possible, and even facilitated, due to the aperture of the active site upon introduction of R226K, E300I, I228V, A289I, F290Y and V331T mutations.

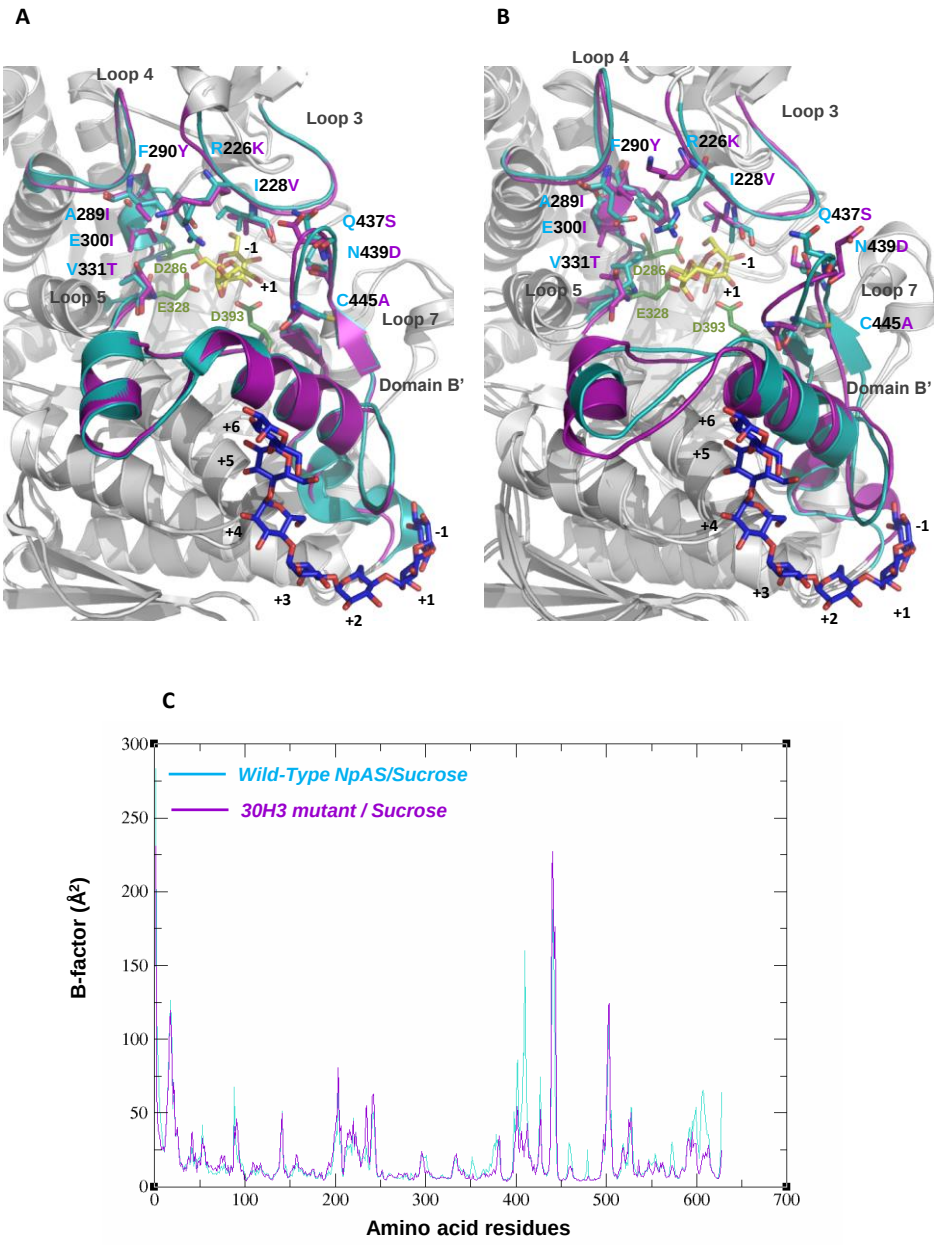


Figure 5. Comparison of *Np*AS (in cyan colour) and mutant 30H3 (coloured in magenta). ***A:*** *Superimposition of X-ray structure of wild-type* Np*AS (PDB code: 1MW0) onto structure of mutant 30H3. For reference, sucrose molecule bound at SB1 site (or OB1) was extracted from structure of* Np*AS in complex with sucrose (PDB code: 1JGI). The maltoheptaose molecule bound at OB2 site is extracted from structure of* Np*AS in complex with maltoheptaose (PDB code: 1MW0). All mutations are indicated on the figure and domain B' is highlighted. In green sticks are shown catalytic residues.* ***B:*** *Superimposition of a snapshot taken at the end of the 50ns MD simulations carried out at 298K on wild-type* Np*AS in complex with sucrose and mutant 30H3 in complex with sucrose. Maltoheptaose bound in OB2 site is shown for reference. It is extracted from x-structure of* Np*AS in complex with maltoheptaose (PDB code: 1MW0). Conformational changes of B' domains during MD simulation are highlighted. In green sticks are shown catalytic residues.* ***C:*** *Calculated B-factors of amino acid residues from 50ns MD simulations carried out at 298K on wild-type* Np*AS and mutant 30H3 in complex with sucrose represented using the same colour code as in panels A and B.*

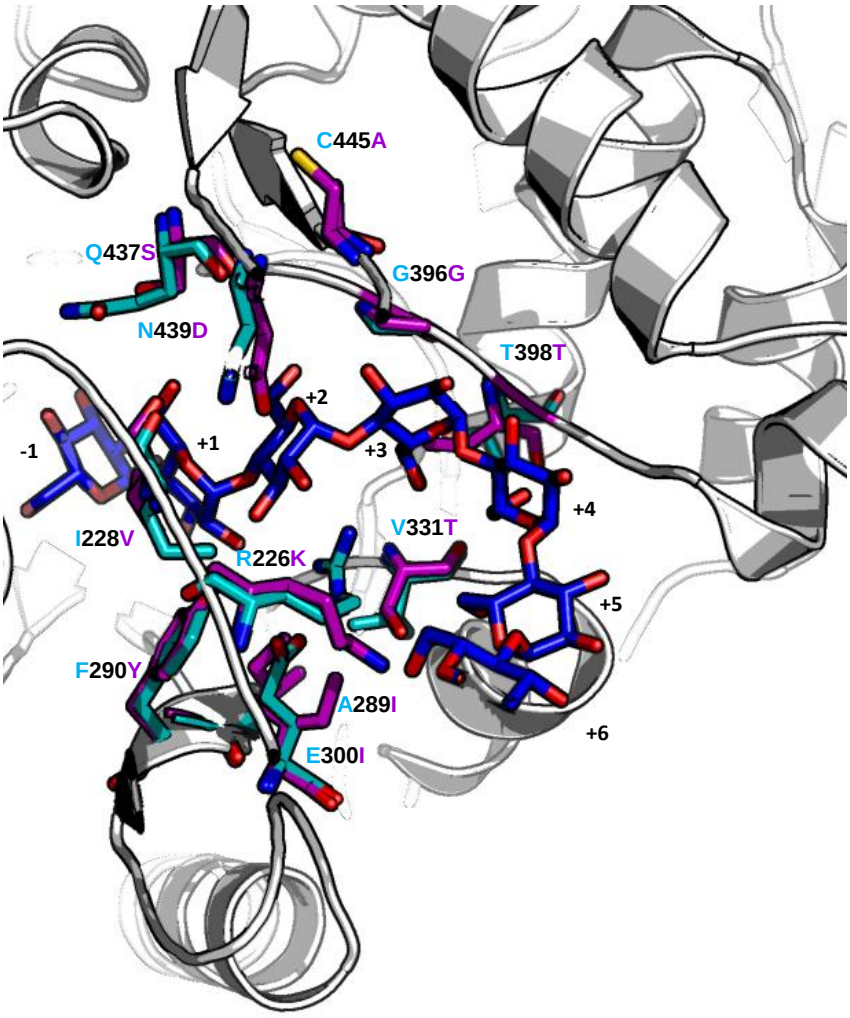


Figure 6. Zoom on active sites of *Np*AS (in cyan colour) and mutant 30H3 (coloured in magenta). *Targeted residues are shown in sticks. For reference, bound maltoheptaose in OB1 site is shown. It was extracted from X-ray structure of complex between* Np*AS and maltoheptaose. Numbering of subsites is shown going from -1 at catalytic site to +1 at the surface of the enzyme.*

CONCLUSIONS

In conclusion, screening of the semi-rational library constructed in the first place to tailor a novel specificity towards a non-natural β-linked disaccharide acceptor, which is a potential intermediate in the synthesis of *Shigella flexneri* cell-surface oligosaccharides, has led to the isolation of several mutants displaying unique catalytic properties and product specificities, among which mutant 30H3, disclosed herein. Indeed, this is the first report of an efficient amylosucrase mutant that produces from sole sucrose, a distribution of short chain from DP 3 to DP 21 and range over a very narrow polydispersity compared to wild-type amylosucrases. Furthermore, in spite of 9 mutations introduced in the active site, the mutant is highly active on sucrose substrate. Finally, this mutant only produces soluble oligosaccharide as no insoluble high molecular weight amylose was seen. This mutant is thus an attractive enzymatic tool that could offer interesting opportunities for the design of amylodextrins with controlled size.

THEORETICAL AND EXPERIMENTAL PROCEDURES

Bacterial strains, plasmids & chemicals.

Plasmid pGST-amylosucrase (pGST-AS) derived from the pGEX-6P-3 (GE Healthcare Biosciences. Piscataway, U.S.A) encoding GST fused to *N. polysaccharea* amylosucrase gene was used for the construction of semi-rational libraries. *E. coli* TOP 10 electrocompetent cells (Invitrogen. Carlsbad, U.S.A.) were used as the host for plasmid transformation and production of recombinant amylosucrase variants. *Dpn*I, EcoRI HF, NotI HF restriction enzymes, T4 DNA ligase and Antartic Phosphatase were purchased from New England Biolabs (Beverly, MA, U.S.A.). Phusion DNA-polymerase was purchased from Finnzymes (Espoo, Finland). Ampicillin, lysozyme and isopropyl-β-D-thiogalactopyranoside (IPTG) were purchased from Eurodemex (Souffelweyersheim, France); Bromothymol blue ordered from Sigma – Aldrich (St Louis, MO, U.S.A.). Oligonucleotides were synthesized by Eurogenetec (Liege, Belgium). DNA plasmid extraction (GenElute HP plasmid Miniprep kit, Quiagen. Chatsworth, CA, U.S.A.), DNA purification and gel extraction (GenElute Gel Extraction kit) columns were purchased from Sigma-Aldrich (St Louis, MO, U.S.A.). DNA sequencing was performed by GATC Biotech AG. (Constance, Germany).

Library construction.

Mutant 30H3 was isolated from computer-aided libraries designed for the recognition and glucosylation of a partially protected β-linked disaccharide acceptor called allyl 2-deoxy-2-*N*-trichloroacetyl-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside[19]. Among the screened 63,000 variants, 58 were isolated as being active on sucrose by a primary solid colorimetric screening based on pH shift[24].

Library screening in microplate format.

The 58 variants were stored in 96-well microplates containing LB medium (200µL), ampicillin (100µg/mL^{-1}) and glycerol (12% w/v). For expression of mutants, these microplates were used to inoculate 96-Deep Well plates. Protocol is well described in the following references[19] (Vergès *et al.* in preparation). After expression of all enzymes contained into supernatants, sucrose consumption was evaluated by mixing 50µL enzyme extract with 50µL sucrose at 292mM during 1h at 30°C under agitation at 450 Rpm. Reactions were stopped by addition of 100µL DNS and heated at 95°C during 10min. After dilution by 5-fold into water for a final volume of 200µL, measurement of reducing sugars release was performed by reading the absorbance at 540nm. 55 mutants were confirmed active on sucrose and their activities were normalized with respect to the wild-type value. A plot of all activities is given in a prior article[19].

Characterization of mutant 30H3 and products.

Production and purification.

Production and purification of the mutant was performed as previously reported (Vergès et al., in preparation). Briefly, *E. coli* TOP10 electrocompetent cells containing genes of *Np*AS and mutant 30H3 were used to inoculate a pre-culture in LB medium with ampicillin (100µg/mL) and stored at 30°C overnight. Then, cultures were inoculated at $OD600_{nm}$ 0.2 and grown at 23°C in the same medium. When $OD600_{nm}$ reached 0.5, cultures were induced by IPTG at 1mM. After 17h incubation, the supernatant was harvested after centrifugation, sonication and a new step of centrifugation. The purification step consisted in the incubation of supernatants on Glutathion Sepharose 4B column

(Dutscher , Brumath, France) to isolate GST-*Np*AS wild-type and mutant 30H3 and by affinity chromatography using the AKTA Express system. The detailed protocol is given elsewhere *(Vergès* et al., in preparation*).* Purified enzymes were stored in Tris buffer (50mM Tris, 150mM NaCl. 1mM EDTA. 1mM DTT. pH7.0) and GST Tag cleaved by PreScission protease (Dutscher, Brumath, France). Protein concentration was estimated using Nanodrop ND-1000 spectrophotometer.

Kinetics of sucrose consumption.

0.3mg/mL of mutant 30H3 (2.8 U/mL) or *Np*AS (0.3 U/mL) were mixed to 146mM sucrose, at 30°C during 24h to follow sucrose depletion. Every 10 min during 1h, samples were withdrawn and heated at 95°C during 5min to stop the reaction. Then, samples were taken every 1h between 2h to 8h of the reactions and at 24h. After centrifugation (10min, 11,000 Rpm) of reaction mixes, supernatants were analyzed by HPLC using RI detection to measure sucrose consumption with a Biorad HPX-87K column and water as eluent (65°C, 0.6mL.min^{-1}).

Product profile.

1U/mL of purified enzyme (mutant 30H3 and wild-type *Np*AS) was mixed with 146mM sucrose at 30°C for 24h. Reactions were stopped by heating at 95°C during 5 min. Half of the mix was centrifuged for 10min at 11,000 Rpm to recover the supernatant for product analysis of the soluble fraction upon dilution in water by 3,500-fold. The remaining fraction corresponding to the crude extract (containing both the soluble and insoluble products) was solubilized in 1M KOH to a final 5-fold dilution. Both fractions were analyzed by HPAEC-PAD (Dionex Carbo-Pack PA100, 30°C, 1mL.min-1) through a gradient of sodium acetate (from 6 to 300mM in 28min) in 150mM NaOH for supernatants and with a gradient of sodium acetate from 6-500mM over 120min for crude extracts using a mobile phase of 150mM NaOH aq.. Detection was performed using a Dionex ED40 module with a gold working electrode and Ag/AgCl reference electrode. As quantitative analysis using HPAEC-PAD requires standards that are not available for maltooligosaccharides of DP>7, we estimated their concentration by difference between the molar amount of consumed sucrose and the molar amount of glucosyl residues into various products quantified using available standards (glucose, turanose, trehalulose, sucrose not consumed, G2, G3, G4, G5, G6 and G7).

Product characterization.

High-Performance Size Exclusion Chromatography (HPSEC) was performed on crude extract of the mutant 30H3 and wild-type *Np*AS after 24h reaction on 146mM sucrose at 30°C in the presence of 1U/mL enzyme. Globally, 100µL of crude extract was solubilized in 25µL NaOH 5M. Then, successive dilutions were performed in 187.5µL water and then neutralized by addition of 312.5µL HCl 0.4M for a final concentration of sugars at 8g/L. A series of Polysep-GFC P6000 and P2000 columns were used for product separation with a flow rate at 0.5mL/min of 8mM Tris – 200mM NaCl - pH7.0 at 50°C. Detection was performed using a refractometer (Shodex RI 101).

Molecular mass distribution.

Concentration of maltooligosaccharides from DP8 to DP 21 was approximated using the linear relationship existing between the detector response per mole obtained by HPAEC-PAD and the degree of polymerization (DP). A linear curve was drawn with standards from DP 2 to 7 to determine coefficients (a=513 and b=55.7) and thus extend the equation to higher DPs.

The number-average DP (DPn), weight-average DP (DPw) and polydispersity index (P) were calculated as follows:

$$\overline{DP}_n = \frac{\sum_i C_i}{\sum_i \frac{C_i}{DP_i}}$$

$$\overline{DP}_w = \frac{\sum_i C_i DP_i}{\sum_i C_i}$$

$$P = \frac{\overline{DP}_w}{\overline{DP}_n}$$

With C_i, product concentration; DP_i, degree of polymerization.

Influence of sucrose concentration.

In order to investigate the impact of sucrose concentration onto product distribution, reactions were performed at various sucrose concentrations (146mM, 300mM, 400mM and 700mM) with 1U/mL of purified enzyme (30H3, wild-type *Np*AS) at 30°C for 24h. Reaction mixes were stopped by heating at 95°C during 5min. As for standard reaction (146mM), supernatant and crude extracts were analyzed by HPAEC-PAD as described before (section "Product profile").

Kinetic parameters determination.

Initial velocity was determined by mixing pure mutant enzyme at 0.02 U/mL and 0.01U/mL at 30°C, under agitation on several sucrose concentrations (from 25 to 200mM). Reactions were stopped by heating at 95°C for 5 min at several intervals for a period corresponding to a maximum sucrose consumption equal to 10%. Mixes were centrifuged (10min, 11,000 Rpm) and the soluble part was analyzed by HPLC, to measure the fructose release using RI detection with a Biorad HPX-87K column and water as eluent (65°C, 0.6mL.min^{-1}).

Stability studies on enzymes.

The melting point (Tm) of amylosucrases (mutant 30H3 and *Np*AS) were determined by Circular Dichroism analyses on a JASCO J815 spectropolarimeter, with a Peltier cell temperature controller and a thermal transition from 20 to 80°C. Purified enzymes were diluted at 2µM in Tris buffer (50mM Tris, 150 mM NaCl, 1mM DTT, 1 mM EDTA, pH7.0).

Structure determination.

Crystallization conditions.

Mutant 30H3 was crystallized using conditions previously described by Skov et al.[25]. Experiments were carried out in 24-well plates at 12°C using the hanging drop vapor diffusion method. 1 volume of protein at 10mg/mL (in Tris 50mM, NaCl 150mM, EDTA 1mM, DTT 1mM, pH 7.3) was mixed to one volume of precipitant solution (PEG 6000 at 25, 30, 35% or 40% (w/v), 0.1M HEPES at pH 6.8, 7.0, 7.3, 7.6, 7.9 or 8.2). Platelet crystals appeared after 1 week. Best crystals were obtained for PEG 6000 at 30% (w/v) and 0.1M HEPES pH 7.3.

Data collection and structure determination.

X-ray experiments were carried out at 100 K. Prior to flash cooling, crystal of mutant 30H3 were soaked for two minutes in the reservoir solution supplemented with 100 mM sucrose. Diffraction data set were collected to a maximum resolution of 2.0 Å on beamline ID14-EH1 at the European Synchrotron Radiation facility (ESRF, Grenoble, France). Diffracted intensities were integrated and scaled using XDS[28] and 5% of the scaled amplitudes were randomly selected and excluded from the refinement procedure. Crystals of mutant 30H3 belong to orthorhombic space group $P22_12_1$ with one molecule per asymmetric unit giving a Matthews coefficient of 2.3 Å^3/Da. Data collection statistics are given in Table 4.

Model Building and refinement.

Structure refinement was performed using the structure of *Np*AS wild type (PDB code: 1G5A) as an initial model and refmac5 from CCP4 GUI[29]. Model was built manually in SigmaA weighed electron density maps using COOT[30]. Water molecules were automatically assigned and ligand molecules were fitted manually into residual maps. The final model was refined to final R and Rfree of 14.8% and 21.3% respectively. It contains 628 amino acids out of 632 theoretical residues with4 missing residues at the N-terminal extremity. Refinement statistics are given in Table 4.

Molecular modelling.

All Molecular Dynamics (MD) simulations were carried out using the AMBER 9 suite of programs[31]. The molecular all-atom ff03[32,33] and carbohydrate GLYCAM06[34] force fields were used for the proteins and the sucrose, respectively. The starting models were derived from the high-resolution crystal structures of *Np*AS in complex with the sucrose (PDB code: 1JGI) and the mutant 30H3. To obtain a neutral charge of the simulated systems, a number of counter-ions were included. Each enzyme/sucrose complex together with the counter-ions was solvated with TIP3P water molecules, using the rectangular parallelepiped box with a minimum distance of 0.12 nm between the solute and the simulation box edge. Preparation of simulations consisted of initial energy minimization steps (steepest descent and conjugate gradient methods). The minimization steps were then followed by a slow heating to 298 K under constant volume over a period of 100 ps. At the final required temperature (298 K), the system was equilibrated under constant volume condition over 10 ps and then it was turned on constant pressure (1 bar) condition over 90 ps. Atomic positions of the protein backbone were first restrained using a harmonic potential during the minimization schedule. The force constant was then progressively diminished until a final unrestrained minimization step. Harmonic constraints applied on sucrose were maintained until the end of the heating phase and thereafter, they were gradually removed along the equilibration in the isothermal-isobaric ensemble. The final production phase of simulations was then carried out for a total of 50 ns at constant temperature (298 K) and pressure (1 bar) conditions. The temperature and pressure were controlled using Langevin[35] thermostat and Berendsen barostat[36] with a collision frequency 2 ps^{-1} and pressure relaxation time as 2 ps. Long-range electrostatic forces were handled by using the particle-mesh Ewald method[37]. The time step of the simulations was 2.0 fs and the SHAKE algorithm was used to constrain the lengths of all chemical bonds involving hydrogen atoms to their equilibrium values[38]. The resulting trajectories were analyzed using the Ptraj module of the AMBER 9 package. The RMSD was calculated for the protein backbone atoms using least squares fitting. Atomic positional fluctuations (Δr_i^2) of protein backbone were calculated. A mass-weighted average value was then calculated for each residue. These parameters are related to the B-factors through the following relationship:

$$B\mathrm{i} = \frac{8\pi^2}{3} \langle \Delta r_i^2 \rangle$$

The simulated B-factors were calculated using the coordinates of the 50 ns trajectories.
All graphics were prepared using Pymol[39] (Schrodinger, Portland, OR, USA).

ACKNOWLEDGMENT

The authors are grateful to N. Monties for her help in the product analysis by HPAEC-PAD and HPAEC-PAD-MS of

the ICEO high-throughput facility, devoted to the engineering and screening of new and original enzymes, of the Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (Toulouse, France). The equipments used for crystallography experiments are part of the Integrated Screening Platform of Toulouse (PICT, IBiSA). The authors thank the European Synchrotron Radiation Facility (ESRF) at Grenoble (France), in particular the beamline ID14eh1 staff. This work was supported by the French National Research Agency (ANR Project GLUCODESIGN, ANR-08-PCVI-002-02). This work was granted access to the HPC resources of the Computing Center of Region Midi-Pyrénées (CALMIP, Toulouse, France).

ABBREVIATIONS

*Np*AS, *Neisseria polysaccharea* amylosucrase; *Mf*AS, *Methylobacillus flagellatus* amylosucrase; *Dg*AS, *Deinococcus geothermalis* amylosucrase; *Dr*AS, *Deinococcus radiodurans* amylosucrase; DNA; deoxyribonucleic acid; IPTG, isopropyl α-D-thiogalactopyranoside; PCR, Polymerase Chain Reaction; LB, lysogeny broth; DNS, DiNitroSalycilic acid; HPLC, High Performance Liquid Chromatography; HPAEC-PAD, High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection; HPSEC, High Performance Size Exclusion Chromatography; GST, Gluthation – S-Transferase; EDTA, EthyleneFiamineTetraacetic Acid; CCP4, Collaborative Computational Project no 4; DTT, DiThioThreitol; PEG, PolyEthylene Glycol; RI, Refractive Index; UV, Ultra Violet; DP, Degree of Polymerization; GH, Glycoside-Hydrolase; Tm, Melting Point; HEPES, 4-(2-HydroxyEthyl)-1-PiperazineEthaneSulfonic acid; DPn, number-average DP; DPw, weight-average DP; Gn, n glucose units; OD; Optical Density; K_m; Michaelis dissociation constant; k_{cat}, catalytic constant; V_m, maximum velocity; MD, Molecular Dynamics.

REFERENCES

(1) Emond, S.; Mondeil, S.; Jaziri, K.; André, I.; Monsan, P.; Remaud-Siméon, M.; Potocki-Véronèse, G. *FEMS Microbiol. Lett.* **2008**, *285*, 25.

(2) Pizzut-Serin, S.; Potocki-Véronèse, G.; Van der Veen, B.; Albenne, C.; Monsan, P.; Remaud-Siméon, M. *FEBS Lett.* **2005**, *579*, 1405.

(3) Potocki de Montalk, G.; Remaud-Siméon, M.; Willemot, R. M.; Sarçabal, P.; Planchot, V.; Monsan, P. *FEBS Lett.* **2000**, *471*, 219.

(4) Potocki de Montalk, G.; Remaud-Siméon, M.; Willemot, R. M.; Planchot, V.; Monsan, P. *J. Bacteriol.* **1999**, *181*, 375.

(5) Guérin, F.; Barbe, S.; Pizzut-Serin, S.; Potocki-Véronèse, G.; Guieysse, D.; Guillet, V.; Monsan, P.; Mourey, L.; Remaud-Siméon, M.; André, I.; Tranier, S. *J. Biol. Chem.* **2012**, *287*, 6642.

(6) Ha, S.-J.; Seo, D.-H.; Jung, J.-H.; Cha, J.; Kim, T.-J.; Kim, Y.-W.; Park, C.-S. *Biosci. Biotechnol. Biochem.* **2009**, *73*, 1505.

(7) Seo, D. H.; Jung, J. H.; Choi, H. C.; Cho, H. K.; Kim, H. G.; Ha, S. J.; Yoo, S. H.; Cha, J.; Park, C. S. *J. Microbiol. Biotechnol.* **2012**, *22*, 1253.

(8) Jeong, J.-W.; Seo, D.-H.; Jung, J.-H.; Park, J.-H.; Baek, N.-I.; Kim, M.-J.; Park, C.-S. *Appl. Biochem. Biotechnol.* **2014**, *173*, 904.

(9) Perez-Cenci, M.; Salerno, G. L. *Plant Sci.* **2014**, *224*, 95.

(10) Skov, L. K.; Mirza, O.; Henriksen, A.; De Montalk, G. P.; Remaud-Siméon, M.; Sarcabal, P.; Willemot, R. M.; Monsan, P.; Gajhede, M. *Acta Crystallogr. Sect.* D **2000**, *56*, 203.

(11) Jensen, M. H.; Mirza, O.; Albenne, C.; Remaud-Siméon, M.; Monsan, P.; Gajhede, M.; Skov, L. K. *Biochemistry* **2004**, *1991*, 3104.

(12) Skov, L. K.; Mirza, O.; Sprogøe, D.; Van der Veen, B.; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *Biocatal. Biotransformation* **2006**, *24*, 99.

(13) Skov, L. K.; Mirza, O.; Henriksen, a; De Montalk, G. P.; Remaud-Siméon, M.; Sarçabal, P.; Willemot, R. M.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2001**, *276*, 25273.

(14) Skov, L. K.; Mirza, O.; Sprogøe, D.; Dar, I.; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *J. Biol. Chem.* **2002**, *277*, 47741.

(15) Mirza, O.; Skov, L. K.; Remaud-Siméon, M.; Potocki De Montalk, G.; Albenne, C.; Monsan, P.; Gajhede, M. *Biochemistry* **2001**, *328*, 9032.

(16) Skov, L. K.; Pizzut-Serin, S.; Remaud-Siméon, M.; Ernst, H. A.; Gajhede, M.; Mirza, O. *Acta Crystallogr. Sect.* F **2013**, *69*, 973.

(17) Albenne, C.; Skov, L. K.; Mirza, O.; Gajhede, M.; Feller, G.; D'Amico, S.; André, G.; Potocki-Véronèse, G.; Van der Veen, B.; Monsan, P.; Remaud-Siméon, M. *J. Biol. Chem.* **2004**, *279*, 726.

(18) Albenne, C.; Potocki de Montalk, G.; Monsan, P.; Skov, L.; Mirza, O.; Gajhede, M.; Remaud-Siméon, M. *Biol. Bratislava* **2002**, 119.

(19) Vergès, A.; Cambon, E.; Barbe, S.; Salamone, S.; Le Guen, Y.; Moulis, C.; Mulard, L. a.; Remaud-Siméon, M.; André, I. *ACS Catal.* **2015**, *5,1186*.

(20) Potocki-Veronese, G.; Putaux, J.; Dupeyre, D.; Remaud-Siméon, M. *Biomacrom*olec. **2005**, *6*, 1000.

(21) Cambon, E.; Barbe, S.; Pizzut-Serin, S.; Remaud-Siméon, M.; André, I. *Biotechnol. Bioeng.* **2014**, *111*, 1719.

(22) Champion, E.; Guérin, F.; Moulis, C.; Barbe, S.; Tran, T. H.; Morel, S.; Descroix, K.; Monsan, P.; Mourey, L.; Mulard, L. A.; Tranier, S.; Remaud-Siméon, M.; André, I. *J. Am. Chem. Soc.* **2012**, *134*, 18677.

(23) Albenne, C.; Skov, L. K.; Tran, V.; Gajhede, M.; Monsan, P.; Remaud-Siméon, M.; André-Leroux, G. *Proteins Struct. Funct. Bioinf* . **2007**, *126*, 118.

(24) Champion, E.; Moulis, C.; Morel, S.; Mulard, L. A.; Monsan, P. *ChemCatChem* **2010**, 969.

(25) Skov, L. K.; Mirza, O.; Sprogøe, D.; Van der Veen, B.; Remaud-Siméon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *Biocatal. Biotransformation* **2006**, *24*, 99.

(26) Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W. *Acta Crystallogr. Sect.* D **2011**, *67*, 271.

(27) Evans, P. *Acta Crystallogr. Sect.* D **2006**, *62*, 72.

(28) Kabsch, W. *J. Appl. Crystallogr.* **1993**, *26*, 795.

(29) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr. Sect.* D **1997**, *53*, 240.

(30) Emsley, P.; Cowtan, K. *Acta Crystallogr. Sect.* D **2004**, *60*, 2126.

(31) Case, D. A.; Darden, T.; Cheatham, T. E.; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornal, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. AMBER 9, **2006**, University of California, San Francisco.

(32) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.

(33) Lee, M. C.; Duan, Y. *Proteins Struct. Funct. Bioinf* . **2004**, *55*, 620.

(34) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonza, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2007**, *29*, 622.

(35) Pastor, R. W.; Brooks, B. R.; Szabo, A. *Mol. Phys.* **1988**, *65*, 1409.

(36) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(37) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

(38) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(39) DeLano, W. L..; The PyMOL Molecular Graphics System**, 2002**, DeLano Scientific, San Carlos, CA, USA

Conclusion

Along this manuscript, we have explored the potential of computer-aided design approaches to engineer carbohydrate-active enzymes gifted with a novel substrate specificity, and thus able to catalyze the transglucosylation of a non-natural acceptor, a precursor in the synthesis of carbohydrate structures mimicking microbial cell-surface oligosaccharides. The objective *in fine* of this work is to develop chemo-enzymatic routes that include enzymatic steps at different stages of the process using engineered enzymes able to act on lightly chemically protected intermediates in order to overcome the lack of stereoselectivity of chemical 1,2-*cis* glucosylations. The purpose of these chemo-enzymatic pathways is to easily access to carbohydrates entering in the composition of vaccines. The focus of our work was set on the synthesis of *Shigella flexneri* oligosaccharides, which are responsible for endemic shigellosis (a diarrheic disease). This study fits into a collaborative ANR funded project called "GlucoDesign" (2009-2012) between the Catalysis and Enzyme Molecular Engineering group of the Laboratory of Biosystems and Chemical Engineering (LISBP) from INSA Toulouse and the GlycoChemistry group led by Laurence Mulard at the Institut Pasteur in Paris. It follows up on a prior work that aimed at engineering on rational basis novel transglycosylases able to catalyze the regio-specific glucosylation of two partially protected monosaccharides, namely the α-L-rhamnopyranoside and the allyl 2-acetamido-2-deoxy-α-D-glucopyranoside. Building up on these previous successes, and aspiring at exploring further the potential offered by state-of-the-art computational design approaches, we have undertaken a greater challenge within the frame of this thesis. This time, the aim was to deeply re-engineer the active site of a sucrose-utilizing enzyme, amylosucrase from *Neisseria polysaccharea* (*Np*AS), in order to enable recognition of a partially protected disaccharide, the allyl 2-deoxy-2-*N*-trichloroacetyl-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranoside (D'A'), and its glucosylation with the desired regiospecificity without activity loss for sucrose donor substrate. Given the complexity of the task, we used for the first time, an automated computational protein design (CPD) approach to guide the redesign of the amylosucrase active site to achieve targeted reaction, that is the transglucosylation of D'A' disaccharide. This was never attempted before for amylosucrases, nor for carbohydrate-active enzymes, generally speaking.

A computer-aided engineering strategy applied to reshape active site and increase the odds to isolate wanted acceptor specificity toward D'A' disaccharide.

First, molecular docking of the disaccharide D'A' in the active site of wild-type *Np*AS was carried out in order to identify amino acid residues that could by mutation enable productive binding of D'A' without perturbing sucrose recognition. Altogether, 23 positions were selected in subsites +1 and +2 of the active site first shell. If considering all 20 possible amino acids at these positions, this would have led to a theoretical combinatorial space of about 20^{23} sequences what would have been impossible to explore, construct or screen experimentally. Using RosettaDesign software, an automated CPD tool, we explored random combinations of mutations introduced *in silico* at these 23 positions. Designed sequences were then scored based on their estimated Enzyme-Ligand interaction energy and ranked. Altogether, 1,500 sequences were retained for further consideration. Analysis was carried out on these sequences to identify amino acid frequency encountered at each of the 23 positions and co-varying residues. On the basis of these results, a library of about 2.7×10^4 sequences was designed that reflects the amino acid diversity found in the 1,500 sequences. At experimental level, degenerate oligonucleotides encoding only for amino acids present in the design at a given position were designed.

To ensure enough diversity coverage, about 63,000 variants (3 times the size of the designed library) were generated and evaluated using multi-level screenings toward donor substrate (sucrose) and non-natural acceptor substrate (D'A'). A total of 55 mutants remained active on sucrose with either equal or higher activity than parental wild-type amylosucrase. These mutants contained in average 8.8 mutated amino acid residues per protein. The hit rate was about 0.08 % for sucrose-active mutants. Regarding glucosylation of exogenous D'A', only one mutant was identified as being positive, although glucosylation yields were quite low. The hit rate was extremely low, about 0.0015% but it was still considered a real tour de force due to the impressive number of mutations introduced in the active site (7 mutations) to enable productive binding of D'A' and its α-1,4 glucosylation, without affecting sucrose recognition. Regrettably, it was not possible within the frame of this thesis to analyze in details the individual and combinatorial effect of these mutations as it would have required extensive work to deconvolute all mutations. Nonetheless, this deconvolution could undoubtedly offer new insights on the role of mutated amino acid residues. In particular, identification of beneficial mutations could bring new ideas to further optimize activity of the mutant for D'A' glucosylation and perhaps provide some hints if one wants to undertake the glucosylation of larger molecules, such as tetrasaccharides, that could be integrated at a later stage of the chemo-enzymatic process.

Novel transglucosylation activities isolated from a deeply mutated active site.

Primary screening on sucrose of the semi-rational library led to the isolation of 55 active mutants, of which 38 had kept their transglucosylating activity. These mutants were characterized further and led to the discovery of novel activities not displayed by wild-type amylosucrase. In particular, 18 mutants revealed their unique ability to produce novel trisaccharides which are not synthesized by parental wild-type *Np*AS. These novel molecules were identified to be erlose (α-D-glucopyranosyl-(1→4)-α-D-glucopyranosyl-(1→2)-β-D-fructose) and panose (α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→4)-α-D-glucose). This is the first report of novel molecules produced by amylosucrases from sole sucrose. Three representative mutants, named 47A10, 39A8 and 37G4, were selected for further characterization. In the presence of 146mM sucrose, they produced for mutants 47A10, 39A8 and 37G4, respectively, 7.4g/L, 0.44g/L, 9.5g/L of erlose and 0.4g/L, 1.9g/L, 3.9g/L of panose. Mutant 47A10 appeared specialized for erlose production, mutant 39A8 for panose and mutant 37G4 synthesized both molecules. Production of these molecules was optimized by increasing sucrose concentration (400 and 600mM). It led to a 10-fold increase of the erlose production by mutant 47A10 and a 3-fold increase of panose synthesis by mutant 37G4 after 24h reaction. These mutants are specialized in short chain oligosaccharides production and lost their ability to transfer glucosyl units onto oligosaccharides of DP higher than 3. Upon sequencing, these mutants revealed a high number of mutations (from 7 to 11) in subsites +1 and +2, enabling a dual accommodation of maltose acceptor at subsite +1 to form both α-1,4 glucosylated maltose as in maltotriose and α-1,6 glucosylated maltose as found in panose. Molecular modelling studies suggested that mutation of residues G396 and T398 into bulkier amino acids hampered binding of oligosaccharides longer than DP3 at subsites +1, +2 and +3 what might favor accumulation of lower size products. The plasticity of the acceptor binding subsites was previously demonstrated by the large promiscuity of amylosucrase to accept a wide range of structurally distinct exogenous acceptors what was further extended by molecular engineering. The work disclosed herein is another example witnessing the plasticity of amylosucrase active site that is able to bind natural acceptors, like maltose, in different catalytically productive manners, leading to a diversification of the products from sole sucrose used as substrate. Levels of production of bioactive erlose and panose molecules, from a

cheap agroresource, could also benefit from further process optimization, in particular in batch or fed-batch reactors.

A controlled production of short soluble oligosaccharides using a multi-site engineered mutant.

Another mutant, named 30H3, was identified as being 6.5-fold more active on sucrose that wild-type amylosucrase during primary screening of the library. Analysis of its reaction products did not reveal any novel molecule, not already synthesized by *Np*AS. However, it showed very interesting features as no insoluble amylose was detected and only soluble matooligosaccharides of controlled short chain size (DP going from 2 to 21) were observed, with a very narrow polydispersity. In comparison to wild-type enzyme, the DPn values indicated smaller sizes for mutant 30H3: 4.8, 7.0, 6.2 and 6.2 for reactions carried out in the presence of 146, 300, 400, 700mM sucrose versus 24, 19 and 17 for *Np*AS in the presence of 100, 300 and 600mM sucrose. Altogether, 9 mutations were found in the active site of the enzyme, located in the different loops of the $(\beta/\alpha)_8$ catalytic barrel: mutations R226K, and I228V are carried by loop 3; A289I, F290Y, E300I by loop 4; V331T by loop 5 and Q437S, N439D and C445A belong to the domain B' (the extended loop 7). The three-dimensional structure of the mutant was determined by X-ray crystallography with a resolution of 2.0Å. The overall organization was found very close to that of *Np*AS. Main differences resided in the 9 mutated side-chains which led to a slightly larger catalytic pocket. This structural information used in combination with molecular dynamics simulations suggested a key role of mutations at positions 396, 398 and 445 from domain B' onto changes of product specificity observed for the mutant. In particular, flexibility of region G396-G428, forming a short α-helix turn in domain B', was found considerably lowered in mutant 30H3 compared to that of parental *Np*AS. It was hypothesized that it could play a role on the proper catalytic positioning of residues D393 (the transition state stabilizer) and D394 (known to be involved in binding interactions with fructosyl moiety from sucrose). Dynamics of the domain B' could also interfere with binding of oligosaccharides at secondary OB2 site and affect connection between OB1 and OB2 sites assumed to be essential for polymer elongation in prior work.

Impact of dynamics of the loops from the $(\beta/\alpha)_8$ catalytic barrel onto substrate and product specificity of amylosucrases.

For a long time, enzyme flexibility has been neglected in the investigation of catalytic mechanisms. Structure-function relationship studies have mostly relied on static protein structures determined by biophysical methods combined to biochemical data. It is only for the past 10 to 15 years that the role of protein flexibility (from local flexibility to large conformational changes, including substrate accessibility and product release) has become an accepted intrinsic property used to better understand enzyme catalysis. More and more studies have revealed that networks of motions of different types are needed to help the protein to perform its biological function(s). In particular, flexibility of the active site has been shown to be essential to lower free energy barriers and accelerate enzymatic reactions. Growing evidences on the connection between flexibility and substrate turnover rate have been proposed. It has also been shown that accuracy of CPD methods could be further improved by better accounting for molecular flexibility of proteins among other things (there is also a need for more accurate energy functions, better solvent consideration and entropy effect…). Due to the high structural complexity of proteins, only limited flexibility resulting from side-chain rotations is usually considered in computational design. However, backbone flexibility and

large conformational rearrangements that might occur during substrate binding of product release have been shown to be essential for catalysis in many cases. This has encouraged many research groups, among which the Catalysis and Enzyme Molecular Engineering group, to actively work on the question in order to propose novel methodological alternatives to take into account greater flexibility in the design of proteins. Concerning amylosucrases, molecular flexibility has been suggested to be important in various ways for the activity of the enzyme. Movements of hairpin loop 7 (G433-G449) encompassed in domain B' was previously suggested to play a role in polysaccharide elongation and sucrose feeding of the enzyme from SB2 site to catalytic SB1 site without product release, although no experimental evidences were provided. It was also recently shown that introduction of mutations at two adjacent positions 289 and 290 enabled creation of a novel acceptor specificity for the allyl 2-acetamido-2-deoxy-α-D-glucopyranoside, likely due to the impact of mutation onto the flexibility of the loops forming the active site topology. Additional results obtained within this work seem to reinforce the idea that the flexibility of loops from the $(\beta/\alpha)_8$ catalytic barrel might be responsible for the novel substrate and product specificity of amylosucrase mutants. More generally, this rises some intriguing questions regarding the plasticity and the malleability of the TIM barrel fold that is one of the most encountered among proteins, presenting nonetheless very low sequence identity. To gain further insight on the role played by loop motions onto amylosucrase activity, we initiated within the frame of this thesis the construction of several mutants containing disulfide bridges introduced at specific positions identified by molecular modelling studies in order to restrain mobility of the loops 2, 3, 4, 7 and 8 forming the active site. The idea was then to put these mutants under oxidative or reducing conditions to control the formation or the rupture of the disulfide bridges and follow the effect on the enzyme activity. Unfortunately, we ran into major difficulties to express properly these mutants and due to the lack of time, we had to put this work on hold (results not given in the manuscript). Nonetheless, this could be a promising direction to pursue in combination with their study by biophysical methods in order to better understand inter-relationships between sequence, structure, dynamics and activity of amylosucrases and *in fine* better guide engineering of their properties.

Résumé en français

Design Computationnel et Ingénierie de Transglucosylases pour la Synthèse d'oligosaccharides.

Combinée à la synthèse par voie chimique, l'utilisation de catalyseurs enzymatiques permet d'étendre la diversité moléculaire accessible. Toutefois, le manque d'enzymes naturelles ayant les propriétés désirées limite fortement le développement de voies de synthèse chimio-enzymatique, notamment pour accéder à des oligosides complexes. Afin de pallier à cela, l'ingénierie protéique s'avère être un outil puissant pour créer à façon des catalyseurs enzymatiques aux propriétés attendues. De plus en plus, les stratégies d'ingénierie reposent sur une compréhension approfondie de la catalyse enzymatique ainsi que des relations structure-activité. Les outils computationnels ont, à cet égard, fortement participé aux avancées majeures dans le domaine notamment de la modélisation de structures tridimensionnelles ou encore l'analyse fine des interactions enzyme-ligand. Depuis quelques années, des progrès fulgurants ont été accomplis grâce au développement de méthodes pour le « redesign » d'enzymes existantes ou le design dit « *de novo* » d'activités enzymatiques n'existant pas dans la nature. Ces techniques computationnelles sont ainsi amenées à occuper une place grandissante dans le domaine de l'évolution d'enzymes. En particulier, elles permettent de mieux cibler les régions structurales à modifier pour conférer au catalyseur la fonction recherchée mais également, de réduire considérablement la taille des librairies à construire ainsi que les efforts de criblage. Cependant, bien que ce type d'approches ait récemment fait ses preuves pour la création de nouvelles activités, elles sont encore peu répandues et n'ont jamais été appliquées aux enzymes actives sur les sucres, telles que les glycoside-hydrolases et les glycosyltransférases.

Au cours de cette thèse, nous avons exploré le potentiel des approches computationnelles pour concevoir de nouvelles transglycosylases dotées de nouvelles spécificités de substrat et capables de catalyser de nouvelles réactions de glycosylation, non rencontrées dans la nature. Ce travail s'inscrit dans la continuité des travaux menés ces dernières années au sein de l'équipe Catalyse et Ingénierie Moléculaire Enzymatique du LISBP à l'INSA de Toulouse en collaboration avec l'équipe de Glycochimie de Laurence Mulard (Institut Pasteur à Paris) dans le cadre du projet de recherche financé par l'ANR « Glucodesign » (2009-2012). L'objectif à terme vise à développer une voie de synthèse chimio-enzymatique de sucres complexes intégrant des étapes enzymatiques à différents stades du procédé, afin de glucosyler régio- et stéréospécifiquement des accepteurs partiellement protégés par voie chimique. Cette glucosylation réalisée par voie enzymatique a pour but de pallier le manque de stéréosélectivité de la 1,2-*cis* glucosylation par voie chimique.

Nous nous sommes intéressés dans ce cadre-là, à la synthèse des oligosaccharides présents en surface de la bactérie *Shigella flexneri*, en vue de leur utilisation dans la composition de vaccins contre une maladie endémique, appelée la Shigellose (une maladie diarrhéique). Des travaux précédents mettant en oeuvre une approche semi-rationnelle avaient conduit avec succès à la création de nouvelles transglucosylases capables de glucosyler régio- et stéréo-spécifiquement deux monosaccharides partiellement protégés : l'α-methyl L-rhamnopyranose et l'allyl 2-acetamido-2-deoxy-α-D-glucopyranose.

Dans le cadre de cette thèse, nous avons cherché à repousser les limites de l'ingénierie en effectuant un remodelage complet du site actif de l'amylosaccharase de *Neisseria polysaccharea*, une enzyme utilisant le saccharose comme substrat donneur, afin de permettre la reconnaissance et la glucosylation régiospecifique d'un disaccharide partiellement protégé,

l'allyl 2-deoxy-2-*N*-trichloroacetyl-β-D-glucopyranosyl-(1→2)-α-L-rhamnopyranose, tout en conservant l'activité sur saccharose. En s'appuyant sur une approche basée sur du design computationnel, nous avons construit une librairie d'environ $2.7x10^4$ mutants, ciblant 23 positions reparties dans la première couronne d'acide aminés autour du site actif. Cette librairie a ensuite été criblée vis-à-vis du saccharose afin d'identifier les clones capables de l'utiliser comme substrat. Les clones positifs ont ensuite été testés en réaction de transglycosylation sur l'accepteur cible. Le criblage a révélé une variété d'activités de transglycosylation très originales par rapport à celles de l'enzyme sauvage. Ainsi, un mutant capable de glucosyler le disaccharide cible a été isolé ainsi que des mutants capables à partir du seul saccharose, de produire de nouvelles molécules ou de synthétiser de façon contrôlée des maltooligosaccharides solubles. Ces mutants ont été caractérisés de façon approfondie d'un point de vue biochimique et structural afin d'élucider les facteurs moléculaires responsables de la modification des activités. Ces travaux ont été décrits dans 3 chapitres résumés ci-après.

1 - Design computationnel de librairies de séquences d'une transglucosylase pour la glucosylation d'un accepteur disaccharidique non-naturel intervenant dans la synthèse d'oligosaccharides bactériens antigéniques.

Ce premier volet de ma thèse a consisté à remodeler le site actif de l'amylosaccharase de *Neisseria polysaccharea* pour lui permettre de reconnaitre un accepteur disaccharidique, l' (2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranose (D'A'), naturellement non reconnu par l'enzyme sauvage, et de le glucosyler en α-1,4 **(Figure 1)**. L'objectif étant d'intégrer cette enzyme dans une stratégie de synthèse chimio-enzymatique de motifs oligosaccharidiques représentatifs des lipopolysaccharides (LPS) présents en surface de la bactérie *Shigella flexneri*, correspondants respectivement aux sérotypes 1a, 1b et 1d.

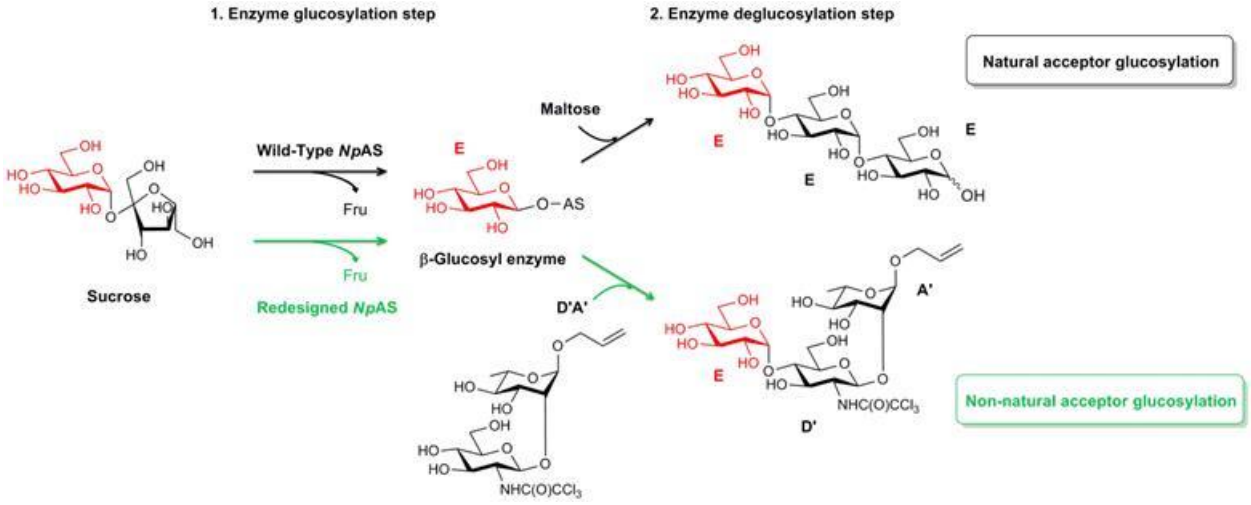


Figure 1: Réactions catalysées à partir du saccharose en présence d'un accepteur naturel, (le maltose) par l'AS*Np* (noir) et en présence de l'accepteur non-naturel (D'A') par l'enzyme redessinée (vert). *Le mécanisme catalytique est décomposé en 2 étapes: la première étape (étape de glucosylation) conduit à la formation du β-glucosyl enzyme, et la seconde (étape de dé-glucosylation) à la transglucosylation de l'accepteur (naturel ou non).* Np*AS :* Neisseria polysaccharea *amylosucrase; suc : sucrose (α-D-glucopyranosyl-(1,2)-β-D-fructofuranoside); Fru : Fructose; Maltose : (Glucopyranosyl-(1-4)-α-D-glucopyranose); A': allyl α-L-rhamnopyranoside; D': 2-deoxy-2-trichloroacetamido-α-D-glucopyranosyl; E : α-D-Glucopyranosyl.*

La stratégie employée à d'abord consisté à identifier les positions les plus favorables à muter pour permettre la reconnaissance du disaccharide D'A' au sein du site actif de l'AS*Np*. Par une étude de modélisation moléculaire, le produit glucosylé, l'allyl α-D-glucopyranosyl-(1→4)-(2-deoxy-2-trichloroacetamido-α-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranose, (ED'A') a été arrimé dans le site actif de l'enzyme afin d'identifier les acides aminés gênant le positionnement productif de la molécule et pouvant être mutés pour améliorer la reconnaissance du ED'A', sans toucher aux résidus clés pour la reconnaissance du saccharose. Au total, 23 positions (226 à 229, 290, 300, 331, 332, 396 à 398, 434 à 439, 444 à 448) des sous-sites +1, +2 et +3 ont été sélectionnées, sans affecter le sous-site -1 qui est extrêmement conservé et spécifique du saccharose. Ces acides aminés sont répartis sur les boucles 3, 4, 5 et 7 du tonneau catalytique $(\beta/\alpha)_8$. Si on considère l'ensemble de toutes les combinaisons possibles des 20 acides aminés naturels sur les 23 positions sélectionnées, l'espace théorique s'éleve à 20^{23} séquences, ce qui est exhorbitant et dépasse les capacités de criblage. Ainsi, afin de réduire la taille des librairies, nous avons utilisé un logiciel dédié au design de protéines, RosettaDesign, pour explorer *in silico* la combinatoire de séquences et sélectionner un nombre limité de séquences prédites avec une énergie favorable pour la liaison du trisaccahride ED'A'. Au total, 1,515 séquences ont été choisies et étudiées plus en détails. La fréquence des acides aminés observés à chacune des positions, ainsi que les corrélations entre positions ont été analysées. Sur cette base, une librairie semi-rationnelle de $2.7x10^4$ séquences couvrant la diversité observée dans les 1,515 séquences a été proposée et construite expérimentalement à l'aide d'amorces dégénérées et dessinées sur mesure pour encoder uniquement les acides aminés proposés par le design. Afin de s'assurer d'une exploration suffisante de la diversité générée, environ 63,000 clones ont été générés et soumis à un criblage primaire colorimétrique sur milieu solide pour identifier les mutants actifs sur saccharose. Après isolement, expression en milieu liquide et vérification par HPLC, 55 variants ont été confirmés comme étant toujours actifs sur le saccharose. Les activités sur saccharose déterminées par test au DNS indiquent un large spectre d'activités allant d'une activité équivalente à l'enzyme sauvage à presque 6.5 fois plus, alors qu'en moyenne, environ 8.8 mutations ont été introduites par gène. La stabilité des 55 variants a également été évaluée par DSF, révélant également un profil hétérogène des valeurs de Tm indiquant des variations allant de -5 à +3 °C par rapport au Tm de l'enzyme sauvage. Ces 55 mutants ont ensuite été évalués en réaction de transglycosylation sur l'accepteur cible (D'A'). Les produits de la réaction ont été analysés par LC/MS, ce qui a conduit à l'identification d'un mutant unique, à savoir le mutant (F3), capable de glucosyler D'A' selon 2 produits correspondants à une forme mono- (Produit P1) et di-glucosylée (Produit P2) **(Figure 2)**, alors que l'enzyme sauvage ne reconnait pas le D'A'. L'analyse RMN de P1 et P2 a révélé que P1 contenait la mono-glucosylation attendue en α-1,4, et que P2 était issu d'une seconde glucosylation en α-1,6 du produit P1 agissant comme accepteur de l'enzyme.

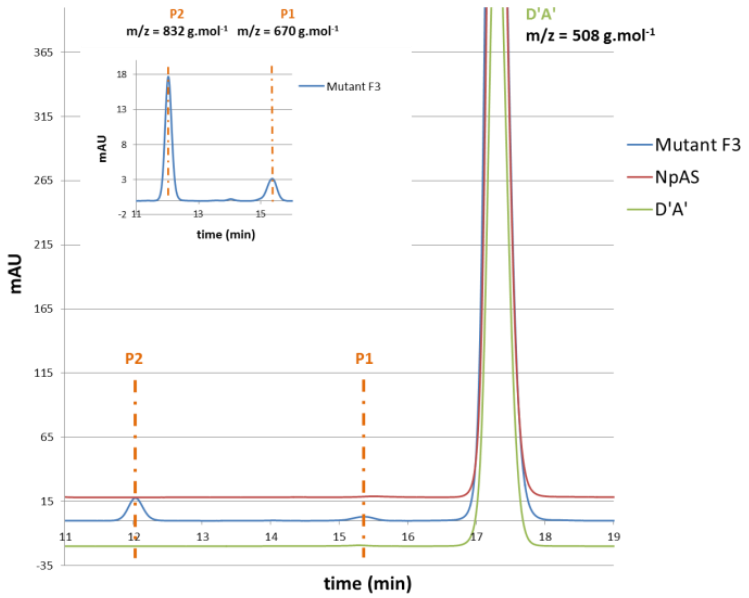


Figure 2 : Comparaison des profils de produits obtenus par l'enzyme sauvage (*Np*AS) en rouge, et le mutant F3 en bleu, après 24h en présence de 146mM de saccharose et 146mM de D'A'. *Les analyses ont été effectuées en HPLC (colonne C18 (Venusil AQ 50mm) avec pour éluant de l'acétonitrile / eau (17:83, v/v), à 30°C, 1mL/min) et une détection UV à 205 et 220nm. Le profil du D'A' (t_r = 17.3 min) est représenté en vert pour référence. P1 (t_r = 15.3 min) et P2 (t_r = 12.1 min): produits de glucosylation ayant pour masses molaires respectives : 670 et 832 g.mol^{-1}. Un zoom sur les produits P1 et P2 est situé en haut à gauche.*

Les produits obtenus par ce mutant en présence de saccharose seul et de saccharose avec ajout de D'A' ont été analysés en détails et comparés à l'enzyme sauvage, AS*Np* **(Figure 3).** En présence du seul saccharose, le mutant F3 synthétise plus de turanose (34.2% versus 25.3%), de maltose (14.9% vs. 6.1%) et de petits oligosaccharides (29.6% vs. 3.1%) que l'enzyme sauvage. L'activité de polymérisation a par contre été annulée (pas d'amylose insoluble, DP>25). En présence de D'A', seul le mutant F3, contenant 7 mutations au niveau du site actif, est capable de produire du ED'A' et du EED'A', bien que le taux de glucosylation demeure faible, seulement 2% des unités glucopyranosyles issues du saccharose sont transférées sur D'A' ou ED'A' **(Figure 3).** Néanmoins, cette activité est totalement inédite, et n'est pas observée pour l'AS*Np* sauvage.

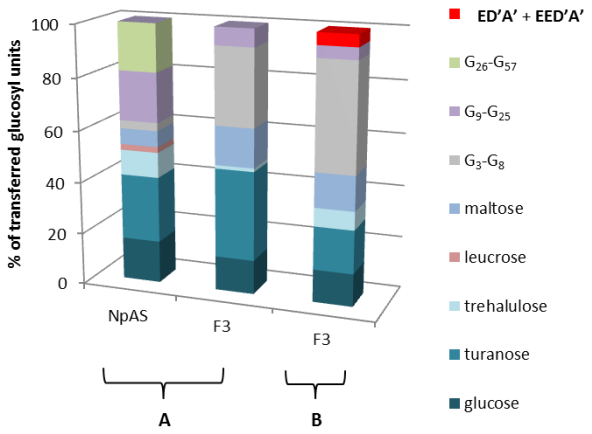


Figure 3 : Pourcentages d'unités glucopyranosyles transférées dans les différents produits réactionnels, analysés par HPAEC-PAD (*Colonne Dionex Carbo-Pack PA100; 30°C; 1 mL/min; gradient d'acétate de sodium (de 6 à 300mM en 28 min) avec 150mM NaOH) de l'enzyme sauvage (*NpAS*) et du mutant F3 après 24h. A: en présence de 146mM de saccharose et B: en présence de 146mM de saccharose et de 146mM de D'A'.*

La caractérisation cinétique sur saccharose (5-600mM) seul témoigne d'un comportement similaire à celui de l'enzyme sauvage, avec toutefois une meilleure affinité pour le donneur (saccharose) ($K_{m,\ apparent}$ diminué d'un facteur 2.3). A l'inverse, en présence de D'A' (200mM), l'efficacité catalytique est impactée suggérant une diminution de l'affinité du mutant pour le saccharose. Concernant l'affinité pour D'A'(5-200mM) en présence de saccharose (200mM), l'affinité pour D'A' est relativement faible au vu de l'efficacité catalytique (3.5×10^{-4} $s^{-1}mM^{-1}$). Néanmoins, bien que faible, cette nouvelle activité catalytique pourrait servir de point de départ pour mener de nouveaux tours d'évolution dirigée en vue d'être améliorée.

Enfin, l'étude par modélisation moléculaire de ce mutant en complexe avec le produit monoglucosylé ED'A' **(Figure 4)** a permis d'éclairer le rôle joué par les mutations introduites sur la reconnaissance de l'accepteur D'A'. En particulier, les mutations au niveau des positions 226 et 300, engagées dans un pont salin dans l'enzyme sauvage, ont conduit à une ouverture du site actif et à une augmentation de sa plasticité, favorisant l'arrimage productif du disaccharide D'A'. Des interactions spécifiques favorables avec les groupements protecteurs de D'A' ont été rendues possibles par introduction de mutations au niveau des positions 398, 396 et 331.

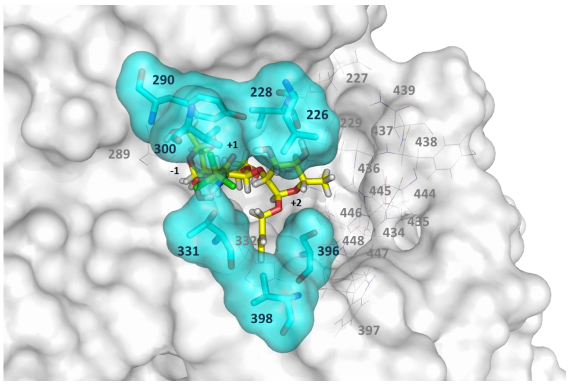


Figure 4 : Arrimage moléculaire de ED'A' dans le site actif du modèle 3D du mutant F3'. *Les 7 mutations (R226L-I228V-F290Y-E300V-V331T-G396S-T398V) sont représentées en cyan (les atomes d'hydrogènes ont été enlevés par clarté). En gris sont représentés les positions ciblées par le design mais non mutées dans F3.*

En conclusion, l'approche computationnelle mise en jeu dans le cadre de cette étude a permis d'explorer la combinatoire de séquences sur 23 positions de la première sphère du site actif et de proposer la construction d'une librairie de taille réduite en adéquation avec la cadence de nos criblages. A partir d'une librairie de $2.7x10^4$ variants, un seul mutant a été isolé avec l'activité recherchée, ce qui témoigne de la complexité de la tâche. Néanmoins, vues les importantes différences structurales entre D'A' et l'accepteur naturel, il a été nécessaire d'introduire 7 mutations dans le site actif de l'enzyme pour lui permettre de reconnaitre l'accepteur non-naturel. Le couplage du design computationnel et de l'ingénierie semi-rationnelle ont ainsi conduit à un réel tour de force puisque outre l'acquisition d'une nouvelle spécificité vis-à-vis de l'accepteur D'A', l'activité originale sur saccharose (substrat donneur) n'a pas été affectée, offrant ainsi un point de départ pour l'optimisation des performances de ce mutant.

2- Nouvelles spécificités de substrat pour la production d'erlose et de panose par des mutants multi-positions de l'amylosaccharase.

Le criblage sur saccharose de la librairie semi-rationnelle a révélé des mutants aux comportements originaux, bien que non-actifs sur D'A'. En particulier, le criblage par HPLC des 55 mutants isolés comme actifs sur saccharose a permis de mettre en évidence la production de nouveaux produits, P1 et P2, non synthétisés par l'enzyme sauvage. Ainsi, quatre groupes ont été définis afin de classer les mutants selon les profils de produits : groupe A (profils similaires à l'enzyme sauvage), groupe B (mutants produisant une nouvelle molécule, P1), groupe C (mutants produisant une nouvelle molécule, P2) et groupe D (mutants produisant les molécules, P1 et P2). Pour chaque groupe, une enzyme représentative a été sélectionnée, produite, purifiée et caractérisée en détails. Puis, les 3 mutants appelés 47A10, 37G4 et 39A8 représentant respectivement les groupes B, D et C ont été mis en présence de 146mM de saccharose pendant 24h, et les produits analysés et quantifiés par HPAEC-PAD couplée à la spectrométrie de masse **(Figure 5).** A l'aide de molécules standards, les produits P1 et P2 ont ainsi été identifiés comme étant des trisaccharides correspondant respectivement à de l'erlose (α-D-Glucopyranosyl-(1→4)-α-D-Glucopyranosyl-(1→2)-β-D-Fructose) et du panose (α-D-Glucopyranosyl-(1→6)-α-D-Glucopyranosyl-(1→4)-α-D-glucose). Il est à noter que ces mutants se sont avérés spécifiques de la synthèse d'oligosaccharides ne dépassant pas un DP3 et ayant complètement perdu leur activité de polymérisation.

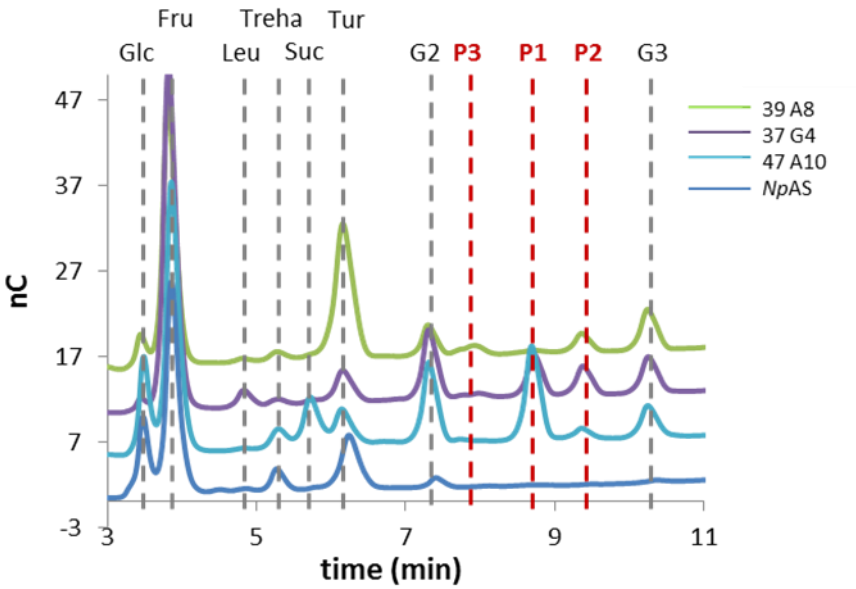


Figure 5 : Profils de produits des mutants 47A10, 37G4, 39A8 et de l'enzyme sauvage après 24h sur 146mM de saccharose et analysés par HPAEC-PAD. *Le profil du mutant 37G4 est en violet, du mutant 39A9 en vert, du mutant 47A10 en cyan et de l'enzyme sauvage en bleu. Glc, Glucose; Fru, Fructose; Treha, Trehalulose; Tur, Turanose; Suc, Sucrose; Leu, Leucrose; G_2, Maltose; P1, produit 1; P2, produit 2; P3, produit 3 ; G_3, maltotriose.*

Comme indiqué sur la **Figure 6**, ces 3 mutants incorporent plus d'unités glucopyranosyle dans le maltose, le maltotriose et ne produisent pas de maltooligosaccharides de degrés supérieurs. Concernant la production d'erlose, le mutant 47A10 incorpore 23% et le mutant 37G4 30% des unités glucopyranosyle disponibles dans la production de la molécule contre 1.7% et 17.7% dans la production de panose. Le mutant 39A8 quant à lui synthétise plus de panose (8.7%) que d'erlose (1.3%) mais une grosse majorité de turanose (46% vs 20% (AS*Np*)). Les concentrations de produits représentent : 6.4 (47A10), 8.5 (37G4) et 0.5g/L (39A8) pour l'erlose et 0.4 (47A10), 3.4 (37G4) et 2.1g/L (39A8) pour le panose.

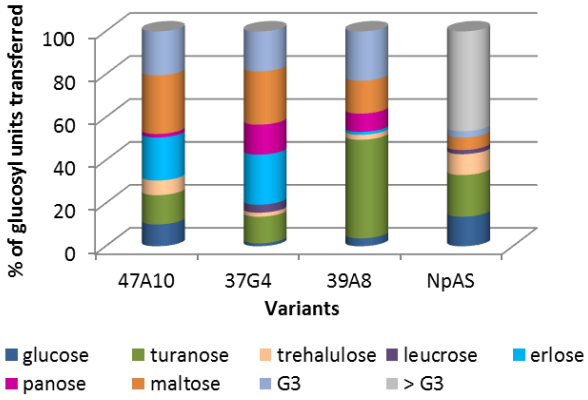


Figure 6 : Pourcentages d'unités glucopyranosyle transférées dans les différents produits réactionnels à partir de saccharose des mutants 47A10, 39A8, 37G4 et du sauvage (*Np*AS) par analyse HPAEC-PAD. *Les réactions ont été conduites sur 146mM de saccharose et 1U/mL d'enzyme. G: Glucose.*

Afin d'augmenter la production de ces 2 molécules, erlose et panose, des concentrations croissantes en saccharose ont été testées : 400 et 600mM. Comme représenté sur la **Figure 7**,

la production d'erlose a ainsi pû être fortement augmentée. A 600mM, 70.6 g/L, 57.5 g/L and 9.6g/L (42%, 46% and 5.2%) d'erlose a été produit. En revanche, l'impact de la concentration de saccharose sur la quantité produite de panose a été moindre.

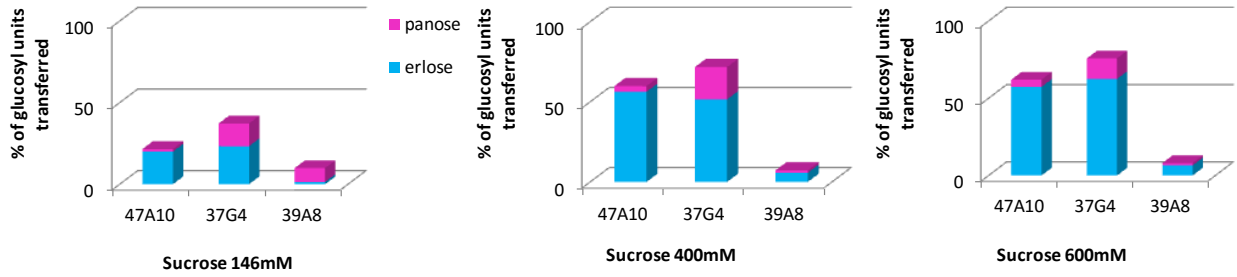


Figure 7 : Incorporation des unités glucopyranosyles dans l'erlose (bleu) et le panose (rose) à partir de saccharose (146, 400 et 600mM) par les mutants 47A10, 37G4 and 39A8.

Le mutant 37G4 a pour principal avantage d'être capable de synthétiser les 2 molécules dans des proportions non négligeables.

Le séquençage des 3 mutants a révélé un nombre important de mutations, 7 (47A10), 10 (37G4) et 11 (39A8) introduites dans leur site actif **(Table 1).**

Table 1: Séquences des mutant sélectionnés. *En orange sont surlignées les positions mutées par rapport à l'enzyme sauvage.*

			Mutants		
Amino acid positions		**wild-type**	**47A10**	**37G4**	**39A8**
226	Loop 3 (domain B)	R	L	K	K
227		E	E	E	E
228		I	V	V	V
229		F	A	F	F
289	Loop 4 (domain A)	A	I	I	I
290		F	Y	Y	Y
300		E	I	I	I
331	Loop 5 (domain A)	V	T	T	T
332		H	H	H	H
396	Loop 7 (Domain B')	G	G	S	S
397		W	W	W	W
398		T	T	V	V
434		V	V	V	V
435		P	P	P	P
436		F	F	F	F
437		Q	Q	R	S
438		Y	Y	Y	Y
439		N	N	D	D
444		D	D	D	D
445		C	C	C	R
446		R	R	R	R
447		V	V	V	V
448		S	S	S	S
Number of mutations			**7**	**10**	**11**

Les efficacités catalytiques sur saccharose des mutants 47A10 et 37G4 ont été diminuées d'un facteur 1.7 et 2.1 suggérant une perte d'affinité pour le saccharose alors que pour le mutant 39A8, une augmentation d'un facteur 2.2 a été calculée.

Afin de mieux comprendre l'impact des mutations sur la production des nouveaux produits, des modèles 3D des mutants ont été construits par modélisation moléculaire en complexe avec plusieurs ligands (turanose, erlose, panose, maltotriose). Ainsi, il a été proposé que la perte de l'activité de polymérisation pouvait être assimilée au remodelage des sous sites +1 et +2 provoquant la perte du pont salin entre les résidus 226 et 300 et des sites d'ancrage fort au niveau des sous-sites +3 (positions 396 et 398), mutés ici par des acides aminés encombrants, et pouvant rendre difficile l'ancrage de longues chaines dans le site actif. Il est intéressant de noter ici que le saccharose joue le rôle à la fois de donneur et d'accepteur tout comme le maltose. La glucosylation en α-1,6 du maltose pour la synthèse du panose rend également originale ces enzymes puisqu'il ne s'agit pas de la spécificité de liaison originelle de l'enzyme sauvage. Les molécules d'erlose et de panose sont d'autant plus intéressantes qu'elles peuvent être utilisées dans l'industrie agroalimentaire en tant qu'agent sucrant, non cariogéniques et peu caloriques. Le panose est relativement bien produit par les dextransaccharases, d'autres transglycosylases utilisant le saccharose comme substrat donneur. En revanche, même si l'intérêt pour l'erlose est grandissant, aucune enzyme n'est encore décrite dans la littérature pour le produire avec de bons rendements, notamment en présence du seul saccharose. Ces mutants pourraient donc présenter un intérêt après optimisation, pour proposer un nouveau procédé de synthèse d'erlose, en réacteur dans des conditions optimisées.

3- Isolement et caractérisation d'un mutant de l'amylosaccharase de *Neisseria polysaccharea* pour la synthèse de maltooligosaccharides de tailles de chaines contrôlées.

Lors du criblage de la banque semi-rationnelle sur saccharose, un autre mutant s'est démarqué de par son activité accrue sur saccharose (facteur 6.5 par rapport à l'enzyme sauvage, avec enzyme non purifiée). Ce mutant nommé 30H3 a révélé la présence impressionnante de 9 mutations au sein de son site actif : R226K, I228V (boucle 3) ; A289I, F290Y, E300I (boucle 4) ; V331T (boucle 5) ; Q437S, N439D and C445A (boucle 7). Après production et purification du mutant 30H3, un suivi cinétique sur 24h en présence de saccharose (146mM) à quantité égale d'enzyme (0.3mg/mL) a été réalisé, ce qui a permis de confirmer le gain d'activité accrue. Comme indiqué sur la **Figure 8**, la totalité du saccharose a été consommée en 2h contre 7h pour l'enzyme sauvage. Une amélioration d'un facteur 5.6 a été calculée par régression linéaire entre 0 et 40 minutes.

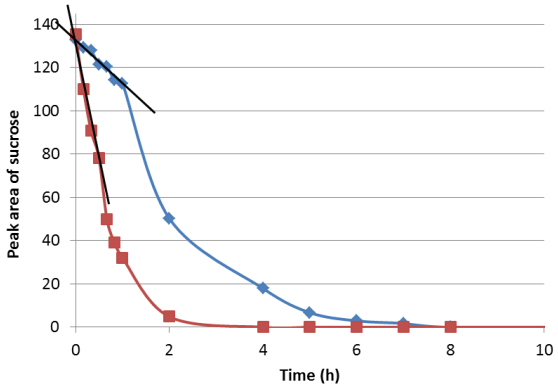


Figure 8 : Suivi cinétique de la consommation de sacharose en fonction du temps par le mutant 30H3 (rouge) et l'enzyme sauvage (bleu) par detection HPLC *(0.3mg/mL d'enzyme purifiée et 146mM de saccharose).*

Son profil de produits sur 146mM de saccharose a été analysé par HPAEC-PAD. Une absence de maltooligosaccharides (MOS) de DP > 21 **(Figure 9 A)** a été observée contrairement à l'enzyme sauvage qui présente un profil de produits contenant des maltooligosaccharides (MOS) de DP allant jusqu'à 90. Par conséquent, ce mutant est capable de synthétiser uniquement des MOS solubles. Quelles que soient les concentrations en saccharose utilisées (300, 400 et 700mM), jamais la synthèse de produits insolubles n'a été observée. En effet l'augmentation de la concentration en saccharose a pour conséquence de favoriser la synthèse de nouveaux oligosaccharides identifiés par les flèches bleues sur la **Figure 9 B** qui ne sont des MOS de DP connus **(Figure 9 B).** Ils pourraient s'agir de formes polyglucosylées du saccharose ou de certains de ses isomères. Cette hypothèse reste à vérifier.

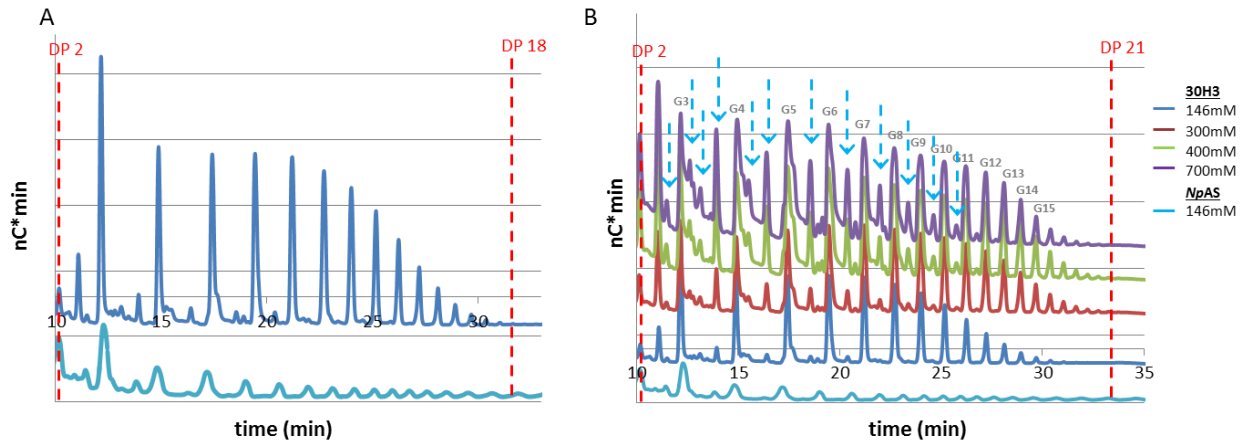


Figure 9 : Comparaison des extraits bruts analysés par HPAEC-PAD ; *(A) mutant 30H3 (bleu) et* Np*AS (cyan) dilué par 5 en présence de 146mM de saccharose à 30°C ; (B) mutant 30H3 en présence de 146mM (bleu), 300mM (rouge), 400mM (vert) et 700mM de saccharose (violet) dilué par 5, 20.52, 13.7 and 16, respectivement. L'enzyme sauvage (*Np*AS) est représentée en cyan et diluée 5 fois. Les flèches bleues représentent les nouveaux oligosaccharides produits en comparaison de l'enzyme sauvage. Gn, maltooligosaccharides de DPn.*

Les DP moyens en nombre (DPn) (de 4.8 à 7.0) et en masse (DPw) (de 6.8 à 8.9) ainsi que la polydispersité des produits formés à partir de différentes concentrations de saccharose ont été calculés et comparés à ceux obtenus avec l'enzyme sauvage. Les résultats révèlent que le mutant 30H3 produit des MOS de tailles bien inférieures à ceux synthétisés par l'enzyme native (DPn = 24, 19, 17 et DPw = 73, 50, 39 à 100, 300 et 600mM de saccharose), et de polydispersité également bien plus faible (1.3-1.4 pour le mutant 30H3 vs 2.4 pour l'AS*Np*). Ce comportement est très éloigné de celui de l'enzyme sauvage qui synthétise préférentiellement de longs MOS dont certains sont insolubles et qui présentent une polydispersité élevée. Le pourcentage d'unités glucosyle incorporées dans les différents produits de la réaction ont aussi été déterminés par HPAEC-PAD. La production de turanose est favorisée pour une concentration standard (146mM). Par contre la concentration en MOS de DP 8 à 21 est favorisée à 300mM. La caractérisation cinétique n'a pas permis de révéler d'amélioration significative des paramètres cinétiques sur saccharose. En vitesse initiale, le mutant n'apparait pas plus efficace que l'enzyme sauvage. Cette observation nous a conduits à formuler l'hypothèse selon laquelle le mutant pourrait être activé par les produits relargués au cours de la réaction, ce qui permettrait de rendre compte de la cinétique de consommation de saccharose présentée en **Figure 8**. Ces produits n'étant pas présents en quantité suffisante

en début de réaction, leur effet sur la vitesse initiale n'est pas détectable. Un phénomène proche a déjà été décrit par Cambon *et al*. pour d'autres mutants de l'amylosaccharase dont le mutant R226N. Pour approfondir la caractérisation de ce mutant et la compréhension des liens structure-activité-dynamique, nous avons entrepris de résoudre la structure tri-dimensionnelle de ce mutant. La protéine a été cristallisée dans les conditions suivantes : PEG 6000, 30% et HEPES 0.1M pH7.3. Une structure cristallographique 3D a été obtenue à 2.0Å, elle est hautement similaire à celle de l'enzyme sauvage **(Figure 10 A)** et présente différents ligands : 1 molécule de Tris dans le site actif (déjà observé auparavant et décrit comme pouvant aider à la stabilisation des boucles formant la topologie du site actif) ; 3 molécules de saccharoses dans les sites SB2 et SB3 (déjà caractérisés par Albenne *et al.*) et sur un quatrième site que nous nommerons SB4 et un maltotriose dans le site OB2 **(Figure 10A).**

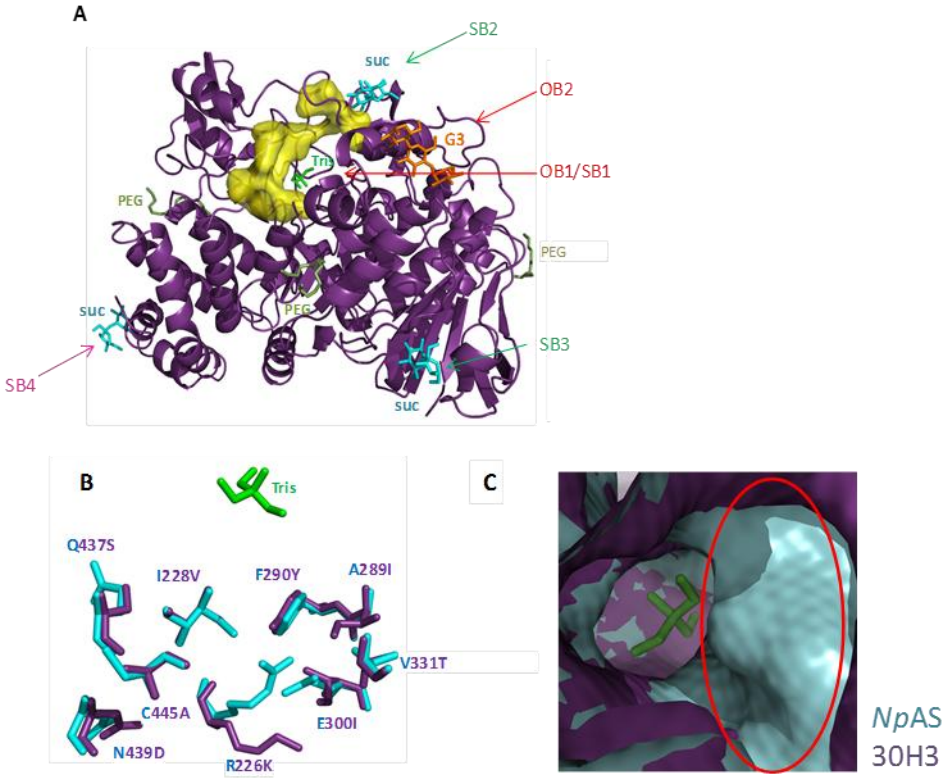


Figure 10 : Représentation de la structure 3D du mutant 30H3 (violet) superposée à celle de l'enzyme sauvage (cyan, code PDB : 1G5A). ***A :*** *organisation du mutant 30H3 avec une molécule de tris (vert) au sein du site actif, deux PEG (vert fonçé), 3 saccharoses (cyan), un maltotriose (orange) et les 9 mutations (halo jaune).* ***B:*** *superposition des 9 acides aminés mutés 226, 228, 289, 290, 300, 331, 437, 439 et 445 en violet sur les acides aminés sauvages en bleu. Une molécule de tris est dockée au sein du site actif.* ***C :*** *représentation en surface du mutant 30H3 en violet et de l'enzyme sauvage en cyan, avec une molécule de saccharose dockée dans le site actif (1G5A).*

Comme illustré en **Figures 10 B et C**, l'ouverture de l'accès à la poche catalytique semble due aux mutations en positions 226 et 300 qui ont conduit à une rupture du pont salin présent dans l'enzyme sauvage. Ce mutant étant capable de synthétiser plus de turanose que l'enzyme sauvage, tout comme le mutant 39A8 décrit précédemment, il est apparu judicieux de comparer leurs mutations. Seules les positions 396, 398 et 445 diffèrent, et parmi elles, seule la position 445 n'est pas de type sauvage pour le mutant 30H3. Ceci nous a amenés à penser

que la mutation en position 445 pouvait jouer un rôle majeur dans la synthèse écourtée de MOS. Cet acide aminé pourrait être responsable d'un changement de flexibilité de la boucle 7 portant la mutation et du domaine B' et par conséquent affecter la liaison des oligosaccharides au niveau du site secondaire OB2. La boucle 7 a été décrite précédemment comme jouant potentiellement un rôle de connexion entre les sous sites OB1 et OB2. Ainsi, une modification en position 445 de la boucle pourrait par conséquent jouer sur l'arrimage des chaines et leur élongation.

Des simulations de dynamique moléculaire ont été menées sur le mutant 30H3 et sur l'enzyme sauvage en complexe avec le saccharose. Les différences les plus significatives ont été observées dans la région G396-G428 du domaine B'. L'introduction des mutations dans le mutant 30H3 a conduit à une diminution significative de la flexibilité de cette région qui porte également deux résidus essentiels pour la catalyse, Asp393 qui stabilise l'état de transition et Asp394 qui participe au positionnement productif du résidu fructosyle dans le saccharose. Ainsi la flexibilité du domaine B' pourrait jouer un rôle déterminant dans la nature des produits formés par le mutant 30H3, en particulier la production de MOS de taille contrôlée pouvant ouvrir de nouvelles voies pour des applications intéressantes.

Conclusions

A partir d'une stratégie semi-rationnelle appliquée à l'amylosaccharase de *N. polysaccharea* pour concevoir de nouveaux catalyseurs pour la glucosylation d'un disaccharide non-naturel, de structure relativement éloignée de celle des substrats naturels, nous avons isolé une variété d'activités originales inattendues.

Tout d'abord un mutant a été isolé nous permettant d'atteindre l'objectif fixé initialement qui était la glucosylation de l'allyl (2-deoxy-2-trichloroacetamido-β-D-glucopyranosyl)-(1→2)-α-L-rhamnopyranose (D'A'). Pour cela, 23 positions du site actif avaient dû être ciblées pour remodeler la topologie de la poche par design computationnel. Suite à une réflexion poussée sur le design de la librairie, une librarie semi-rationnelle de taille réduite a été construite et criblée, révélant le mutant F3 qui a requis l'action combinée de 7 mutations pour permettre la glucosylation de D'A'. Ce mutant, unique par sa spécificité, représente une excellent point de départ pour des travaux ultérieurs d'évolution visant, en particulier, à augmenter ses performances catalytiques.

En explorant la diversité issue de cette librairie, trois mutants, appelés 39A8, 37G4 et 47A10 contenant également entre 7 et 11 mutations dans le site actif, ont conduit à la production à partir du seul saccharose, de deux trisaccharides non synthétisés par l'enzyme sauvage : l'erlose et le panose. Au vu des propriétés edulcorantes et acariogène de ces molécules, les mutants obtenus pourraient s'avérer utiles pour la production de nouveaux sirops de sucres à forte teneur en erlose et panose à partir de saccharose.

Pour terminer, le mutant 30H3, isolé à partir de cette même librairie, a conduit à une activité fortement améliorée vis-à-vis du saccharose en dépit des 9 mutations introduites dans le site actif. Ce mutant a également conduit à une activité originale en comparaison de l'enzyme sauvage pour la synthèse de MOS solubles de longueur de chaines contrôlée.

Les résultats obtenus au cours de cette étude ont permis de démontrer le potentiel des approches de design computationnel couplé à l'évolution d'enzymes pour créer des activités enzymatiques nouvelles. Cette stratégie appliquée pour la première fois à des transglycosylases a permis d'élargir le répertoire de réactions de glucosylation catalysées par

les enzymes et la création de nouvelles spécificités vis-à-vis de substrats non-naturels. Le nombre élévé de mutations introduites dans le site actif sans perdre l'activité pour le substrat originel soulève des questions intriguantes notamment sur la plasticité et la maléabilité du repliement protéique $(\beta/\alpha)_8$, un des plus répandus et retrouvé pour des séquences présentant très peu de similarité. En particulier, le rôle de la flexibilité des boucles du tonneau catalytique sur la reconnaissance des substrats et la synthèse des produits, et qui semblerait être modulée par l'introduction de mutations, demeure un point clé à élucider pour mieux comprendre les liens entre la séquence, la structure, la dynamique et l'activité des amylosaccharases, et des enzymes plus généralement, et pouvoir à terme améliorer le pouvoir prédictif des méthodes de design computationnel pour proposer rapidement les mutations à apporter pour concevoir des catalyseurs d'intérêt notamment pour les domaines de la glycochimie et de la glycobiologie synthétique.