Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de Toulouse

Présentée et soutenue par

AO LI

Le 25 juin 2020

Découverte de nouvelles glycoside-phosphorylases pour la synthèse d'oligosaccharides

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

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :
TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering

Thèse dirigée par
Gabrielle VERONESE et Elisabeth LAVILLE

Jury

Mme Estelle Bonnin, Rapporteure
M. Jean-Guy Berrin, Rapporteur
Mme Caroline Remond, Rapporteure
M. Vincent Lombard, Examinateur
Mme Magali Remaud-Simeon, Examinatrice
Mme Gabrielle Potocki-Veronese, Directrice de thèse
Mme Elisabeth Laville, Co-directrice de thèse

Thèse

Présentée devant

L'Institut National des Sciences Appliquées de Toulouse

En vue de l'obtention du

Doctorat

Spécialité : Ingénieries Microbienne et Enzymatique

Par **Ao Li**

Discovery of new glycoside-phosphorylases for oligosaccharide synthesis

Directrices de thèse

Mmes Gabrielle POTOCKI-VERONESE et Elisabeth LAVILLE

NOM: LI **PRÉNOM**: Ao

TITRE: Découverte de nouvelles glycoside-phosphorylases pour la synthèse d'oligosaccharides.

SPÉCIALITÉ: Ingénieries Microbienne et Enzymatique

ANNÉE: 2020 **LIEU**: INSA, TOULOUSE

DIRECTRICES DE THÈSE: Mmes Gabrielle POTOCKI-VERONESE et Elisabeth LAVILLE

Résumé:

Les glycoside-phosphorylases sont des enzymes actives sur les hydrates de carbone (CAZymes), capables de catalyser à la fois la dégradation des glycosides en utilisant le phosphate inorganique pour rompre les liaisons osidiques, et leur synthèse en utilisant les sucre-phosphates comme donneurs de glycosyles. Très peu de glycoside-phosphorylases ont été caractérisées à ce jour par rapport aux autres CAZymes, malgré leur implication dans d'importants processus biologiques, en particulier dans le système digestif des mammifères, et leur potentiel pour la synthèse de glycosides à haute valeur ajoutée. Dans les génomes et les métagénomes, leurs séquences sont en effet difficiles à distinguer de celles des glycosides-hydrolases et des transférases, avec lesquelles elles partagent de nombreuses similitudes structurales et mécanistiques. De nombreuses glycoside-phosphorylases sont donc probablement encore cachées dans la fraction non cultivée des écosystèmes microbiens.

Dans ce travail de thèse, nous avons développé une nouvelle approche pour accélérer la découverte de glycoside-phosphorylases, et pour analyser leur diversité dans les microbiomes. Cette approche générique combine l'exploration de larges espaces de séquences (méta)génomiques, et le criblage d'activités glycoside-phosphorylases. Pour établir la preuve du concept, nous avons d'abord ciblé la famille CAZy GH130, qui contient à la fois des mannoside-phosphorylases et des mannosidases ciblant des mannosides de structures et d'origines diverses. Nous avons analysé 6,308 séquences GH130, dont 4,714 provenant des microbiomes humains, bovins, porcins et murins. En utilisant des réseaux de similarité de séquences, nous avons divisé la diversité des séquences en 15 groupes principalement isofonctionnels ; parmi ceux-ci, neuf ne contenaient aucun membre caractérisé expérimentalement. En analysant les alignements de séquences pour chaque groupe, nous avons pu prédire les déterminants du mécanisme phosphorolytique et de la spécificité de liaison osidique. Ces prédictions ont été testées en caractérisant quatre membres de cette famille, dont les séquences sont parmi les séquences GH130 les plus répandues et les plus abondantes dans le métagénome intestinal humain. Leurs fonctions précises ont été identifiées grâce à une combinaison d'analyses chromogéniques, chromatographiques, de résonance magnétique nucléaire, de spectrométrie de masse et mobilité ionique. Nous avons découvert la première β-1,4-mannosyl-acide glucuronique-phosphorylase connue, et une mannoside-phosphorylase/transmannosylase très originale. Toutes deux ciblent des motifs glycosidiques trouvés dans des levures et bactéries pathogènes. Cette approche a ensuite été appliquée à l'analyse des familles GH65, GH94, GH112, GH149 et GH161. Nous avons montré que la diversité de séquences et de fonctions des familles GH65, GH94 et GH112 est déjà bien couverte par les données génomiques et biochimiques actuellement disponibles. En revanche, les séquences GH149 et GH161, en particulier issues des métagénomes, sont probablement une source de nouveauté fonctionnelle. Au total, onze cibles ont été sélectionnées à partir de groupes de séquences non caractérisés, représentant, potentiellement, onze nouvelles fonctions, ou du moins des fonctions qui ne sont pas décrites pour ces familles. Ce travail de thèse a ainsi permis de développer une stratégie efficace pour la découverte de nouvelles glycoside-phosphorylases et l'évaluation de leur diversité dans les microbiomes. Il a également révélé des interactions possibles entre les bactéries intestinales, et permis d'identifier de nouveaux outils biotechnologiques pour la synthèse d'oligosaccharides antigéniques.

Mots-clés: Glycoside phosphorylases, Sequence similarity network, Microbiome intestinal, Glucides, Synthèse d'oligosaccharides

---

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

Unité de recherche: TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering

SURNAME: LI **FIRST NAME**: Ao

TITLE: Discovery of new glycoside-phosphorylases for oligosaccharide synthesis.

SPECIALITY: Enzymatic and Microbial Engineering

YEAR: 2020 **PLACE**: INSA, TOULOUSE

THEIS DIRECTOR: Mrs Gabrielle POTOCKI-VERONESE and Elisabeth LAVILLE

Summary:

Glycoside-phosphorylases are particular carbohydrate active enzymes (CAZymes), able to catalyze both glycoside degradation by using inorganic phosphate to breakdown osidic linkages, and synthesis by using sugar-phosphates as glycosyl donors. Very few glycoside-phosphorylases have been characterized to date compared to the other CAZymes, despite their involvement in important biological processes, in particular in the gut of mammals, and their potential for the synthesis of high-added value glycosides. In genomes and metagenomes, their sequences are indeed difficult to discriminate from those of glycoside-hydrolases and transferases, with which they share many structural and mechanistic similarities. Many glycoside-phosphorylases are thus probably still hidden in the uncultured fraction of microbial ecosystems.

In this thesis work, we developed a novel approach to boost the discovery of glycoside-phosphorylases, and to analyze their diversity in microbiomes. This generic approach combines sequence-based mining of (meta)genomes and activity-based screening of GP activity. To establish the proof of concept, we first targeted the CAZy family GH130, which contains both mannoside-phosphorylases and mannosidases targeting mannosides of various structures and origins. We analyzed 6,308 GH130 sequences, including 4,714 from the human, bovine, porcine and murine microbiomes. Using sequence similarity networks, we divided the diversity of sequences into 15 mostly isofunctional meta-nodes; of these, nine contained no experimentally characterized member. By examining the multiple sequence alignments in each meta-node, we predicted the determinants of the phosphorolytic mechanism and linkage specificity. These predictions were tested by characterizing four members of this family, of which the sequences are among the most prevalent and abundant GH130 sequences of the human gut microbiome. Their functions were proven by using chromogenic assays, high-performance anion exchange chromatography, nuclear magnetic resonance and cyclic ion mobility mass spectrometry. We discovered the first known β-1,4-mannosyl-glucuronicacid phosphorylase, and a very original dual mannoside-phosphorylase/transmannosylase. Both of them target glycosidic motifs found in pathogenic yeasts and bacteria. This approach was then applied to the analysis of families GH65, GH94, GH112, GH149 and GH161. We showed that the sequence and functional diversity of the GH65, GH94 and GH112 families is already well covered by the presently available genomic and biochemical data. In contrast, the GH149 and GH161 sequences, in particular the metagenomic ones, are probably a source of functional novelty. In total, eleven targets were selected from uncharacterized meta-nodes, representing, potentially, eleven novel functions, or at least functions which are not described for these families. In this thesis work, we thus developed an efficient strategy for the discovery of glycoside-phosphorylases and assessment of their diversity in microbiomes. It also revealed possible interactions between gut bacteria, and allowed us to identify novel biotechnological tools for the synthesis of antigenic oligosaccharides.

Keywords: Glycoside phosphorylases, Sequence similarity network, Gut microbiome, Glycosides, Oligosaccharide synthesis

---

DOCTORAL SCHOOL: SEVAB - Ecological, Veterinary, Agronomic Sciences and Bioingeneering

RESEARCH INSTITUTE: TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering

Acknowledgements

This thesis would not have been possible without the patient instruction, insightful criticism and expert guidance that I received from my talented supervisor, Research Director Gabrielle Potocki-Veronese, who lead me to deal with the huge sequence datasets and discovery novel fantastic and naughty enzymes. I will never forget the first day Gaby took me into her office and gave me a wonderful lab-tour. Her guidance and advice will not only help me in all the time of research and writing of the thesis, but also will inspire my entire research career. At the same time, I would like to express my sincere gratitude to my co-supervisor Elisabeth Laville, for giving the encouragement and sharing insightful suggestions particularly in genomic and metagenomic analyzes. Thanks you for being there, always accompany me when I have tons of questions.

Apart from my great supervisors, I won't forget to express the gratitude to rest members in DiscOmics team, for their help in the experiment and discussion, especially, I am grateful for Laurence Tarquis, who is an excellent technician, for her devoted work in the screening assay and purification of these naughty enzymes, for giving me valuable suggestion about enzyme characterization, and My sincere thanks should also go to Alexandra Tauzin, Emna Bouhajja, Julien Durand and Mounir Benkoulouche, all of whom provide lots of useful help in my research. Utmost gratitude and heartfelt appreciation for all the members in Molecular Modeling and Design Group, especially Isabelle André and Jeremy Esque, who gave me permission to the highly functional computers. It is really means a lot that Jeremy provide me with various of magical script and technical support. I am also greatly indebted all the members in Catalysis and Enzyme Molecular Engineering group for providing a kind and pleasant working atmosphere.

I would also like to thank all the collaborators: Vincent Lombard, Nicolas Terrapon and Bernard Henrissat from the CAZy team in Marseille, annotated the GPs-containng family enzymes in the targeted metagenomes; Diego Morgavi from INRAE Saint-Genès-Champanelle, provided the bovine rumen metagenomic dataset; David Guieysse from TBI, performed the NMR analysis; Simon Charnock and Darren Cook from Prozomix in Northumberland, provided me with some enzymes. Simon Ollivier and David Ropartz from the BIBS platform at INRAE of Nantes, performe the mass spectrometry, which are of great help for the discovery the β-1,4 mannosyl-glucuronicacid and Man-β1,2-Man-α1,2-Man.

I would like to express my deepest gratitude to the members of my jury: Estelle Bonnin, Jean-Guy Berrin, Bernard Cathala, Magali Remaud-Siméon and Caroline Rémond-Zilliox for accepting to judge this work.

Thanks to the financial support from China Scholarship Council (CSC) in the frame of CSC-UT/INSA, provide me an excellent opportunity to study in France and develop my capability of scientific thinking.

Lastly, my special thanks would go to my beloved family and girlfriend, who give me consistent support and encouragement during this challenging period.

Table of contents

Abbreviations list

αGlcNAc1P	α-D-N-acetyl-glucosamine-1-phosphate
αGal1P	α-D-galactose-1-phosphate
αGlc1P	α-D-glucose-1-phosphate
αMan1P	α-D-mannose-1-phosphate
βGlc1P	β-D-glucose-1-phosphate
βGlcNAc1P	β-D-N-acetyl-glucosamine-1-phosphate
AAs	Auxiliary activities
Ara	L-arabinose
ATD	Arrival time distributions
CAT	Comparative annotation toolkit
CAZymes	Carbohydrate-active enzymes
CBMs	Carbohydrate-binding modules
CEs	Carbohydrate esterases
d4-TSP	Sodium 2,2,3,3-tetradeuterio-3-trimethylsilylpropanoate
DNPGlc	2,4-dinitrophenyl β-D-glucoside
DP	Degree of polymerization
EFI-EST	Enzyme function initiative-enzyme similarity tool
EIM	Extracted ion mobiligrams
ESI	Electrospray ion source
Fru	D-fructose
Gal	D-galactose
GalA	D-galacturonic acid
GalNAc	N-acetyl-D-galactosamine
GHs	Glycoside hydrolases
Glc	D-glucose
GlcA	D-glucuronic acid
GlcN	D-glucosamine
GlcNAc	N-acetyl-D-glucosamine
GLNBP	Galacto-N-biose/lacto-N-biose I phosphorylase
GNBP	Galacto-N-biose phosphorylase

GPs	Glycoside phosphorylases
GSs	Glycosynthases
GTs	GlycosylTransferases
HMM	Hidden Markov models
HMOs	Human milk oligosaccharides
HPAEC-PAD	High performance anion exchange chromatography with pulsed amperometric detection
HPLC	High performance liquid chromatography
LPMOs	Lytic polysaccharide monooxygenases
IBD	Inflammatory bowel diseases
IMS-MS	Ion mobility-mass spectrometry
Man	D-mannose
ManNAc	N-acetyl-D-mannosamine
NGS	Next generation sequencing
OPME	One-pot multi-enzymatic
PDB	Protein data bank
PLs	Polysaccharide lyases
PUL	Polysaccharide utilization loci
RG-I	Rhamnogalacturonan I
Rha	L-rhamnose
RI	Refractive index detection
SSNs	Sequence similarity networks
Sus	Starch utilization system
TGs	Transglycosidases
TLC	Thin layer chromatography
Uhgb_MP	Unknown human gut bacterium_mannoside-phosphorylase
Xyl	D-xylose

General introduction

Nature spent millions of years evolving powerful biological systems. Among them are found enzymes, able to catalyze and speed up all the reactions necessary for a living organism to thrive. While our knowledge on enzymes greatly expanded over the last decades, there is still a lot to discover. The latest technological advances, in particular in the 'omics' and bioinformatics fields to explore huge data sets, is a gateway toward exploitation of the untapped natural diversity of enzymes and the understanding of their catalytic mechanisms. Not only this is absolutely necessary to improve our knowledge on the very complex interplays that occur in cells and in complex ecosystems, but these new enzymes are also very interesting tools for a large number of applications in several fields, whether it is for pharmaceutics, diagnostics, food and feed, cosmetics, energy…

Of special interest are carbohydrate active enzymes (or CAZymes), able to synthesize, modify and degrade glycosidic linkages. Glycosides are involved in numerous cell processes, for example energy storage, cell protection or recognition. Their structures are exquisitely diversified and can be very complex. The understanding of their synthesis and degradation is important, both at the fundamental and applied levels. Many glycosides indeed have a great potential for preventing diseases (*e.g.* antigenic glycosides), improving our gut health and comfort (*e.g.* prebiotics), or shifting towards more sustainable practices (*e.g.* biofuels and biosourced polysaccharide-based materials). This often requires devising suitable synthetic routes to access such glycosides, involving more and more the usage of CAZymes, which are not necessarily known or adapted for biotechnological processes. While laboratory evolution is now a common strategy to tweak, fine-tune, or even design *de novo* biocatalysts, there is still a lot that Nature has to offer.

This is especially true when considering one class of CAZyme called glycoside-phosphorylases (GPs), my model of study. Very few GPs have been characterized to date compared to the other CAZymes, despite their involvement in important biological processes, in particular the breakdown of glycosides in the gut of mammals, and their potential as synthetic tools for glycoside synthesis. In genomes and metagenomes (the collective genomes of a microbiota), their sequences are indeed difficult to discriminate from those of glycoside-hydrolases and transferases, with which they share many structural and mechanistic similarities. A lot of GPs are thus probably still hidden in the uncultured fraction of microbial ecosystems.

In order to better understand the context of my work, my thesis manuscript will therefore start with a bibliographic introduction, that will give insights into the known diversity of GPs, their mechanisms and biotechnological applications. I will then present the tools currently available to explore large (meta)genomic datasets with activity- and sequence-based approaches, and their respective limitations. In the next three chapters, I will present the results of this thesis work, in particular the development of a novel approach to analyze the diversity of GP sequences in (meta)genomes, and the discovery of novel GPs from the human gut microbiome. This manuscript will end with a general conclusion and perspectives to this work.

Scientific production

Research papers

- Li A., Laville E., Tarquis L., Lombard V., Ropartz D., Terrapon N., Henrissat B., Guieysse D., Esque J., Durand J., Morgavi D. P., & Potocki-Veronese G. (2020). Analysis of the diversity of the glycoside hydrolase family 130 in mammal gut microbiomes reveals a novel mannoside-phosphorylase function. Microbial Genomics.

- Ollivier S*, Li A*, Laville E, Tarquis L, Guieysse D, Durand J, et al. A dual glycoside-phosphorylase/transmannosylase enzyme from an uncultured gut bacteria, associated with inflammatory bowel diseases. In preparation.* Contributed equally to this work.

Review paper

-Li A, Benkoulouche M, Potocki-Veronese G. Functional diversity of glycoside phosphorylases in glycoside hydrolase families. In preparation.

Oral and poster communications

Li A, Laville E, Pompon D, Lombard V, Henrissat B, and Potocki-Veronese G. Diversity and abundance of glycoside-phosphorylases in gut microbiomes, 3-6 October 2017, Siena, Italy, (short talk and poster presentation).

Li A, Tarquis L, Laville E, Pompon D, Lombard V, Henrissat B, and Potocki-Veronese G. Discovery of new glycoside-phosphorylases from mammal gut microbiomes. Functional metagenomics conference, Trondheim, Norway, June 16th–19th, 2019(poster)

Introduction

1. Carbohydrates: from monosaccharides to polysaccharides

Carbohydrates are ones of the most abundant organic compounds and are widely distributed in Nature. They are essential components of all living organisms, in combination with other macromolecules such as lipids, proteins and nucleic acids. Based on their degree of polymerization (DP), carbohydrates can be divided into monosaccharides (DP = 1), disaccharides (DP = 2), oligosaccharides (DP = 3 to 10) and polysaccharides (DP > 10) (Cummings and Stephen 2007). Monosaccharides are the most basic carbohydrates, made of one single carbohydrate monomer, also named glycosyl unit. Glycosides are composed of at least one glycosyl unit, linked by an O-glycosidic bond to another glycosyl unit or to a non-glycosyl, hydroxylated group. Glycosides include oligosaccharides, which are made up of a few covalently linked glycosyl units. They also include polysaccharides, which are classified as homopolysaccharides, when they consist of one glycosyl type, or as heteropolysaccharides when they have more than one type of glycosyl moieties. The molecular masses of polysaccharides are very diverse, yielding, for some of them, to millions of Daltons. In Nature, polysaccharides play various roles (Stick and Williams 2010). First, polysaccharides can act as energy storage molecules. For instance, starch is the most common energy storage polysaccharide in plants, while it is glycogen in animals, both being composed of only glucosyl moieties (Roach 2002; Lal 2018). Secondly, polysaccharides can play important roles in cell wall structuration. Cellulose is the most abundant organic polymer found on Earth and the main constituent of plant cell walls (Delmer and Amor 1995). Fungal cell walls are made up of chitin, which also serve as the fundamental material in arthropod exoskeletons (*e.g.* crustaceans, insects…) (Muzzarelli 2013). Finally, oligosaccharides and polysaccharides can be covalently linked to lipid molecules to form glycolipids, that are the common components of biological membranes. They can also be linked to proteins to constitute glycoproteins, which are essential transmembrane and extracellular components of plant, animal, and bacterial cells, being involved in cellular recognition. Both glycoproteins and glycolipids are known as glycoconjugates, which also include a battery of smaller molecules (Allen and Kisailus 1992). The term ‘glycan’ includes all oligosaccharides and polysaccharides, whether or not they are linked to non-carbohydrate moieties.

Glycosides are used in a broad range of applications in food & feed (Cui 2005; Embuscado 2014), pharmaceutical (Seeberger and Rademacher 2014; Tiwari 2020), cosmetic (Se-Kwon Kim 2014), chemical and biofuel industries (Raimo 2018; Azad and Rasul 2019). They are also crucial for the functioning of ecosystems, since they fuel cells and mediate their interactions. Therefore, it is of particular importance to investigate and try to understand, in deep details, their synthesis, modification and degradation.

1.1. Structural diversity

The structural diversity of carbohydrates is great, being several orders of magnitude higher than that of any other biological macromolecule (Roger A. Laine 2008). This is due to the diversity of constitutive glycosyl moieties (>100), each of them having various absolute configuration and anomeries, and also to the existence of multiple hydroxyl groups that can be further elongated or branched by other glycosyl moieties, resulting in a huge diversity of more or less complex structures. Branched glycosides are a unique feature that distinguishes them from other linear polymers such as proteins and nucleic acids, and contribute to the wide diversity of carbohydrate structures. It has been estimated by Laine that the possible linear and branched isomers of one single hexasaccharide could lead to more than 10^{12} possible structures (R. A. Laine 1994), although it is known that carbohydrates would adopt preferred conformations in solution, involving possible cyclization (Rao 2019). However, the real number of

oligosaccharide and polysaccharide structures in Nature is still unknown. It has recently been suggested by Lapebie and colleagues that the diversity of glycan structures could be much smaller than the theoretical one, given that 'only' a few thousand enzyme combinations have been elaborated by Bacteroidetes (bacteria found in many different environments) to breakdown glycans (Lapébie et al. 2019).

2. Carbohydrate active enzymes

The enzymes which synthesize, degrade and modify glycosides are called Carbohydrate-Active enZymes (CAZymes). They are listed in the CAZy database (http://www.cazy.org) (Lombard et al. 2014).

2.1. Classification and mechanisms

The classification of CAZymes was established in 1991 by B. Henrissat, based on structural and mechanistic similarities of these enzymes. It is now a standard in the glycoscience field (Lombard et al. 2014). The CAZy database covers sequences from archaea, bacteria and eukaryotes, in large majority from cultured species of which the genome has been sequenced (André et al. 2014) .

The CAZy database is daily updated from public GenBank sequences (ftp://ftp.ncbi.nih.gov/genbank/daily-nc) (Benson et al. 2012), Swiss-Prot sequences (https://www.uniprot.org/) (Boutet et al. 2016) and the Protein Data Bank (PDB; http://www.rcsb.org) (Berman et al. 2000).

CAZymes contain at least one catalytic domain, which can be associated to carbohydrate binding modules, and/or to additional catalytic domains. Catalytic domains and carbohydrate binding modules are described in the CAZy classification as 'modules', and constitute the base of the CAZy classification. The CAZy database provides an updated classification of CAZyme modules by family since 1998. A family is created when at least one member is biochemically characterized. Sequences that share low sequence similarities with already known CAZy sequences, and which are not sufficiently numerous to create a novel family, are listed as 'non-classified modules'. As of March 2020, the CAZy database counted 436 families, including 167 glycoside hydrolases (GHs), 110 glycosyltransferases (GTs), 40 polysaccharide lyases (PLs), 17 carbohydrate esterases (CEs), 86 carbohydrate-binding modules (CBMs) and 16 auxiliary activities (AAs), of which the specific traits are detailed hereafter.

- Glycoside Hydrolases (GHs)

GHs were the first described CAZymes in the classification by Henrissat and colleagues (Henrissat and Davies 1997). They are widely distributed in living organisms, and very often used for biorefineries (Rakotoarivonina et al. 2016; Bourlieu et al. 2020). Most GHs can only hydrolyse the glycosidic bonds between two or more glycosyl moieties, or between a glycosyl and a non-glycosyl moiety. They usually act following one of the two mechanisms involving the formation of oxocarbenium ion-like transition states as described by Koshland (Koshland 1953), but several variations exist for specific GH families, involving, for example, substrate-assisted catalysis (Terwisscha van Scheltinga et al. 1995). The two main mechanisms are known as single-displacement mechanism (occurring for inverting GHs, which invert the anomeric configuration of the donor) and double-displacement mechanism (occurring for retaining GH which retain the anomeric configuration of the donor) (Figure 1). A third mechanism was recently described, which uses NADH as a cofactor (Rajan et al. 2004). Nevertheless, some retaining

GHs (*e.g.* from the GH13, GH70 and GH77 families) naturally display transglycosidase activity (Plou, Segura, and Ballesteros 2007) thus being able to both hydrolyze and synthesize glycosidic bonds. On the other hand, some GHs were specifically engineered to increase their glycoside synthesis ability (Figure 2). These enzymes are called glycosynthases and will also be briefly presented in Section 2.4.

- GlycosylTransferases (GTs)

GlycosylTransferases (GTs) (Coutinho et al. 2003) catalyze the transfer of sugar moieties from an activated donor to an acceptor molecule, forming regio- and stereo-specific glycosidic bonds. The activated donor is a sugar mono- or diphosphonucleotide for Leloir-type GTs, or a non-nucleotide donor, which may be a polyprenol pyrophosphate, a polyprenol phosphate, a sugar-1-phosphate, or a sugar-1-pyrophosphate for non-Leloir type GTs. GTs, just like GHs, can invert or retain the anomeric configuration of the donor. Inverting GTs act following a single nucleophilic substitution step, usually with the assistance of a divalent cation. The mechanism of retaining GTs is still unclear and remains to be solved definitely. They could act following either a double displacement close to the retaining mechanism described by Koshland (Koshland 1953), or by "front-side", internal return mechanism (S_Ni-type) (Persson et al. 2001; S. S. Lee et al. 2011). GTs are usually responsible for the assembly of glycosides *in vivo* (Lairson et al. 2008).

- Polysaccharide Lyases (PLs)

Polysaccharide Lyases (PLs), like GHs, can cleave glycosidic linkages, but they act by β-elimination to produce a terminal unsaturated hexenuronic acid moiety and a new reducing end. They are specific to uronic acid-containing polysaccharides (Lombard et al. 2010). PLs only account for a very small proportion in the vast majority of genomes. It merely corresponds to 3~5% of the GHs amount, which could be explained by a smaller proportion of uronic acids-containing polysaccharides in Nature.

- Carbohydrate Esterases (CEs)

Carbohydrate esterases (CEs) can remove *O*- or *N*-acyl substituents on esterified glycans (Lombard et al. 2014) and hence promote the subsequent carbohydrate degradation by GHs and/or PLs. There are two classes of substrates for carbohydrate esterases, the carbohydrate playing the role of the acid (*e.g.* pectin methyl esters) or of the alcohol (*e.g.* acetylated xylan) (Biely 2012) .

- Auxiliary Activities (AAs)

Auxiliary Activities (AAs) is the latest created group in the CAZy classification, with the recent identification of lytic polysaccharide monooxygenases (LPMOs) by an oxidative mechanism (Vaaje-Kolstad et al. 2010; Levasseur et al. 2013). It is reported that auxiliary activities (AAs) are related to two main families of enzymes: lignolytic enzymes, assisting the degradation activity of other CAZymes by oxidizing the surrounding lignin, or LPMOs, which display an oxidative degradation mechanism (Bissaro et al. 2017).

- Carbohydrate-Binding Modules (CBMs)

Carbohydrate-Binding Modules (CBMs) are non-catalytic modules, defined as contiguous amino acid sequences within CAZymes with a discreet fold and having carbohydrate-binding activity. They play a key role in mediating protein-carbohydrate associations by specific binding, especially for some insoluble polysaccharides (Boraston et al. 2004). In rare cases, some CBMs can be independent or found in cellulosomal scaffolding proteins.

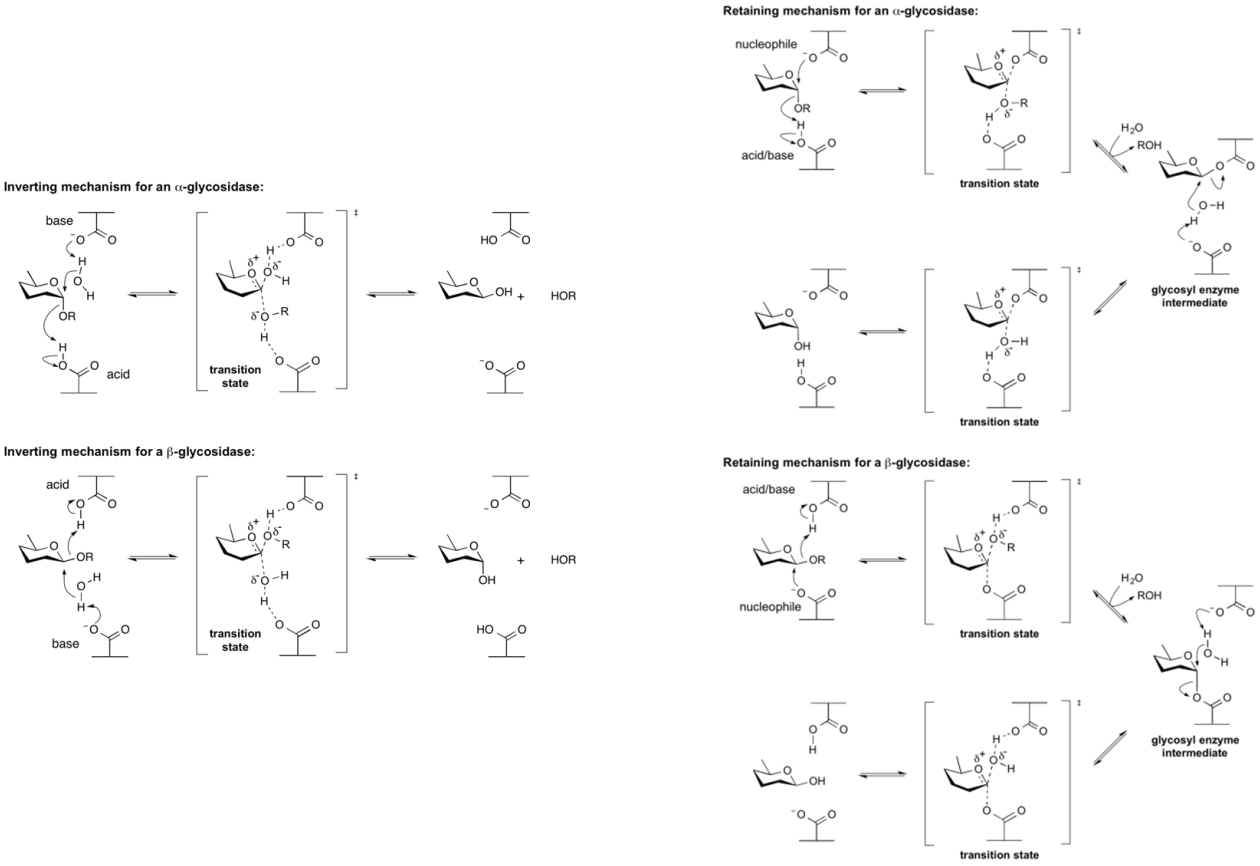


Figure 1: The two main mechanisms of inverting and retaining glycosidases (from https://www.cazypedia.org)

As previously explained, the CAZy classification is based on sequence similarities, the corresponding enzymes thus sharing similar folds and catalytic machineries. However, the sequences within a same family can display large differences. Structural identity is classified into the following hierarchical levels, which can be used to enhance the prediction accuracy in the description of the catalytic mechanisms (Lombard et al. 2014):

The **clan** level gathers families that share some common secondary structure elements leading to a similar fold. They also hold their catalytic machinery on the same secondary structure elements.

The **family** level includes sequences of enzymes sharing an identical catalytic mechanism, with a conserved catalytic machinery, and which display a common fold.

In addition, to deal with the problem of poly-specific families, **subfamilies** have been created in some of them. They are subgroups that share a more recent common ancestor and are more uniform in their molecular function, with a highly conserved catalytic site. The creation of subfamilies is frequently based on phylogenetic analyses but has not been carried out for all the CAZy families, especially when there is a lack of biochemical characterization. It focuses on rather multispecific families including for example GH5 , GH13 (Stam et al. 2006), GH30 (John, González, and Pozharski 2010), GH43 (Mewis et al. 2016) and a lot of PL families (Lombard et al. 2010). More subfamilies classification is currently in progress internally in the CAZy team and will be released in the CAZy database when an increasing

number of sequences will confirm the subfamilies stability. A novel method for the creation of subfamilies, based on sequence similarity networks (presented in Section 4.3) has recently been tested for the GH16 family (Viborg et al. 2019).

In CAZy, sequences are automatically or manually added in the database after comparison with its internal BLAST and HMM libraries of modules, all including a validation step by a curator. The high-quality manual curation of the data makes CAZy annotation the standard in the glycoscience field. To offer auto-annotation capabilities for the increasing number of genome and metagenome sequencing projects, the dbCAN(2) web server and the Comparative Annotation Toolkit (CAT) have been developed. dbCAN2 is based on the identification of hidden Markov models (HMMs) of the already known CAZyme families (Han Zhang et al. 2018). Contrary to daily updated databases such as the CAZy database, this webserver is only updated once a year (Han Zhang et al. 2018), which prevents the annotation of the most recent CAZy families and could generate errors. CAT (B. H. Park et al. 2010) utilizes reference-free, duplication-aware multiple genome alignment to symmetrically annotate multiple genomes at the same time, which provides an effective way to annotate entire genome sequences and identify evolution relationships (Fiddes et al. 2018).

2.2. Biological functions of glycoside-degrading enzymes

CAZymes are widely distributed in all living organisms (around 1–3% of the total gene content), making it particularly important to understand their biological functions and their roles in the metabolism.

The organisms that feed on complex glycosides have evolved a huge amount of CAZymes for their breakdown into metabolisable monosaccharides. It is the case of most of the microorganisms on Earth, in particular those of the human gut microbiota. The CAZymes encoded by the human genome only account for a small group of all the CAZymes contained in the human body, which is in fact a superorganism formed by the host and the microorganisms inhabiting it. Human CAZymes can only degrade sucrose, lactose and the readily digestible starch (Kaoutari et al. 2013). The breakdown of more complex glycosides is attributed to the CAZymes produced by the numerous bacterial cells in the body, which harvest them to colonize and survive in the gut. Dietary glycosides are the main carbon sources for gut bacteria. By fermenting them, they produce short-chain fatty acids (butyrate, acetate and propionate), which are used by the host cells as energy sources and are involved in multiple physiological processes, reducing the risk of cardiovascular and inflammatory bowel diseases, cancer and type 2 diabete (Den Besten et al. 2013). Bacterial CAZymes are also involved in the breakdown of host glycans, such as the N- (Briliūtė et al. 2019) and O-glycans (Bell et al. 2019) which line the intestinal epithelium to protect it against bacterial invasion, and in microbial glycosides (Cuskin, Lowe, et al. 2015), playing key roles in microbial interactions in the gut.

2.3. Polysaccharide Utilization Loci

Bacteroidetes are Gram negative bacteria found in all ecosystems investigated, in particular in mammalian guts where they dominate, with Firmicutes (Ley et al. 2006). Bacteroidetes have evolved a complex mechanism to degrade polysaccharides, involving binding of the polysaccharides at the bacterial surface, followed by their degradation in oligosaccharides. The latter are transported into the periplasmic space to be further degraded in monosaccharides, which enter the central metabolism. In Bacteroidetes, the genes encoding carbohydrate transporters and the various CAZymes required to breakdown glycans cluster in specific genomic loci known as polysaccharide utilization loci (PULs), in which their expression is co-regulated (Foley, Cockburn, and Koropatkin 2016; Terrapon et al. 2018).

The starch utilization system (Sus system, Figure 2) of *Bacteroides thetaiotaomicron,* a prominent member of the human gut microbiota, is the most well-studied PUL-encoded glycan-uptake system (Tancula et al. 1992; Karunatilaka et al. 2014). The Sus system is formed by multiple-associated proteins involved in sensing, binding, transporting and then assisting the several CAZymes that utilize starch. It contains eight proteins (SusR, A, B, C, D, E, F, and G). Some of these Sus-proteins have redundant or complementary functions, in order to retain activity when one is lacking or knocked out, which could have a detrimental effect on starch degradation (Koropatkin et al. 2008; Cameron et al. 2014). SusD, E and F are outer membrane-binding proteins that bind starch at the cell surface. Starch is first roughly degraded to smaller malto-oligosaccharides by the amylase SusG, then the SusC TonB-dependent membrane transporter transports these malto-oligosaccharides into the periplasm. The final degradation into glucose is performed by the SusA and SusB glucosidases. The regulatory function of this interplay is encoded by SusR, which controls the expression of the other Sus proteins, of which the encoding genes are up-regulated in the presence of starch. Among all Sus-proteins, only some were shown as essential for the degradation of starch (Cameron et al. 2014; Foley, Cockburn, and Koropatkin 2016).

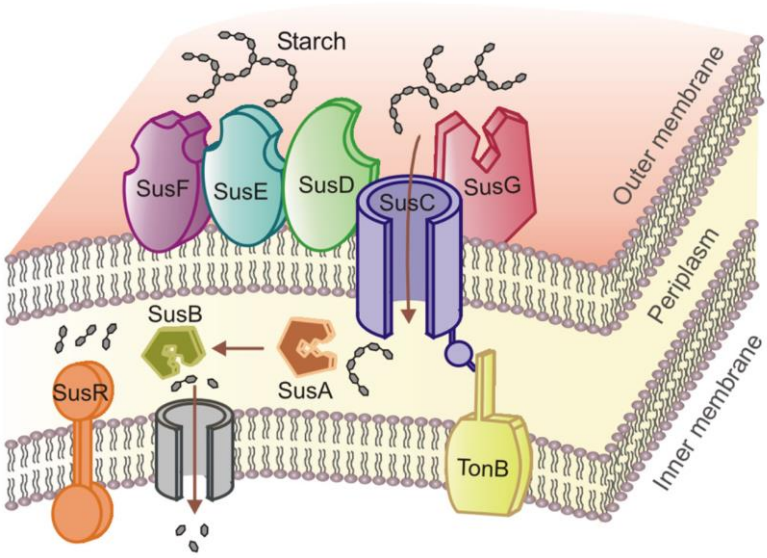


Figure 2: Model for starch catabolism by the B. thetaiotaomicron *Sus system (From (Karunatilaka et al. 2014)). The Sus system is a good model system to understand the utilization of glycans by* Bacteroidetes*, and the functioning of dozens Sus-like systems targeting various glycans have been partly or fully deciphered in the last decade (Larsbrink et al. 2014; Cockburn and Koropatkin 2016; Tingirikari 2018; Ndeh and Gilbert 2018; Déjean et al. 2020).*

Some *Bacteroidetes* species have around 20% of their genome that could be responsible for the degradation of several particular glycan structures. The intricacy of PULs mostly depends on their cognate substrates (Hamaker and Tuncil 2014; Martens et al. 2014; Grondin et al. 2017), and might include sulfatases for sulfated glycans (Mann et al. 2013), phosphatases for phosphorylated glycan such as mannans (Abbott et al. 2015), and peptidases for the release of the glycan part in glycoproteins (Renzi et al. 2015; Terrapon et al. 2018; Lapébie et al. 2019). With the development of (meta)genomic and transcriptomic approaches, an increasing number of PULs is being identified from the mammalian, insect, soil, marine and fresh water microbiota. This sometimes arose with surprises such as the lack of *susC-susD* pair in the genome of the gliding soil bacterium *Cytophaga hutchinsonii* (Xie et al. 2007), showing the complexity of these systems which remain to be fully understood. The experimentally

characterized (at least by transcriptomics) Bacteroidetes PULs are listed in the PULDB(Terrapon et al. 2018) , together with 35,632 predicted PULs identified in the genomes of 1,154 Bacteroidetes species (March 2020 release). The PULDB allows searching PULs by taxon, by (combination of) CAZy module(s), by locus tag, protein accession number, or by substrate. Basically, the principle of PUL prediction is based on, (i) the detection of every susCD-like gene pair, and (ii) on the extension of the PUL boundaries to operonic genes. The PULDB only lists Bacteroidetes PULs, because the genetic signatures used for PUL prediction (such as intergenic distances) are those, experimentally validated, of Bacteroidetes. Nevertheless, the PUL name is also extended to Gram-positive bacteria, known as Gram-positive PULs (gpPULs) (O. Sheridan et al. 2016). They are operons involved in glycan catabolism, but do not contain the canonical elements of the Sus-like systems.

2.4. CAZymes for glycoside synthesis

As numerous glycosides are used in food, feed, cosmetics, health and chemical industries, glycoside-synthesizing enzymes are very interesting tools to produce these compounds *in vitro* (Figure 2) (André et al. 2014; Benkoulouche et al. 2019). Indeed, while today most glycosides are synthesized using chemical processes, biocatalysts can bypass or be used in combination of glycochemistry (W. Li, McArthur, and Chen 2019; Rudroff et al. 2018), which often face multiple limitations, such as numerous protection and deprotection steps leading to poor yields and experimental hurdles, the use of solvents and harsh conditions etc. Since enzymes are highly regio- and stereo- specific, and do not require organic solvents, they are especially sought-after especially when considering the tremendous efforts put into the development of greener processes, and the variety of complex glycans with interesting properties targeted that still lack synthesis routes.

The formation of glycosidic bonds *in vivo* is mainly performed by GTs, which catalyze the transfer of sugar moieties from activated donor to acceptor molecules (often a carbohydrate) (Lairson et al. 2008).The first nucleotide sugar-dependent GT was identified by Luis F. Leloir (Nobel Prize in chemistry in 1970), and these GTs are often named as Leloir enzymes. The other type of GTs that utilizes non-nucleotide donors are non-Leloir GTs (Lairson et al. 2008). GTs are used in the synthesis of various oligosaccharides and glycoconjugates. The advantages of GTs are the high synthetic yield they often offer(Wong 1996) , while major drawbacks are i) the requirement of expensive donors (Overkleeft and Seeberger 2015), which can nevertheless be produced *in cellulo* together with the target GTs to directly produce and excrete oligosaccharides (Fierfort and Samain 2008; Samain and Priem 2001), and ii) the low number of available (stable, produced as soluble forms) GTs for *in vitro* synthesis purposes (McArthur and Chen 2016).

Transglycosidases (TGs) catalyze the transfer of a glycosyl unit from a donor glycoside (often readily available and cheap, unlike nucleotide sugars) onto an acceptor molecule to produce another type of carbohydrate or glycoconjugates, thanks to their wide acceptor promiscuity [69]. The mechanism of transglycosidases resemble that of retaining glycosidases. More and more attention is given to the discovery of new transglycosidases (Matsuzawa 2019) or their engineering (Teze et al. 2014; Daudé et al. 2019; Zeuner et al. 2019; Tran et al. 2019), as they are often more interesting than GTs for oligosaccharide synthesis *in vitro*. Few native transglycosidases are described and their specificities are restricted to some substrates. They are classified into the CAZy families GH2 (Wallenfels 1951; Juers, Matthews, and Huber 2012), GH13 (Skov et al. 2001; Leemhuis, Kelly, and Dijkhuizen 2010) GH16 (Eklöf and Brumer 2010; Blanco et al. 2015; Qin et al. 2017), GH23 (Höltje et al. 1975; Blackburn and Clarke 2001; Scheurwater, Reid, and Clarke 2008), GH31 (Mukai et al. 2004; Larsbrink et al. 2012), GH70 (Monchois, Willemot, and Monsan 1999; Moulis, André, and Remaud-Simeon 2016; Meng et al. 2016), GH77 (T. Takaha et al. 1993; Terada et al. 1999; Kaper et al. 2005), and GH102, GH103 and

GH104(Ursinus and Höltje 1994; Blackburn and Clarke 2001; Scheurwater, Reid, and Clarke 2008). The main drawback of transglycosidases is that often, accumulating products will suffer from secondary hydrolysis, leading to low yields. Monsan, Remaud and colleagues extensively described the utilization of native and engineered transglucosidases from the GH70 family, which are able to use sucrose as a cheap donor substrate for the production of high value-added gluco-oligosaccharides and various α-glucans for the food, health and material industries(Daudé et al. 2014; Grimaud et al. 2018; Malbert et al. 2018; Salamone et al. 2015).

Glycosynthases (GSs) are engineered glycoside hydrolases, in which the substitution of the catalytic nucleophile by a non-nucleophilic residue (Alanine, Glycine or Serine) abolishes the innate hydrolytic activity. The role of the nucleophile is played by an activated donor such as fluorine-containing substrates, providing the energy necessary to cross the barrier for the reaction to occur in the synthesis direction. They can therefore synthesize glycosidic bonds instead of hydrolyzing them. The first described glycosynthase was obtained in the group of Withers from a GH1 from *Agrobacterium* sp. (Mackenzie et al. 1998). Glycosynthases can be obtained from both retaining (Mackenzie et al. 1998; Malet and Planas 1998) and inverting (Honda and Kitaoka 2006) GHs, thus remarkably broadening the donor variety and providing new synthesis tools (Hayes and Pietruszka 2017). For example, the use of glycosynthases was explored for applications in the pharmaceutical field with the chemoenzymatic synthesis of potential therapeutic and diagnostic glycans (Cobucci-Ponzano and Moracci 2012). Another glycosynthase was more recently applied in a scalable chemo-enzymatic process for the synthesis of the trisaccharide lacto-N-triose II, a core oligosaccharide of HMO, with high yields and purity (Schmölzer et al. 2019). However, the fluorine donor substrates are expensive or have to be prepared for each study, and they are unstable (Albert et al. 2000; Okuyama et al. 2002) although they can be replaced by sugar oxazoline (Higuchi et al. 2017) or (more toxic) glycosyl azide (Cobucci-Ponzano et al. 2009) donors. Finally, the synthesis yields remain in general quite low (Hayes and Pietruszka 2017).

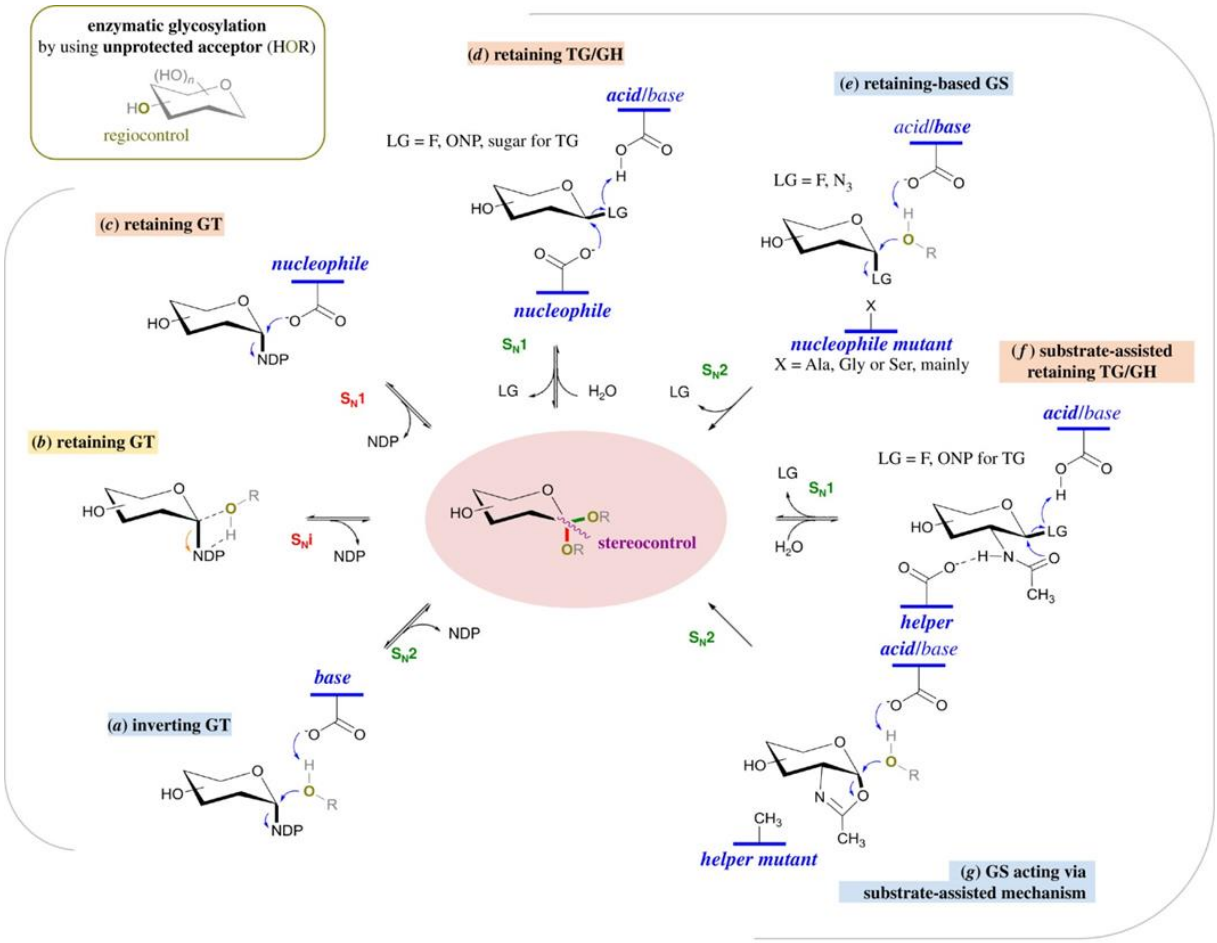


Figure 3: Enzymatic pathways for glycosidic bond formation by glycosyltranferases (GTs), transglycosidases (TGs) and glycosynthases (GSs) (Adapted from (Benkoulouche et al. 2019)). Glycoside-phosphorylases (GPs) will be presented in the next section.

Glycoside phosphorylases (GPs) catalyze both glycoside degradation by using phosphate to breakdown osidic linkages (phosphorolysis), and synthesis by using sugar-phosphates as glycosyl donors (reverse phosphorolysis). Initially the term "phosphorylases" was used to describe the enzyme that can produce α-D-glucose-1-phosphate (αGlc1P) from the glycogen present in the liver and in skeletal muscle fibers. The glycogen phosphorylases were discovered by the 1947 physiology and medicine Nobel Prize recipients Carl F. and Gerty T. Cori (C. F. Cori and Cori 1936), who greatly contributed to the understanding of the glycogen metabolism. Due to the multiple and ambiguous meaning of phosphorylases, it is preferable to name them according to the glycoside substrate which is naturally phosphorolyzed. Since GPs are the CAZymes targeted in this thesis work, more details on these enzymes are given in the next part.

3. Glycoside Phosphorylases

As explained above, GPs are a class of CAZymes that are involved in the formation and cleavage of glycosidic linkages, with characteristics shared with both glycoside hydrolases (GHs) and glycosyl transferases (GTs). GPs can indeed catalyze both phosphorolysis and reverse phosphorolysis, with an

equilibrium between these two reactions (Figure 4). This is due to the fact that the free energy required for the cleavage of the glycosidic linkage is relatively close to that required for the cleavage of the ester linkage in glycosyl phosphates, as proved with a cellobiose phosphorylase (Alexander 1961).

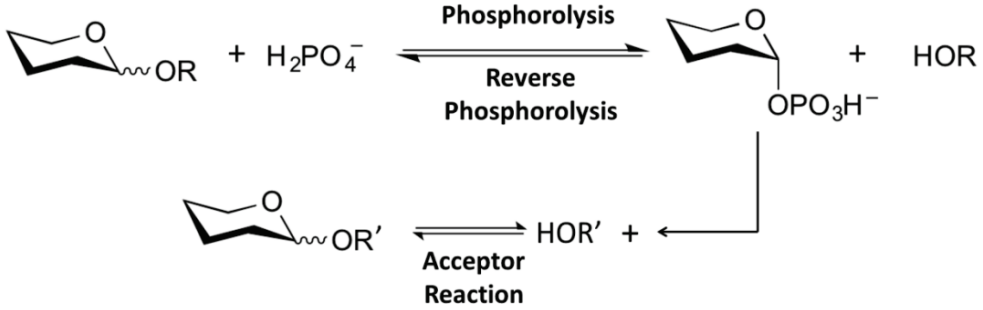


Figure 4: Reactions catalyzed by glycoside phosphorylases (From (Ladevèze 2015)

GPs can be found in both GH and GT families. The GPs biochemically characterized as of April 2020 are listed in Table 1.

Table 1: List of functionally characterized GPs. Some of the enzymes (labelled with a star) referenced in this table are not listed in the CAZy Database. * In the direction of phosphorolosis, the donor is the product of the reverse-phosphorolyolys reaction, and the acceptor is inorganic phosphate.

CAZy family	Mechanism	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity*			Reference
				Donor	Acceptor	Product	
GH3	retention	AAQ05801.1	bifunctional N-acetyl-β-glucosaminidase/β-glucosidase/ β-N-acetylglucosaminide phosphorylases(Nag3) (tested only in phosphorolysis)	βGlcNAc 1P/ βGlc1P	*p*NP/DNP	*p*-Nitrophenyl*N*-acetyl-β-D-glucosaminide/*p*-nitrophenyl β-D-glucopyranoside/ 2,4-Dinitrophenyl β-D-glucopyranoside	(Macdonald, Blaukopf, and Withers 2015)
GH3		AUG44408.1	β-glucoside phosphorylase (BglP) (tested only in phosphorolysis)	βGlc1P	DNP	2,4-Dinitrophenyl β-D-glucopyranoside	(Macdonald et al. 2018)
GH13_18	retention	2207198A	sucrose phosphorylase(SPase)	αGlc1P	D-fructose	sucrose (α-D-glucopyranosyl-(1→2)-β-D-fructofuranoside)	(Kawasaki et al. 1996)
GH13_18		AAD40317.1	sucrose phosphorylase	αGlc1P	D-fructose	sucrose	(Trethewey et al. 2001)
GH13_18		AAN24362.1	sucrose phosphorylase (Spl;BL0536)	αGlc1P	D-fructose	sucrose	(M. Kim et al. 2003)
GH13_18		AAN58596.1	sucrose phosphorylase (GtfA;SMU.881)	αGlc1P	D-fructose	sucrose	(Shemesh, Tam, and Steinberg 2007)
GH13_18		AAO21868.1	sucrose phosphorylase (LaSP;GtfA2;LBA1437)	αGlc1P	D-fructose	sucrose	(Barrangou et al. 2003)
GH13_18		AAO33821.1	sucrose phosphorylase (SucP;SP;BaSP)	αGlc1P	D-fructose	sucrose	(Van den Broek et al. 2004)
GH13_18		AAO84039.1	sucrose phosphorylase(SplP)	αGlc1P	D-fructose	sucrose	(M. Kim et al. 2003)
GH13_18		AAX33736.1	sucrose phosphorylase(LmSP1)	αGlc1P	D-fructose	sucrose	(J.-H. Lee et al. 2006)
GH13_18		ABS59292.1	sucrose phosphorylase(742sp)	αGlc1P	D-fructose	sucrose	(J.-H. Lee et al. 2008)
GH13_18		ADL69407.1	6F-P-sucrose phosphorylase (SPP;TtSPP;Tthe_1921)	αGlc1P	D-fructose 6-P	sucrose 6-phosphate (α-D-glucopyranose-(1→2)-β-D-fructofuranose 6-phosphate)	(Verhaeghe et al. 2014)
GH13_18		CCA61958.1 (FQ790378)	6F-P-sucrose phosphorylase (RgSPP)	αGlc1P	D-fructose 6-P	sucrose 6-phosphate	(Tauzin et al. 2019)
GH13_18		BAN03569.1	sucrose 6(F)-phosphate phosphorylase(YM304_32550)	αGlc1P	D-fructose 6-P	sucrose 6-phosphate	(Franceus et al. 2019)
GH13_18		AGK37834.1	sucrose phosphorylase(ScrP)	αGlc1P	D-fructose	sucrose	(Teixeira et al. 2013)
GH13_18		BAA14344.1	sucrose phosphorylase (LmSPase)	αGlc1P	D-fructose	sucrose	(Koga et al. 1991)
GH13_18		BAF62433.1	sucrose phosphorylase(Spl)	αGlc1P	D-fructose	sucrose	(Hui Zhang et al. 2018)

CAZy family	Mechanism	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
				Donor	Acceptor	Product	
GH13_18	retention	CAA30846.1	sucrose phosphorylase (GftA;SmSP)	αGlc1P	D-fructose	sucrose	(Russell et al. 1988)
GH13_18		CAA80424.1*	sucrose phosphorylase	αGlc1P	D-fructose	sucrose	(Gödl et al. 2009)
GH13_18		ADP98617.1	glucosylglycerol phosphorylase(HP15_2853)	αGlc1P	D-glycerol	glucosylglycerol (α-D-glucopyranosyl-(1→2)-glycerol)	(Franceus et al. 2018)
GH13_18		ADH62582.1	glucosylglycerate phosphorylase(Mesil_0665)	αGlc1P	D-glycerate	glucosylglycerate (α-D-glucopyranosyl-(1→2)-D-glycerate)	(Franceus, Pinel, and Desmet 2017)
GH13_18		AEJ61152.1	glucosylglycerate phosphorylase (Spith_0877)	αGlc1P	D-glycerate	glucosylglycerate	(Franceus, Pinel, and Desmet 2017)
GH13_18		AAC74391.2	glucosylglycerate phosphorylase (YcjM)	αGlc1P	D-glycerate	glucosylglycerate	(Franceus, Pinel, and Desmet 2017)
GH65	inversion	AAV43670.1	maltose phosphorylase (MalP;LBA1870)	βGlc1P	D-glucose	maltose (α-D-glucosyl-(1→4)-D-glucose)	(Andersen et al. 2012)
GH65		ADH99560.1	maltose phosphorylase (Bsel_2056)	βGlc1P	D-glucose	maltose	(Nihira, Saito, Kitaoka, Otsubo, et al. 2012)
GH65		BAC54904.1	maltose phosphorylase(MPase)	βGlc1P	D-glucose	maltose	(Inoue, Yasutake, et al. 2002)
GH65		BAD97810.1	maltose phosphorylase(MapA)	βGlc1P	D-glucose	maltose	(Y. Hidaka et al. 2005)
GH65		CAA11905.1	maltose phosphorylase	βGlc1P	D-glucose	maltose	(Ehrmann and Vogel 1998)
GH65		AAO80764.1	maltose phosphorylase (MalP;EF0957)	βGlc1P	D-glucose	maltose	(Nihira, Nishimoto, et al. 2014)
GH65		Q7SIE1	maltose phosphorylase	βGlc1P	D-glucose	maltose	(Hüwel et al. 1997)
GH65		ABX42243.1	nigerose phosphorylase (Cphy_1874)	βGlc1P	D-glucose	nigerose (α-D-glucosyl-(1→3)-D-glucose)	(Nihira, Nakai, et al. 2012)
GH65		AAE30762.1	kojibiose phosphorylase (KojP;KPase)	βGlc1P	D-glucose	kojibiose (α-D-glucosyl-(1→2)-D-glucose)	(Chaen et al. 1999; Yamamoto et al. 2004)
GH65		ABP66077.1	kojibiose phosphorylase (CsKP;Csac_0439)	βGlc1P	D-glucose	kojibiose	(Yamamoto et al. 2011)
GH65		AAC74398.1	kojibiose phosphorylase(YcjT)	βGlc1P	D-glucose	kojibiose	(Mukherjee, Narindoshvili, and Raushel 2018)
GH65		BAB97299.1	trehalose phosphorylase(TreP)	βGlc1P	D-glucose	trehalose (α-D-glucosyl-(1→1)-D-glucose)	(T. Nishimoto et al. 1996)
GH65		BAC20640.1	trehalose phosphorylase(TPase)	βGlc1P	D-glucose	trehalose	(Inoue, Ishii, et al. 2002)
GH65		AAK04526.1	trehalose-6-phosphate phosphorylase (TrePP;YeeA;L39593;LL0428)	βGlc1P	D-glucose 6-P	trehalose 6-P (α-D-glucosyl-(1→1)-D- glucopyranoside 6-P)	(Andersson, Levander, and Rådström 2001)
GH65		ABX41399.1	3-O-α-glucopyranosyl-L-rhamnose phosphorylase (Cphy_1019)	βGlc1P	L-rhamnoe	α-D-glucopyranosyl-(1→3)-L-rhamnose	(Nihira, Nakai, and Kitaoka 2012)
GH65		ADI00307.1	1,2-alpha-glucosylglycerol phosphorylase(Bsel_2816)	βGlc1P	D-glycerol	α-D-glucopyranosyl-(1→2)-glycerol	(Nihira, Saito, et al. 2014)
GH94	inversion	AAB95491.2	cellobiose phosphorylase(CbpA)	αGlc1P	D-glucose	cellobiose (β-D-glucopyranosyl-(1→4)-D-glucopyranose)	(Yernool et al. 2000)
GH94		AAC45510.1	cellobiose phosphorylase(CepA)	αGlc1P	D-glucose	cellobiose	(Reichenbecher, Lottspeich, and Bronnenmeier 1997)
GH94		AAD36910.1	cellobiose phosphorylase (CepA;TM1848;Tmari_1863)	αGlc1P	D-glucose	cellobiose	(Sun-Ki Kim et al. 2018)
GH94		AAL67138.1	cellobiose phosphorylase(Cbp)	αGlc1P	D-glucose	cellobiose	(Y.-K. Kim et al. 2002)
GH94		AAQ20920.1	cellobiose phosphorylase(CbP)	αGlc1P	D-glucose	cellobiose	(Nidetzky et al. 2004)
GH94		ABD80580.1	cellobiose phosphorylase 94A (Cbp;SdCBP;Sde_1318;Cep94A)	αGlc1P	D-glucose	cellobiose	(Ha et al. 2013)
GH94		ABN51514.1	cellobiose phosphorylase (Cbp;CtCBP;Cthe_0275)	αGlc1P	D-glucose	cellobiose	(Wilson et al. 2013)

CAZy family	Mechanism	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
				Donor	Acceptor	Product	
GH94	inversion	ACL76454.1	cellobiose phosphorylase (CbpA;Ccel_2109)	αGlc1P	D-glucose	cellobiose	(Xu et al. 2013)
GH94		ADU20744.1	cellobiose phosphorylase (CBP;RaCBP;Rumal_0187)	αGlc1P	D-glucose	cellobiose	(Sawano et al. 2013)
GH94		BAA25846.1	cellobiose-phosphorylase(Cbp)	αGlc1P	D-glucose	cellobiose	(Cate et al. 2014)
GH94		BAA28631.1	cellobiose phosphorylase (Cbp;CgCBP)	αGlc1P	D-glucose	cellobiose	(Liu et al. 1998)
GH94		CAB16926.1*	cellobiose phosphorylase(CepA)	αGlc1P	D-glucose	cellobiose	(Dakhova et al. 1993)
GH94		ADU22883.1	cellodextrin phosphorylase (CDP;RaCDP;Rumal_2403)	αGlc1P	$[(1\rightarrow4)$- β-D-glucosyl$]_n$	cellodetrin $[(1\rightarrow4)$- β-D-glucosyl$]_{n+1}$	(Sawano et al. 2013)
GH94		BAB71818.1	cellodextrin phosphorylase (Cdp-ym4)	αGlc1P	$[(1\rightarrow4)$- β-D-glucosyl$]_n$	$[(1\rightarrow4)$- β-D-glucosyl$]_{n+1}$	(Sawano et al. 2013)
GH94		ABN54185.1	cellodextrin-phosphorylase (Cdp;CtCDP;Cthe_2989)	αGlc1P	$[(1\rightarrow4)$- β-D-glucosyl$]_n$	$[(1\rightarrow4)$- β-D-glucosyl$]_{n+1}$	(Kawaguchi et al. 1998)
GH94		ADZ85667.1*	cellodextrin phosphorylase (Cdp;ClCDP;Clole_3989)	αGlc1P	$[(1\rightarrow4)$- β-D-glucosyl$]_n$	$[(1\rightarrow4)$- β-D-glucosyl$]_{n+1}$	(Ha et al. 2013)
GH94		AAC45511.1	cellodextrin phosphorylase (CepB;CsCdP)	αGlc1P	$[(1\rightarrow4)$- β-D-glucosyl$]_n$	$[(1\rightarrow4)$- β-D-glucosyl$]_{n+1}$	(Reichenbecher, Lottspeich, and Bronnenmeier 1997)
GH94		BAJ10826.1	laminaribiose phosphorylase (LbpA)	αGlc1P	D-glucose	laminaribiose (β-D-glucopyranosyl-(1→3)-D-glucopyranose)	(Kuhaudomlarp, Walpole, et al. 2019)
GH94		ABX81345.1	laminaribiose phosphorylase (ACL_0729;ACL0729)	αGlc1P	D-glucose	laminaribiose	(Nihira, Saito, Kitaoka, Nishimoto, et al. 2012)
GH94		AAG23740.1	diacetylchitobiose phosphorylase (ChbP)	αGlcNAc 1P	N-acetyl-D-glucosamine	N,N'-diacetylchitobiose (N-acetyl-β-D-glucosaminyl-(1→4)-N-acetyl-D-glucosamine)	(Keyhani, Li, and Roseman 2000)
GH94		BAC87867.1	chitobiose phosphorylase (ChbP;VpChBP)	αGlcNAc 1P	N-acetyl-D-glucosamine	N,N'-diacetylchitobiose	(Honda, Kitaoka, and Hayashi 2004)
GH94		CAC97070.1	β-1,2-oligoglucan phosphorylase (LiSOGP;Lin1839)	αGlc1P	$[(1\rightarrow2)$- β-D-glucosyl$]_n$	$[(1\rightarrow2)$- β-D-glucosyl$]_{n+1}$	(Nakajima et al. 2014)
GH94		ABX41081.1	β-1,2-oligoglucan phosphorylase (LpSOGP;Cphy_0694)	αGlc1P	$[(1\rightarrow2)$- β-D-glucosyl$]_n$	$[(1\rightarrow2)$- β-D-glucosyl$]_{n+1}$	(Nakajima et al. 2017)
GH94		AAM43298.1	cellobionic acid phosphorylase (CelAP;NdvB;XCC4077)	αGlc1P	D-gluconate	β-D-glucopyranosyl-(1→4)-D-gluconate	(Nihira, Saito, et al. 2013)
GH94		ABD80168.1	cellobionic acid phosphorylase 94B (Cep94B;CBAP;SdCBAP;Sde_0906)	αGlc1P	D-gluconate	β-D-glucopyranosyl-(1→4)-D-gluconate	(Nihira, Saito, et al. 2013)
GH94		EAA28929.1	cellobionic acid phosphorylase (CelAP;NdvB;NCU09425)	αGlc1P	D-gluconate	β-D-glucopyranosyl-(1→4)-D-gluconate	(Nihira, Saito, et al. 2013)
GH94		ACD71661.1	cyclic β-1,2-glucan synthetase (Cgs)/ β-1,2-oligoglucan phosphorylase	αGlc1P	$[(1\rightarrow2)$- β-D-glucosyl$]_n$	$[(1\rightarrow2)$-β-D-glucosyl$]_{n+1}$	(Iñón de Iannino et al. 1998; Guidolin et al. 2009)
GH112	inversion	ACB74662.1	D-galactosyl-β-1,4-L-rhamnose phosphorylase (GalRhaP;Oter_1377)	αGal1P	L-rhamnose	β-D-galactosyl-(1→4)-L-rhamnose	(Nakajima, Nishimoto, and Kitaoka 2010)
GH112		ABX42289.1	D-galactosyl-1,4-L-rhamnose phosphorylase (Cphy_1920)	αGal1P	L-rhamnose	β-D-galactosyl-(1→4)-L-rhamnose	(Nakajima, Nishimoto, and Kitaoka 2009b)
GH112		AAO07997.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase (GalGlyNAcP;VV2_1091)	αGal1P	N-acetyl-D-glucosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine	(Nakajima and Kitaoka 2008)
GH112		ABX40964.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase (Cphy_0577)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Nakajima, Nishimoto, and Kitaoka 2009b)
GH112		ABX43387.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase (Cphy_3030)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Nakajima, Nishimoto, and Kitaoka 2009b)
GH112		BAD80751.1	galacto-N-biose / lacto-N-biose phosphorylase (LnpA1;LnbP;GLNBP;BLLJ_1623)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Kitaoka, Tian, and Nishimoto 2005)

CAZy family	Mechanism	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
				Donor	Acceptor	Product	
GH112	inversion	BAD80752.1	lacto-N-biose phosphorylase (LnpA1;LnbP)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Kitaoka, Tian, and Nishimoto 2005)
GH112		ACV29689.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase / galacto-N-biose/lacto-N-biose I phosphorylase (GLNBP;Apre_1669)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Chao, Wim, and Tom 2011)
GH112		ZP_05748149.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase / galacto-N-biose phosphorylase (GNBP;HMPREF0357_1319)	αGal1P	N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Chao, Wim, and Tom 2011)
GH112		ACZ00636.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase / galacto-N-biose phosphorylase (GNBP;Smon_0146)	αGal1P	N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Chao, Wim, and Tom 2011)
GH112		ABG83511.1	galacto-N-biose phosphorylase (CPF_0553)	αGal1P	N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Nakajima et al. 2008)
GH112		BAH10636.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase (GnpA)	αGal1P	N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Nakajima, Nishimoto, and Kitaoka 2009a)
GH130	inversion	CAC96089.1	β-1,2-mannobiose phosphorylase (Lin0857)	αMan1P	D-mannose	β-1,2-mannobiose	(Tsuda et al. 2015)
GH130		ABY93074.1	β-1,2-mannobiose phosphorylase (Teth514_1789)	αMan1P	D-mannose	β-1,2-mannobiose	(Chiku et al. 2014)
GH130		ABY93073.1	β-1,2-oligomannan phosphorylase(Teth514_1788)	αMan1P	$(\beta\text{-}1,2\text{-D-mannose})_n$	$(\beta\text{-}1,2\text{-mannose})_{n+1}$	(Chiku et al. 2014)
GH130		CAZ94304.1	β-1,3-mannooligosaccharide phosphorylase (zobellia_231)	αMan1P	$(\beta\text{-}1,3\text{- D-mannose})_n$	$(\beta\text{-}1,3\text{-mannose})_{n+1}$	(Awad et al. 2017)
GH130		AAS19693.1	β-1,4-mannosylglucose phosphorylase (Unk1)	αMan1P	D-glucose	β-mannopyranosyl-(1→4)-D-glucopyranose	(Ye et al. 2016)
GH130		ADU21379.1	β-1,4-mannosylglucose phosphorylase (RaMP1;RaMGP;Rumal_0852)	αMan1P	D-glucose	β-mannopyranosyl-(1→4)-D-glucopyranose	(Jaito et al. 2014)
GH130		CAH06518.1	β-1,4-mannosylglucose phosphorylase (MGP;BF0772)	αMan1P	D-glucose	β-mannopyranosyl-(1→4)-D-glucopyranose	(Senoura et al. 2011)
GH130		VCV21228.1	β-1,4-mannosylglucose phosphorylase (RIL182_01099;ROSINTL182_07685)	αMan1P	D-glucose	β-mannopyranosyl-(1→4)-D-glucopyranose	(La Rosa et al. 2019)
GH130	inversion	AAO76140.1	β-1,4-mannosyl-N-acetyl-glucosamine phosphorylase (BT1033)	αMan1P	N-acetyl-D-glucosamine	β-mannopyranosyl-(1→4)-N-acetyl-D-glucosamine	(Nihira, Suzuki, et al. 2013)
GH130		ADD61463.1	β-1,4-mannopyranosyl-[N-glycan] phosphorylase /β-1,4-mannopyranosyl-chitobiose phosphorylase (Uhgb_MP)	αMan1P	N-acetyl-D-glucosamine-N-acetyl-D-glucosamine	β-1,4-mannopyranosyl-chitobiose (β-mannopyranosyl-(1→4)-N-acetyl-β-D-glucosaminyl-(1→4)-N-acetyl-D-glucosamine)	(Ladevèze et al. 2013)
GH130		ADU20661.1	β-1,4-mannooligosaccharide phosphorylase (MOP;RaMP2;Rumal_0099)	αMan1P	$(\beta\text{-}1,4\text{-D-mannose})_n$	$(\beta\text{-}1,4\text{-mannose})_{n+1}$	(Ye et al. 2016)
GH130		VCV21229.1	β-1,4-mannooligosaccharide phosphorylase (RIL182_01100;ROSINTL182_05474)	αMan1P	$(\beta\text{-}1,4\text{-D-mannose})_n$	$(\beta\text{-}1,4\text{-mannose})_{n+1}$	(La Rosa et al. 2019)
GH130		AAD36300.1	β-1,4-mannooligosaccharide phosphorylase(TM1225)	αMan1P	$(\beta\text{-}1,4\text{-D-mannose})_n$	$(\beta\text{-}1,4\text{-mannose})_{n+1}$	(Grimaud et al. 2019)
GH130		WP_026485574.1	CalpoDRAFT_0075	αMan1P	$(\beta\text{-}1,4\text{-D-mannose})_n$	$(\beta\text{-}1,4\text{-mannose})_{n+1}$	(Chekan et al. 2014)
GH130		WP_026486530.1	CalpoDRAFT_1209	αMan1P	$(\beta\text{-}1,4\text{-D-mannose})_n$	$(\beta\text{-}1,4\text{-mannose})_{n+1}$	(Chekan et al. 2014)
GH149	inversion	/	β-1,3-glucan phosphorylase(Pro_7066)	αGlc1P	$[(1\rightarrow3)\text{-}\beta\text{-D-glucosyl}]_n$	$[(1\rightarrow3)\text{-}\beta\text{-D-glucosyl}]_{n+1}$	(Kuhaudomlarp et al. 2018)
GH149		AUO30192.1	β-1,3-glucan phosphorylase(EgP1)	αGlc1P	$[(1\rightarrow3)\text{-}\beta\text{-D-glucosyl}]_n$	$[(1\rightarrow3)\text{-}\beta\text{-D-glucosyl}]_{n+1}$	(Kuhaudomlarp et al. 2018)
GH161	inversion	WP_019688419.1	β-1,3-glucan phosphorylase (PPT_RS0121460; PapP)	αGlc1P	$[(1\rightarrow3)\text{- }\beta\text{-D-glucosyl}]_n$	$[(1\rightarrow3)\text{- }\beta\text{-D-glucosyl}]_{n+1}$	(Kuhaudomlarp, Pergolizzi, et al. 2019)
GH161		ACJ76363.1	β-1,3-glucan phosphorylase (THA_1941; TaCDP)	αGlc1P	$[(1\rightarrow3)\text{-}\beta\text{-D-glucosyl}]_n$	$[(1\rightarrow3)\text{-}\beta\text{-D-glucosyl}]_{n+1}$	(Kuhaudomlarp, Pergolizzi, et al. 2019)
GT4	retention	BAA31350.1	trehalose synthase	αGlc1P	D-glucose	trehalose (α-D-glucosyl-(1→1)-D-glucose)	(K. Saito et al. 1998)
GT4		AAF22230.1	trehalose phosphorylase	αGlc1P	D-glucose	trehalose	(S.-E. Han et al. 2003)
GT4		ABC84380.1	trehalose phosphorylase	αGlc1P	D-glucose	trehalose	(Goedl et al. 2006)

CAZy family	Mechanism	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
				Donor	Acceptor	Product	
GT35	retention	ABP51432.1	glucan/maltodextrin phosphorylase (GlgP;PyglgP;Pars_1881)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Jarrell et al. 2014)
GT35		AAL81659.1	α-glucan/maltodextrin phosphorylase (PF1535)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(H.-S. Lee et al. 2006; Mizanur, Griffin, and Pohl 2008)
GT35		AAD28735.1	maltodextrin phosphorylase (MalP)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Karina B. Xavier et al. 1999)
GT35		AAD03471.1	glycogen phosphorylase(GlgP)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Juan E Ugalde et al. 1998)
GT35		AAC06896.1	glycogen phosphorylase (GlgP;Aq_717)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Bhuiyan et al. 2003)
GT35		AAC00218.1	glycogen phosphorylase (GlgP;BSU30940)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Kiel et al. 1994)
GT35		AAM24997.1	α-glucan phosphorylase (GlgP;TTE1805)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(S. Chen et al. 2007)
GT35		AAM52219.1	glycogen phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Griessler et al. 2004)
GT35		BAB98701.1	maltodextrin/glycogen phosphorylase (MalP;GlgP1;cg1479)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Seibold, Wurst, and Eikmanns 2009)
GT35		BAB99480.1	glycogen phosphorylase (GlgP;GlgP2;cg2289)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Seibold, Wurst, and Eikmanns 2009)
GT35		AAC76453.1	glycogen phosphorylase (GlgP;b3428)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Rybak et al. 2008)
GT35		AAC76442.1	maltodextrin phosphorylase (MalP;b3417)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Palm, Goerl, and Burger 1985; O'Reilly et al. 1997)
GT35		BAA19592.1	glycogen phosphorylase (GlgP)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Takata et al. 1997)
GT35		ABN51595.1	α-glucan phosphorylase (α-GP;Cthe_0357)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Wilson et al. 2013)
GT35		AAD53957.1	glycogen phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Igarashi and Meyer 2000)
GT35		AAN59210.1	maltodextrin phosphorylase (GlgP;SMU.1564)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Sato, Okamoto-Shibayama, and Azuma 2013)
GT35		AAL26558.1	glycogen phosphorylase (Glg)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Koksharova , Brandt, and Cerff 2004)
GT35		AGL50099.1	α-glucan phosphorylase (AgpA;Tmari_1175)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Bibel et al. 1998)
GT35		BAB11741.1	α-glucan phosphorylase (GlgP)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(T. (Ezaki G. C. L. Takaha et al. 2001)
GT35		CAC93400.1	glycogen phosphorylase (glgP,YPO3938)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Alonso-Casajús et al. 2006)
GT35		AAB46846.1	glycogen phosphorylase (myophosphorylase)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Seiichi Tsujino et al. 1996)
GT35		ABB88567.1	plastidial starch phosphorylase (PhoB;Sta4)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Dauvillée et al. 2006)
GT35		CAA44069.1	glycogen phosphorylase 1 (GlpV;GP1)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Rogers et al. 1992)
GT35		AAA33211.1	glycogen phosphorylase 2 (GlpD;GP2)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Jones and Wright 1970)
GT35		AAD46887.1	glycogen phosphorylase (GlyP;GLYP;CG7254;Dmel_CG 7254)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Gabriella Tick et al. 1999)
GT35		AAN17338.1	glycogen phosphorylase-2	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Wu and Müller 2003)
GT35		AAL23578..1	glycogen phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Wu and Müller 2003)
GT35		AAP33020.1	glycogen phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Takeyasu, Kawase, and Yoshimura 2003)
GT35		AAK69600.1	glycogen phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Wu and Müller 2003)
GT35		AAB60395.1	glycogen phosphorylase (brain) (bGP)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Newgard et al. 1988)
GT35		CAA75517.1	glycogen phosphorylase (liver)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Newgard et al. 1986)
GT35		AAC17451.1	glycogen phosphorylase (muscle) (PigM)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Burke et al. 1987)

CAZy family	Mechanism	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
				Donor	Acceptor	Product	
GT35	retention	BAK00834.1	plastidial α-1,4-glucan phosphorylase (Pho1;HvPho1)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Cuesta-Seijo et al. 2017)
GT35		AAA63271.1	α-glucan phosphorylase L	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(C. T. Lin et al. 1991)
GT35		AAK01137.1	starch phosphorylase (fragment)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Lu et al. 2006)
GT35		AAL23577.1	glycogen phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Wu and Müller 2003)
GT35		AAD30476.1	glycogen phosphorylase (muscle) (PygM)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Schliselfeld and Danon 2002)
GT35		AAG00588.1	glycogen phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(McInerney et al. 2002)
GT35		ACJ76617.1	glycogen phosphorylase (PygM) (muscle)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Koide et al. 1978; Hermann et al. 1978)
GT35		AAK15695.1	α-1,4-glucan phosphorylase L	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Nishi et al. 2001; Jha and Dubey 2004)
GT35		BAB92854.1	α-1,4-glucan phosphorylase (Os01g0851700)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Nishi et al. 2001; Jha and Dubey 2004)
GT35		AAV87308.1	brain glycogen phosphorylase (PYGB)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Philips et al. 2014)
GT35		AAB68800.1	glycogen phosphorylase (muscle)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Tan et al. 1997)
GT35		AAA41252.1	glycogen phosphorylase (Brain)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Hudson, Hefferon, and Crerar 1993)
GT35		AAH70901.	glycogen phosphorylase (liver)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Schiebel, Pekel, and Mayer 1992)
GT35		AAA41253.1	glycogen phosphorylase (muscle)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Hudson, Hefferon, and Crerar 1993)
GT35		AAB68057.1	glycogen phosphorylase (Gph1;YPR160w)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Hwang and Fletterick 1986)
GT35		AAA33809.1	α-glucan phosphorylase H	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Mori, Tanizawa, and Fukui 1991)
GT35		BAA00407.1	α-glucan phosphorylase L1	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Mori, Tanizawa, and Fukui 1991)
GT35	retention	CAA52036.1	α-glucan phosphorylase L2	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Sonnewald et al. 1995)
GT35		CAA59464.1	α-glucan phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Duwenig, Steup, and Kossmann 1997)
GT35		AAL23579.1	glycogen phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Wu and Müller 2003)
GT35		AAF82787.1	α-glucan phosphorylase	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Schupp and Ziegler 2004)
GT35		CAA84494.1	α-glucan phosphorylase (Pho2;VfPho2)	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Buchner, Borisjuk, and Wobus 1996)
GT35		CAA85354.1	α-glucan phosphorylase L	αGlc1P	$[(1\rightarrow4)$-α-D-glucosyl$]_n$	$[(1\rightarrow4)$-α-D-glucosyl$]_{n+1}$	(Buchner, Borisjuk, and Wobus 1996)
GT108	inversion	CBZ24448.1	MTP3 (LMXM_10_1250)	αMan1P	(β-1,2-D-mannose)$_n$	(β-1,2-mannose)$_{n+1}$	(Sernee et al. 2019)
GT108		CBZ24449.1	MTP4 (LMXM_10_1260)	αMan1P	(β-1,2-D-mannose)$_n$	(β-1,2-mannose)$_{n+1}$	(Sernee et al. 2019)
GT108		CBZ24451.1	MTP6 (LMXM_10_1280)	αMan1P	(β-1,2-D-mannose)$_n$	(β-1,2-mannose)$_{n+1}$	(Sernee et al. 2019)
GT108		CBZ24452.1	MTP7 (LMXM_10_1290)	αMan1P	(β-1,2-D-mannose)$_n$	(β-1,2-mannose)$_{n+1}$	(Sernee et al. 2019)

3.1. GPs catalytic mechanisms

Due to the high similarities in the structure and mechanisms, GPs perform the phosphorolysis and reverse phosphorolysis reactions using the same catalytic features as GHs and GTs. GPs can be divided into retaining GPs and inverting GPs based on the change in the anomeric carbon configuration after the

reaction (Koshland 1953). For example, inverting phosphorylases change the anomeric configuration in the product, yielding α-glycosyl phosphates from β-glycosides or β-glycosyl phosphates from α-glycosides.

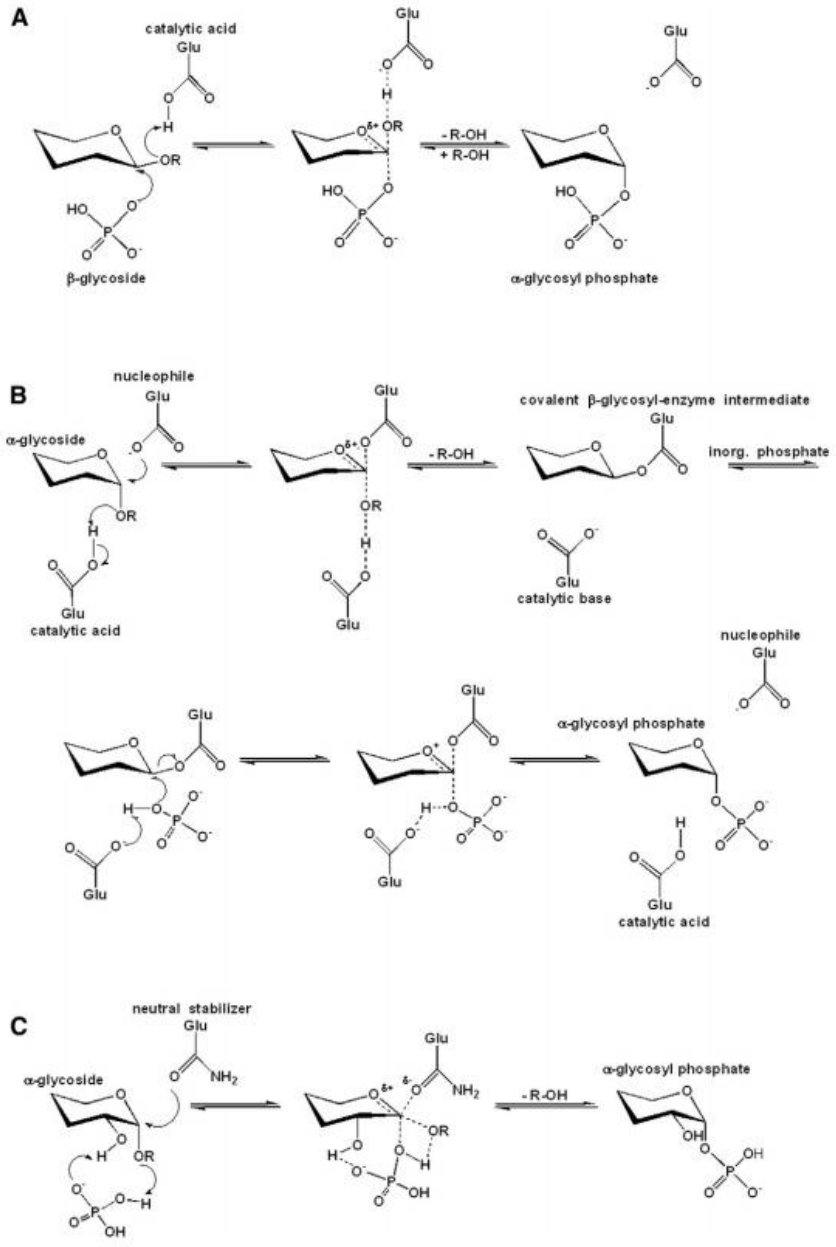


Figure 5: Schematic presentation of the catalytic mechanisms of inverting GPs (A, example of a GP acting on β-glycosides), retaining GPs operating via double displacement mechanism (B, example of a GP acting on α-glycosides), and retaining glycoside phosphorylases using front-side displacement (internal return) mechanism (C, example of a GP acting on α-glycosides). (From (Puchart 2015))

Both inverting GHs and GPs follow a direct displacement SN_2-like reaction mechanism (Figure 5) but with an important difference regarding the catalytic machinery that is required to perform the reactions (Puchart 2015). For the inverting GHs, two catalytic amino acid residues are required: a proton donor that donates a proton to the anomeric carbon, and a catalytic base that can remove the proton from a

water molecule. In inverting GPs (classified in both GT and GH families, Table 1), only one catalytic residue is required, and the role of the catalytic base is played by the inorganic phosphate (Pi). The inversion of the anomeric configuration is achieved in one step: the C_1 position at the non-reductive end of the sugar is attacked by the catalytic base , the proton is then captured by the inter-osidic oxygen on the proton donor, leading to the formation of an oxocarbenium ion-like transition state. Finally, bond cleavage is achieved between the two glycosyl residues with the inter-osidic oxygen being held on the monomer at the reductive-end, which yields a glycosyl-phosphate of inverted configuration and an oligosaccharide of reduced chain length.

Most retaining GPs (classified in both GT and GH families, Table 1) act on α-glycosides. For these enzymes, the anomeric configuration between the newly formed α-glycosidic bond of the product and the α-glycosyl phosphate substrate is maintained. Some members of the GH3 family are the only reported retaining GP which could convert β-glycosidic substrates to β-glycosidic phosphate (Macdonald, Blaukopf, and Withers 2015). The mechanism of retaining GPs is close to that of retaining GTs. Retaining GPs can operate either via a double displacement mechanism to form a glycosyl-enzyme intermediate, or via a front-side displacement mechanism (Puchart 2015) (Figure 5).

Double displacement mechanism occurs in two steps. First, the nucleophile attacks the C_1 of the sugar at the non-reductive end, concomitantly to the proton capture by the inter-osidic oxygen on the proton donor. Contrary to inverting GPs, a catalytic nucleophilic residue is involved here. The same transition state as for inverting GPs is observed, which leads to the cleavage of the inter-osidic oxygen-C_1 bond, and yields in that case a glycosyl-enzyme intermediate plus a glycoside of reduced chain length that may leave the catalytic site. Then, the proton donor recovers its proton by activating the inorganic phosphate which is involved in a concomitant nucleophilic attack onto the C_1 of the glycosyl-enzyme intermediate. The formed transition state is closely related to the one formed in the first step, stabilized by the delocalisation of an electronic lone pair of the ring oxygen. The final step is the cleavage of the glycosyl-enzyme bond, yielding a glycoside-phosphate of retained configuration.

The likely front-side displacement mechanism (also known as internal return-like or *SNi*-like) used by GPs of the GT4 and GT35 families, only requires one catalytic amino acid (Goedl and Nidetzky 2009). Only the nucleophile is required since the role of proton donor is played by the inorganic phosphate itself. One phosphate hydroxyl is providing the proton caught by the glycosidic oxygen, which is accompanied by a nucleophilic attack. It is important to note that both the protonation and the nucleophilic attack occur on the same side of the sugar ring to ensure the retention of configuration. The oxocarbenium-ion transition state, is stabilized by the nucleophile which is, in that case, a glutamine or an asparagine.

3.2. Classification and substrate specificity

In the CAZy database, GPs are classified into a total of 11 CAZymes families (Table 1). Let's note that a N-acetylglucosaminidase from the GH84 family has recently been transformed into an efficient GP by a single point mutation (Teze et al. 2020). As no native GH84 enzyme has been reported to date to act as a GP, the GH84 family will not be mentioned anymore in this chapter.

3.2.1. Retaining GPs

Retaining GPs are found in two GHs families (GH3 and GH13_18) and two GTs families (GT4 and GT35).

Only two characterized GPs have been identified, to date, in the large GH3 family, which contain 30,514 members (April 2020). The Nag3 from *Cellulomonas fimi*, which was previously characterized as a bifunctional N-acetyl-β-glucosaminidase and β-glucosidase (Mayer et al. 2006) was in fact found to be also a β-N-acetylglucosaminide phosphorylase that could produce β-D-N-acetyl-glucosamine-1-phosphate (βGlcNAc1P). This new enzyme activity is due to the presence of a His as a catalytic acid/base residue, instead of the canonical Glu or Asp. The preferred reaction of Nag3 is the cleavage of the disaccharide GlcNAc-anhydro-MurNAc for the recycling of peptidoglycan (Reith and Mayer 2011; Macdonald, Blaukopf, and Withers 2015). Another β-glucoside phosphorylase, BglP, was further discovered by activity-based metagenomics. BglP phosphorolyses the β-1,4-glucosidic linkages of cellulose and cello-oligosaccharides (Macdonald et al. 2018). The other retaining GPs display an α-linkage specificity, such as the retaining GPs from the GH13_18 subfamily, which act on α-glucosides to produce α-glucose 1-phosphate (αGlc1P).

The GH13_18 subfamily was thought for a long time to contain only sucrose phosphorylases. Nevertheless, sucrose 6'-phosphate phosphorylases (releasing αGlc1P and α-D-fructose-6-phosphate) were recently discovered from *Thermoanaerobacterium thermosaccharolyticum* (Verhaeghe et al. 2014) and *Ruminococcus gnavus* (Tauzin et al. 2019). Later on, a novel glucosylglycerate phosphorylase from *Meiothermus silvanus* was discovered (Franceus, Pinel, and Desmet 2017). This enzyme is involved in the metabolism of α-1,2-glucosylglycerate, a molecule known to protect cells from water loss (Ferjani et al. 2003). Then, an α-1,2-glucosylglycerol phosphorylase from *Marinobacter adhaerens* was identified, which could allow the bacterium to tolerate high salt concentrations, thanks to the protective effect of glucosylglycerol (Franceus et al. 2018). The sucrose phosphorylases and related enzymes from GH13, their applications and engineering were recently reviewed (Franceus and Desmet 2020).

The GPs from the GT4 family are all trehalose phosphorylases (Kitamoto et al. 1988; Koki Saito et al. 1998; S.-E. Han et al. 2003; Goedl et al. 2006). Finally, all the characterized members of the GT35 family are α-1,4-glucan phosphorylases acting on glycogen, starch, and maltodextrins (G. T. Cori and Cori 1943; Y. P. Lee 1960).

3.2.2. Inverting GPs

Inverting GPs are currently grouped in 7 GHs families (GH65, GH94, GH112, GH130, GH149, GH161) and one GT family harboring dual glycosyltransferase/phosphorylase activities (GT108) (Sernee et al. 2019).

All those of the GH65 family are involved in the cleavage of α-glucosides to produce β-glucose 1-phosphate (βGlc1P). The first GP from this family was identified from *Lactobacillus sanfranciscensis*. This enzyme could effectively convert maltose and phosphate into βGlc1P and D-glucose (Ehrmann and Vogel 1998). Other enzymes with new specificities, playing key roles in various metabolic pathways, were further identified, such as the trehalose-6-phosphate phosphorylase from *Lactococcus lactis subsp. Lactis* (Andersson, Levander, and Rådström 2001) , the trehalose phosphorylase from *Thermoanaerobacter brockii* (Maruta et al. 2002), the kojibiose phosphorylase from *Thermoanaerobacter brockii*(Yamamoto et al. 2004), the nigerose phosphorylase and the 3-O-α-D-Glucopyranosyl-L-rhamnose phosphorylase from *Clostridium phytofermentans* (Nihira, Nakai, et al. 2012; Nihira, Nakai, and Kitaoka 2012), and the 2-O-α-D-glucosylglycerol phosphorylase from *Bacillus selenitireducens*, which utilizes polyols like glycerol as acceptors instead of longer oligosaccharides or polysaccharides, and could be involved in both the biosynthesis and cleavage of 2-O-α-D-

glucosylglycerol, allowing the microorganism adaptation to high concentration of salts in the environment (Nihira, Saito, et al. 2014).

The inverting GPs from the GH94 family act on various β-linked glycosides to produce αGlc1P or α-D-N-acetyl-glucosamine-1-phosphate (αGlcNAc1P). The first reported GPs in this family were the cellobiose phosphorylase and cellodextrin phosphorylases from *Clostridium stercorarium* (Reichenbecher, Lottspeich, and Bronnenmeier 1997). Then, a N,N'-diacetylchitobiose phosphorylase was identified from *Vibrio furnissii*, which is able to generate α-D-N-acetyl-glucosamine-1-phosphate (J. K. Park, Keyhani, and Roseman 2000). All the other GH94 GPs which were further discovered are specific for αGlc1P. They include enzymes acting on various β-glucosidic linkages. The laminaribiose phosphorylase from *Paenibacillus* sp. (Kitaoka et al. 2012)) plays a key role in the metabolism of paramylon, an intracellular storage form of β-1,3-glucan usually found in *Euglena* species (Kiss, Vasconcelos, and Triemer 1987), while the laminaribiose phosphorylase from *Acholeplasma laidlawii* could effectively produce shorter 1,3-β-D-glucosyl disaccharides (Nihira, Saito, Kitaoka, Nishimoto, et al. 2012). The fungal (*Neurospora crassa*) and bacterial (*Xanthomonas campestris*) cellobionic acid phosphorylases catalyze the reversible phosphorolysis of 4-O-β-D-glucopyranosyl-D-gluconic acid, a motif found in recalcitrant cellulosic biomass (Nihira, Saito, et al. 2013). Finally, the β-1,2-oligoglucan phosphorylase from *Listeria innocua* could be involved in the metabolism of sophorose-containing glucans and/or exogenous 1,2-β-glucans in specific environmental conditions (Nakajima et al. 2014).

The GH112 family is constituted of inverting phosphorylases acting on β-galactosides to release α-D-galactose-1-phosphate (αGal1P). To date, all the characterized GPs from GH112 family contained D-galactopyranosyl-1,3-N-acetylhexosamine phosphorylases and D-galactosyl-β-1,4-L-rhamnose phosphorylases from bacterial origin. β-D-galactopyranosyl-1,3-N-acetylhexosamine phosphorylases were further subdivided based on their substrate preference into galacto-N-biose phosphorylase (GNBP, with a preference for galacto-N-biose) (Nakajima et al. 2008; Nakajima, Nishimoto, and Kitaoka 2009a), lacto-N-biose I phosphorylase (LNBP, with a preference for lacto-N-biose I) (Nakajima and Kitaoka 2008), and galacto-N-biose/lacto-N-biose I phosphorylase (GLNBP, with no clear preference for either galacto-N-biose or lacto-N-biose I) (Derensy-Dron et al. 1999; Nakajima, Nishimoto, and Kitaoka 2009a). GNB is the core component of gastrointestinal mucin and LNB I is a main building block of human milk oligosaccharides (HMOs) and both are present in the intestinal environment (Kitaoka, Tian, and Nishimoto 2005). Besides, the D-galactosyl-β-1,4-L-rhamnose phosphorylase from *Clostridium phytofermentans* probably plays a role in the degradation of rhamnogalacturonan I (RG-I), an important component of pectin (Nakajima, Nishimoto, and Kitaoka 2009a).

The GH130 family contains inverting phosphorylases that act on β-mannosides to generate α-D-mannose 1-phosphate (αMan1P). The identification of a 4-O-β-D-mannosyl-D-glucose phosphorylase activity led to the creation of the GH130 family (Senoura et al. 2011). Then, a lot of attention was paid to this family to discover mannoside phosphorylases with new specificities, in particular in the group of G. Potocki-Veronese, where I performed my PhD work. In 2016, when I started to work on GPs, GH130 GPs with many different substrate specificities had already been identified, highlighting their role in the catabolism of plant, mammal and fungal mannosides, in particular in the human gut where these bacterial enzymes are abundant (Figure 6) (Ladevèze 2015).

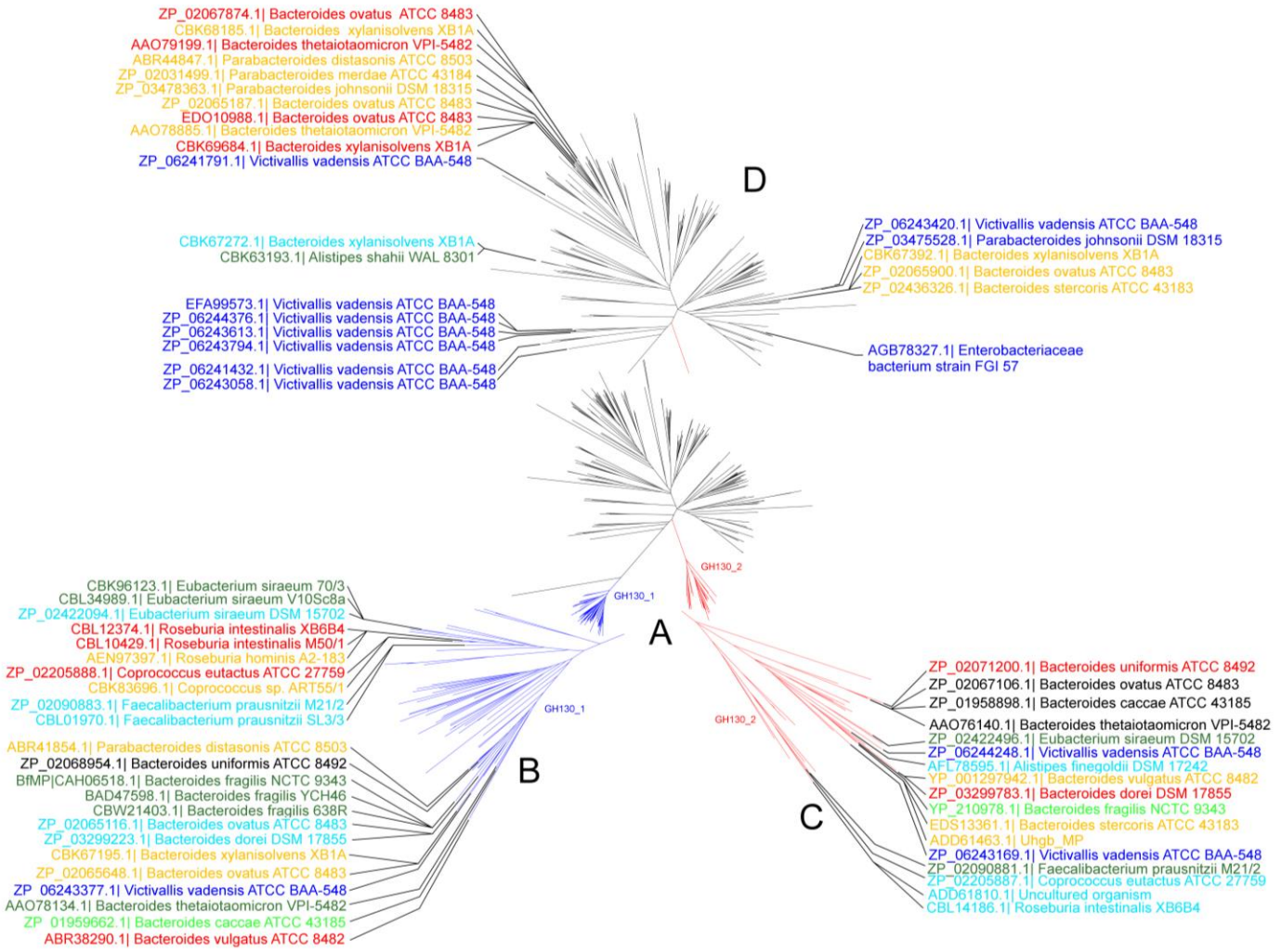


Figure 6: A: Phylogenetic tree of the GH130 family, in 2013. In this radial phylogram, branches corresponding to subfamilies GH130_1 and GH130_2 are coloured in blue and red respectively. Branches in black correspond to proteins that are not classified into any subfamily (GH130_NC). B, C, D: Detailed views of the phylogentic tree of the GH130_1, GH130_2 subfamilies, and of not assigned sequences (GH130_NC). Only sequences corresponding to proteins from human gut bacteria are shown. Label colour is relative to their prevalence in the human gut metagenome of 301 different individuals: Blue: sequences found in the metagenome of 0 to 5 individuals. Light blue: sequences found in the metagenome of 24 to 40 individuals. Green: sequences found in the metagenome of 44 to 59 individuals. Light green: sequences found in the metagenome of 62 to 76 individuals. Orange: sequences found in the metagenome of 83 to 118 individuals. Red: sequences found in the metagenome of 121 to 153 individuals. Black: Sequences found in the metagenome of 173 to 197 individuals. (Modified from (Ladevèze et al. 2013))

The β-1,4 mannooligosaccharide phosphorylases firstly discovered by Kawahara and colleagues (Kawahara et al. 2012), and the 4-O-β-D-mannosyl-D-glucose phosphorylases, participate in the breakdown of plant cell wall mannans and glucomannans. The β-1,4-mannosyl-N-acetyl-glucosamine phosphorylase and β-1,4-mannopyranosyl-chitobiose phosphorylase are both involved in the catabolism of eukaryotic N-glycans (Nihira, Suzuki, et al. 2013; Ladevèze et al. 2013). The β-1,2-mannobiose phosphorylases (Chiku et al. 2014; Tsuda et al. 2015) and β-1,2-oligomannan phosphorylases (Chiku et

al. 2014) act on exogenous yeast mannan. During my PhD project, a β-1,3-mannobiose phosphorylase has been discovered (Awad et al. 2017), but its physiological function was not described.

In 2013, Ladevèze and colleagues indicated that based on sequence similarities, the GH130 family could be subdivided into at least two subgroups (Figure 6). The GH130_1 one would include enzymes strictly acting on β-D-Man*p*-1,4-D-Glc, and the GH130_2 one, those acting on β-1,4-mannosides. The authors also mentioned the presence of a large heterogeneous sequence cluster, named GH130_NC, which was further shown to include β-1,2-mannosidases (Cuskin, Lowe, et al. 2015) and β-1,3-mannobiose phosphorylases (Awad et al. 2017). This highlights the need to re-evaluate the sequence and functional diversity of these enzymes, as described further in Chapter 2.

GH149 and GH161 are newly created families, with enzymes from the photosynthetic excavate *Euglena gracilis*, the heterokont *Ochromonas* spp. and from bacteria that act on β-1,3-glucans, producing αGlc1P. These glycosides can be found in bacteria, fungi, plants, and algae, and are used in a wide range of biotechnological applications, including as ingredients in agricultural, food, cosmetic, and therapeutic products. The members of the GH149, GH161 and GH94 families have the consensus amino acid characteristic of a GP activity, likely due to the diversification by evolution from a common ancestor. However, their sequences share only around 20 % identity, which led to the creation of the distinct GH149 (Kuhaudomlarp et al. 2018) and GH161 families (Kuhaudomlarp, Pergolizzi, et al. 2019).

The GT108 is also a newly created CAZy family, in which were discovered four GPs acting on mannogen, a β-1,2-mannan found in in Leishmania parasites for energy storage. Interestingly, these four enzymes are dual glycosyltransferase-phosphorylases, which participate to the recycling of mannogen. These Leishmania enzymes have the same fold and similar reaction mechanism as bacterial GH130 mannan phosphorylases or hydrolases, although they significantly differ at the sequence level to be classified in another family(Sernee et al. 2019).

As described above, GPs can thus degrade and synthesize a large variety of glycosidic linkages, except α-/β-1,6 and β-1,1 linkages. As of March 2020, only 159 phosphorylases were characterized (Table 1), which represents around 1% of all the characterized CAZymes (Garron and Henrissat 2019). Taking into consideration the huge number of GTs and GHs that are listed in the CAZy database, there might be many GPs with novel specificities that are yet to be discovered.

3.3. Tertiary and quarternary structures

With the exception of the GT4 retaining trehalose phosphorylases and GH161 inverting β-1,3 glucan phosphorylases, for which no structure has been elucidated yet, at least one GP crystallographic structure is available for each family in the protein data bank (PDB) (Table 2).

Thanks to these structures, in particular ones in complex with substrates and products, the catalytic residues have been identified, together with the residues involved in phosphate binding. This complements the site-directed mutagenesis studies that have been performed to identify the residues that are critical for GP activity (for example, (Ladevèze et al. 2013) for the GH130 family). In addition, if several GP members from a same family have been functionally and structurally characterized, structural comparison allowed to identify the molecular determinants of GP activity and of the donor and acceptor specificities ((Ladevèze 2015; Cuskin, Baslé, et al. 2015) for the GH130 family, for example).

The vast majority of GPs crystallize and act in solution as dimers. However, several GPs crystallize as monomers or as large homooligomeric complexes, like the Uhgb_MP enzyme, for which the catalytic pocket results from the assembly of the different monomers (Table 2). The topology of GP active sites

is thus difficult to predict (Ladevèze et al. 2015). Since my thesis work does not deal with tertiary and quaternary structures, this section will not be detailed.

Table 2: Structurally characterized GPs. Only the enzymes proved to have GP activity are listed. ACJ76617.1 have hundreds of PDB, only list one PDB with the best solution.

CAZy family	GenBank/UniProtKB accession	PDB accession number	Protein Name	Fold	Quaternary structure	Reference
GH3	AUG44408.1	5VQD[A], 5VQE[A]	β-glucoside phosphorylase BglP	(β/α)8	monomer	(Macdonald et al. 2018)
GH13_18	AAO33821.1	1R7A[A,B], 2GDU[A,B] 2GDV[A,B], 5C8B[B] 5M9X[B], 5MAN[B] 5MB2[B], 6FME[A,B]	sucrose phosphorylase (SucP;SP;BaSP)	(β/α)8	dimer	(Sprogøe et al. 2004; Mirza et al. 2006; Kraus, Grimm, and Seibel 2016; 2017)
GH13_18	BAN03569.1	6S9U[A]	sucrose 6(F)-phosphate phosphorylase (YM304_32550, IcSPP)	(β/α)8	monomer	(Franceus et al. 2019)
CAZy family	**GenBank/UniProtKB accession**	**PDB accession number**	**Protein Name**	**Fold**	**Quaternary structure**	**Reference**
GH13_18	ADL69407.1	6S9V[A,B]	sucrose 6(F)-phosphate phosphorylase (SPP;TtSPP;Tthe_1921)	(β/α)8	monomer	(Franceus et al. 2019)
GH65	Q7SIE1	1H54[A]	maltose phosphorylase (Lb-MP)	(α/α)6	dimer	(Egloff et al. 2001)
GH65	ABP66077.1	3WIQ[A], 3WIR[A,B,C,D]	kojibiose phosphorylase (CsKP;Csac_0439)	(α/α)6	dimer	(Okada et al. 2014)
GH65	/	6W0P[B,C,D], 6W0P[A]	kojibiose phosphorylase	(α/α)6	dimer	unpublished
GH65	ADI00307.1	4KTP[A,B], 4KTR[A,B,C,D,E,F,G,H]	2-O-α-glucosylglycerol phosphorylase (GGP;Bsel_2816)	(α/α)6	dimer	(Touhara et al. 2014)
GH65	BAB97299.1	/	trehalose phosphorylase (TreP)	(α/α)6	dimer	(Van Hoorebeke et al. 2010)
GH94	BAA28631.1	2CQS[A,B],2CQT[A,B] 3ACS[A,B],3ACT[A,B] 3AFJ[A,B],3QFY[A,B] 3QFZ[A,B],3QG0[A,B]	cellobiose phosphorylase (Cbp;CgCBP)	(α/α)6	dimer	(M. Hidaka et al. 2006; Fushinobu et al. 2011)
GH94	AAQ20920.1	3RRS[A,B],3RSY[A,B] 3S4A[A,B],3S4B[A,B] 3S4C[A],3S4D[A]	cellobiose phosphorylase (CbP)	(α/α)6	dimer	unpublished
GH94	AAL67138.1	3QDE[A,B]	cellobiose phosphorylase (Cbp)	(α/α)6	dimer	(Bianchetti et al. 2011)
GH94	ABN54185.1	5NZ7[A,B] 5NZ8[A,B]	cellodextrin-phosphorylase (Cdp;CtCDP;Cthe_2989)	(α/α)6	dimer	(E. C. O'Neill et al. 2017)
GH94	ABD80168.1	4ZLE[A], 4ZLF[A] 4ZLG[A], 4ZLI[A]	cellobionic acid phosphorylase 94B (Cep94B;CBAP;SdCBAP;Sde_0906)	(α/α)6	dimer	(Nam et al. 2015)
GH94	BAC87867.1	1V7V[A], 1V7W[A] 1V7X[A]	chitobiose phosphorylase (ChbP;VpChBP)	(α/α)6	dimer	(M. Hidaka et al. 2004)
GH94	BAJ10826.1	6GGY[A,B],6GH2[A,B] 6GH3[A,B]	laminaribiose phosphorylase (LbpA)	(α/α)6	dimer	(Kuhaudomlarp, Walpole, et al. 2019)
GH94	ABX41081.1	5H3Z[A,B], 5H40[A,B] 5H41[A,B], 5H42[A,B]	β-1,2-oligoglucan phosphorylase (LpSOGP;Cphy_0694)	(α/α)6	monomer	(Nakajima et al. 2017)
GH112	BAD80751.1	2ZUS[A,B,C,D], 2ZUT[A,B,C,D] 2ZUU[A,B,C,D], 2ZUV[A,B] 2ZUW[A,B,C,D], 3WFZ[A,B,C,D]	galacto-N-biose / lacto-N-biose phosphorylase (LnpA1;LnbP;GLNBP;BLLJ_1623)	(β/α)8	dimer	(M. Hidaka et al. 2009; Koyama et al. 2013)
GH130	CAH06518.1	3WAS[A,B], 3WAT[A,B] 3WAU[A,B], 4KMI[A,B]	β-1,4-mannosylglucose phosphorylase (MGP;BF0772)	5-fold β-propeller	hexamer	(Nakae et al. 2013)
GH130	ADU21379.1	5AY9[A], 5AYC[A]	β-1,4-mannosylglucose phosphorylase (RaMP1;RaMGP;Rumal_0852)	5-fold β-propeller	trimer	(Ye et al. 2016)
GH130	ADU20661.1	5AYD[A,B,C,D,E,F], 5AYE[A,B,C,D,E,F]	β-1,4-mannooligosaccharide phosphorylase (MOP;RaMP2;Rumal_0099)	5-fold β-propeller	hexamer	(Ye et al. 2016)
GH130	AAD36300.1	1VKD[A,B,C,D,E,F]	β-1,4-mannooligosaccharide phosphorylase (TM1225)	5-fold β-propeller	dimer	unpublished
GH130	ADD61463.1	4UDG[A,B,C,D,E,F] 4UDI[A,B,C,D,E,F] 4UDJ[A,B,C,D,E,F] 4UDK[A,B,C,D,E,F]	β-1,4-mannopyranosyl-chitobiose phosphorylase (UhgbMP;Uhgb_MP)	5-fold β-propeller	hexamer	(Ladevèze et al. 2015)
GH130	CAC96089.1	5B0P[A,B], 5B0Q[A,B] 5B0R[A,B], 5B0S[A,B]	β-1,2-mannobiose phosphorylase (Lin0857)	5-fold β-propeller	dimer	(Tsuda et al. 2015)
GH149	/	6HQ6[A,B], 6HQ8[A,B] 6HQ6[A]	β-1,3-oligosaccharide phosphorylase(Pro_7066)	(α/α)6	dimer	(Kuhaudomlarp, Stevenson, et al. 2019)
GT35	AAM52219.1	2C4M[A,B,C,D]	glycogen phosphorylase	GT-B	tetramer	unpublished
GT35	AAC76442.1	1AHP[A], 1E4O[A] 1L5V[A], 1L5W[A] 1L6I[A], 1QM5[A] 2ASV[A,B], 2AV6[A,B] 2AW3[A,B], 2AZD[A,B] 2ECP[A]	maltodextrin phosphorylase (MalP;b3417)	GT-B	dimer	(O'Reilly et al. 1997; Kimberly A Watson et al. 1999; Geremia et al. 2002; Campagnolo et al. 2008; O'Reilly, Watson, and Johnson 1999)
GT35	AAN59210.1	4L22[A]	maltodextrin phosphorylase (GlgP;SMU.1564)	GT-B	monomer	unpublished
GT35	CAB61943.1	4BQE[A,B], 4BQF[A,B] 4BQI[A,B]	α-glucan phosphorylase (PHS2;AtPHS2;At3g46970)	GT-B	dimer	(E. O'Neill et al. 2014)
GT35	AAB60395.1	5IKO[A], 5IKP[A]	glycogen phosphorylase (brain) (bGP)	GT-B	tetramer	(Mathieu et al. 2016)

GT35	CAA75517.1	1EM6[A,B], 1EXV[A,B] 1FA9[A], 1FC0[A,B] 1L5Q[A,B], 1L5R[A,B] 1L5S[A,B], 1L7X[A,B] 1XOI[A,B], 2ATI[A,B] 2QLL[A], 2ZB2[A,B] 3CEH[A,B], 3CEJ[A,B] 3CEM[A,B], 3DD1[A,B] 3DDS[A,B], 3DDW[A,B]	glycogen phosphorylase (liver)	GT-B	dimer	(Rath, Ammirati, Danley, et al. 2000; Rath, Ammirati, LeMotte, et al. 2000; Ekstrom et al. 2002; Wright et al. 2005; Klabunde et al. 2005; Anderka et al. 2008; Onda et al. 2008; Thomson et al. 2009)
GT35	AAC17451.1	1Z8D[A]	glycogen phosphorylase (muscle) (PigM)	GT-B	dimer	(Lukacs et al. 2006)
GT35	BAK00834.1	5LR8[A,B], 5LRA[A,B] 5LRB[A,B]	plastidial α-1,4-glucan phosphorylase (Pho1;HvPho1)	GT-B	dimer	(Cuesta-Seijo et al. 2017)
GT35	ACJ76617.1	2GJ4[A]	glycogen phosphorylase (PygM) (muscle)	GT-B	dimer	(Whittamore et al. 2006)
GT35	AAB68057.1	1YGP[A,B]	glycogen phosphorylase (Gph1;YPR160w)	GT-B	dimer	(K. Lin et al. 1996)
GT108	CBZ24449.1	6Q50[A]	MTP4 (LMXM_10_1260)	5-fold β-propeller	monomer	(Sernee et al. 2019)

3.4. Biotechnological applications

The different approaches for glycoside production include polysaccharide extraction from natural sources (*e.g.* microorganisms or plants) (Bhaumik and Dhepe 2015; Barcelos et al. 2020), hydrolysis and size-fractionation (Dashtban, Schraft, and Qin 2009), chemical synthesis (Overkleeft and Seeberger 2015), (chemo-)enzymatic synthesis (Overkleeft and Seeberger 2015; R. Chen 2018), and whole cell or cell-free synthesis (Anderson, Islam, and Prather 2018; R. Chen 2018). Due to the high diversity of glycosides already existing in Nature, extraction from the natural sources remains one of the most used method, especially at high scale. However, this often requires several purification steps, the use of harsh chemical or mechanical extraction methods, and when produced from plant sources it can compete with food and feed industries. This method is thus used for extraction of abundant glycosides, such as sucrose (for the food industry) or cellulose (paper industry, biofuels) (Bhaumik and Dhepe 2015; Raimo 2018).

Chemical synthesis is particularly interesting for the synthesis of structurally simple but also more complex carbohydrates. Thanks to its versatility, it is indeed possible to either introduce rare glycosyl units in glycosides or to assemble them in modular and complex ways. Recently, several automated systems were investigated (Blow 2009; Panza et al. 2018) and led to the synthesis of complex glycoside libraries (Eller et al. 2013; Walvoort et al. 2012). Even if they are being addressed (Nielsen and Pedersen 2018), major drawbacks arise from the use of these chemical syntheses, such as the tedious manipulation of protecting groups leading to multi-step pathways, the use of (often toxic) solvents and harsh temperature conditions, as well as the lack of reaction selectivity (Overkleeft and Seeberger 2015). While some carbohydrates still cannot be accessed through chemical synthesis or with low yields, it remains one of the most used method to produce specialty and high added value glycosides at the industrial scale.

Finally, enzyme-based approaches, sometimes in combination with chemical steps to take advantage of the best of both worlds, are more and more considered for the synthesis of glycosides (Wang et al. 2013; L. Li et al. 2015; Overkleeft and Seeberger 2015). Enzymes indeed offer exquisite regio- and stereo-specificities, while being able to work under mild pH and temperature conditions in aqueous media. Enzymatic glycoside synthesis now benefits from an increasing number of characterized biocatalysts in many CAZy families, and they can even be further engineered when lacking the requisite properties (McArthur and Chen 2016; Benkoulouche et al. 2019). Whole-cell biocatalysts also showed an increasing interest, especially for the *in vivo* coupling of carbohydrates to lipids or proteins allowing to produce complex glycoconjugates for the pharmaceutical industries (R. Chen 2018).

As explained above, GPs can be used for the synthesis of high-value oligosaccharides, polysaccharides and glycoconjugates, by reverse-phosphorolysis from glycosyl phosphates. In addition, the

phosphorolysis reaction can be used to regenerate glycosyl phosphates for multi-step enzymatic syntheses, for which a review was recently published (Pergolizzi et al. 2017). Several examples of glycoside structures accessed through enzymatic synthesis using GPs were previously reported and reviewed (Kitaoka 2015; O'Neill and Field 2015). Let's mention as example the synthesis reaction using the GH130 Uhgb_MP enzyme, which can use the relatively cheap *N*-acetyl-D-glucosamine and αMan1P substrates to produce the highly expensive β-1,4-mannopyranosyl-chitobiose, of major interest as the core part of human *N*-glycans (Ladevèze et al. 2013). Other mannan featuring β-1,4 linkages could be accessed thanks to GPs, such as crystalline linear β-1,4-mannan, a major component of plant cell walls hemicelluloses, which has been successfully produced using the GH130 β-1,4-mannooligosaccharide phosphorylase TM1225 (Grimaud et al. 2019). Similarly, sucrose-phosphorylases use a cheap and renewable donor substrate, to synthesize α-1,2 glucosylglycerol, a compound used in cosmetics (Goedl et al. 2010). As previously mentioned, the ability of GPs to generate sugar phosphates by phosphorolysis of low-cost substrates can advantageously be used in enzymatic cascades to regenerate the nucleotide sugar donors used by GTs. This was exemplified *in vitro*, with the kg-scale production of LNB using an elegant one-pot 4-enzymes system (M. Nishimoto and Kitaoka 2007; M. Nishimoto 2020), and also *in cellulo* for the production of phenolic glucosides (De Bruyn et al. 2015).

When using GPs in the reverse phosphorolysis direction for glycoside synthesis, an issue is the reversibility of the reaction that could lead to low production yield when reaching an equilibrium. A usual solution is to favor the reaction in the desired direction by removing the reverse phosphorolysis product, which can be easily achieved in biphasic systems when the product presents low solubility in the aqueous medium. For example, cellodextrin and cellobiose produced with the GH94 cellodextrin phosphorylase show a low solubility in the reaction medium and precipitated, so it could be easily recovered (Samain et al. 1995; Hiraishi et al. 2009). Another method to shift the reaction towards substrate consumption is to replace of glycosyl phosphates by glycosyl fluorides as donor substrates. The release of fluoride does not cause an attack of the glycoside, thus preventing phosphorolysis to happen and significantly increasing the production yields (Nakai et al. 2010). However, glycoside fluorides are more expensive than their phosphate counterpart, and just like fluoride donors for the glycosynthases mentioned above, they are often unstable. Finally, with continuous flow phosphorylation of glucose using immobilized sucrose phosphorylase (Britton, Majumdar, and Weiss 2018), as recently exemplified for GPs with cellobiose oligosaccharides synthesis by an immobilized multi-enzymes system (Zhong et al. 2020), or with one-pot multi-enzymatic (OPME) synthesis (Yu and Chen 2016; Pergolizzi et al. 2017).

Finally, when native enzymes do not meet the criteria to be considered for industrial scale production of carbohydrates, they can be modified using enzyme engineering. It is therefore possible to broaden the range of application of GPs, by altering specificity towards donors or acceptors, or even improve the enzyme (thermo)stability or resistance to solvents, making them more suited or compatible with certain industrial steps. Reviews illustrating several approaches and examples of GPs engineering were recently published (Desmet and Soetaert 2012; Benkoulouche et al. 2019; Franceus and Desmet 2020). For example, a lactose phosphorylase was engineered by mutating the acceptor binding site to extend the range of recognized acceptors (Manu R. M. De Groeve et al. 2009). Similarly, the modification of a cellobiose phosphorylase allowed altering the acceptor specificity, the engineered enzyme prefering methyl β-glucoside as an acceptor instead of the natural glucose acceptor (Groeve et al. 2010). The temperature of industrial-operating conditions are usually performed at around 60 °C to avoid microbial contamination (Waites et al. 2009), therefore several studies were focused on the discovery of GPs from thermophilic microorganisms (Grimaud et al. 2019). For instance, the sucrose phosphorylase from *Bifidobacteriom adolescentis* showed high-temperature tolerance properties compared to the other

characterized sucrose GPs, therefore it was a suitable template for thermostability engineering and understanding (Van den Broek et al. 2004).

The diversity of known GPs is unfortunately restricted to one or a few enzymes for each family, limiting the understanding of their structure-specificity-stability relationships, and thus, rational engineering studies. Efforts are needed to discover new GP candidates for industrial applications. The two next sections will therefore focus on the approaches that are used to characterize their activity and to discover them.

4. Methods for glycoside phosphorylase activity detection

Enzyme engineering and discovery from natural sources require the usage of suitable screening methodologies. The number of candidates to be tested can amount to hundreds of millions, especially when mining from metagenomic or combinatorial mutagenesis libraries. One recurrent limiting step is the screening of large libraries to identify the activity or property of interest.

The determination of GP activity can rely on chromogenic assays, chromatographic or spectroscopic methods.

4.1. Chromogenic assays

The inorganic phosphate released during the reverse-phosphorolysis reaction, and consumed during phosphorolysis, directly reflects GP activity, and is very convenient to screen it by chromogenic assays. The molybdenum blue activity assay was initially developed by Fiske and Subbarow, but not specifically for detection of GP activity (Fiske and Subbarow 1925). With this method, the measurement of phosphate consumption or release, depending on the reaction screened for, is based on its conversion to a phosphomolybdate complex in presence of the molybdate reactant. In acidic conditions, phosphomolybdate is reduced in molybdenum, of which the blue color intensity can be quantified by absorbance measurement at 655 nm (Figure 7). This method was further modified in several ways to adapt it to micro-plate format and automate it (Drueckes, Schinzel, and Palm 1995; Cogan, Birrell, and Griffith 1999; Gawronski and Benson 2004), and to make it compatible with the detection of GP-catalyzed reverse phosphorolysis activity in crude cellular extracts of *E. coli*, which cause high background and consumes glycosyl phosphates (M. R. M. De Groeve et al. 2010; Macdonald et al. 2019).

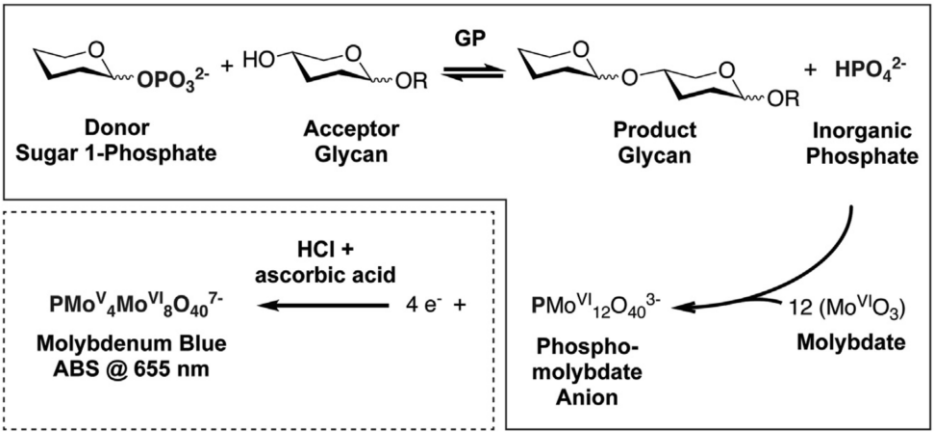


Figure 7: Synthetic scheme of the screening of glycoside phosphorylase activity using the molybdenum blue assay (from (Macdonald et al. 2019)

Phosphate concentration can also be measured with another chromogenic assay in liquid medium, using coupled enzyme systems. For instance, sucrose phosphorylase activity can be detected using a coupled assay with a phosphoglucomutase (which isomerizes G1P to G6P) and a glucose-6-phosphate dehydrogenase for the release of 6-phosphogluconate and NADH. A final step involves a 6-phosphogluconic dehydrogenase for the release of an additional NADH molecule as well as ribulose-5-phosphate (Tedokon et al. 1992).
Similarly, the glyceraldehyde-3-phosphate dehydrogenase uses glyceraldehyde-3-P, NAD^+ and inorganic phosphate to form 1,3-bisphosphoglycerate, releasing H^+ and NADH. The latter can then be detected at 340 nm (Serrano and Löffelhardt 1994). It could therefore be possible to couple this reaction to a reaction that involves the release of inorganic phosphate, the absorbance readout of NADH being directly related to the activity of the phosphorylase. Finally, a coupled enzymatic assay usable at pH 6 to 9 was recently developed and demonstrated in the context of DNA polymerase. It is based on a three-enzyme system: after the concerted action of an inorganic pyrophosphatase, a purine nucleoside phosphorylase and a xanthine oxidase, uric acid is generated and its absorbance is read at 293 nm and linked to the release of phosphate (Suárez et al. 2012). These two assays have yet never been used to detect GP activity.

More recently, another assay in liquid format was used for the screening of GP-catalyzed phosphorolysis reactions. It relies on the cleavage, in the presence of inorganic phosphate, of a 2,4-dinitrophenyl glycoside used as donor, releasing phenols that could be revealed by absorbance measurement at 400 nm. Even though it was proved to be compatible with detection of activity for some GPs (Macdonald et al. 2018), others can be inactive or weakly active (Ladevèze et al. 2013) on such chemically modified substrates which do not exactly mimic the natural glycoside targets.

4.2. Chromatographic and spectroscopic methods for carbohydrate detection and characterization

Thin layer chromatography (TLC) is a qualitative separation method that is widely applied for the separation of monosaccharides, oligosaccharides and sugar phosphates. It is easy to implement and perform, cheap and fast. The limitations are a relative low separation efficiency and lack of information when dealing with unknown spots (Qureshi et al. 2011).

High performance liquid chromatography (HPLC) can be used as a quantitative analysis methodology. It relies most of the time on a column-based separation, for example with copolymer resins in protonated

or metal ion forms that are suited for carbohydrate separation. While it offers good separation of carbohydrates (often after optimization of the elution gradient), one main roadblock is the carbohydrate detection. Commonly, carbohydrates can be detected using low-wavelength UV. However, due to the similar absorbance between charged sugars and organic mobile phase, UV detection requires derivatization to add suitable chemical groups that can be detected in a different wavelength (Harvey 2011). More common system for carbohydrate detection is based on the refractive index difference using refractive index detection (RI), although it often suffers from low sensitivity.

High-performance anion-exchange chromatography (HPAEC) offers a highly sensitive separation for mono-, oligo- and polysaccharides based on their structural features. For example, several DP2 or DP3 oligosaccharides differing only by their osidic linkage types can be separated by HPAEC. It is often coupled with pulsed amperometric detection (PAD) which is a highly specific and sensitive detection method (in the nanomole range) for carbohydrates (Corradini, Cavazza, and Bignardi 2012), HPAEC-PAD remains one of the most attractive and powerful methods for carbohydrate analysis. Nevertheless, as the response factor depends on the carbohydrate structures, it is impossible to determine the concentration of a product for which no pure standard is available. In addition, if HPAEC allows separating glycosyl phosphates from low DP glycosides and monosaccharides, the PAD response factor is very low for glycosyl phosphates, which makes difficult the simultaneous determination of monosaccharide, oligosaccharide and glycosyl phosphate concentrations.

The determination of oligosaccharide structures requires information about the anomers, linkages, and their component glycosyl units. Permethylation analysis was developed in the late 1960s and used to determine the position of glycosidic linkages (Hellerqvist et al. 1968). Typically, the carbohydrates are derivatized to form acid-stable methyl ethers, then hydrolyzed and converted to alditol acetates to be finally separated by gas chromatography and analyzed by mass spectrometry (MS) (Ciucanu 2006). The methylated alditol acetates are identified based on a combination of the relative retention times of the analytes and their fragmentation spectra.

Mass spectrometry (and more specifically, tandem MS) is a powerful strategy allowing the identification of ionized species according to their mass-to-charge (m/z) ratios (L. Han and Costello 2013), and it is widely used to determine glycoside structures.

Finally, NMR spectroscopy can provide the configuration of glycosidic linkages, by comparing of the chemical shifts with those of pure standard molecules. It is non-destructive, highly sensitive and allows an analysis of products in mixtures, even in crude cell extracts (Irague et al. 2011). NMR is extensively used to determine the structure of homooligosaccharides and homopolysaccharides synthesized by GPs (Ladevèze et al. 2013; Chiku et al. 2014; Grimaud et al. 2018).

Recently, new methods such as ion-mobility based separation arose, offering exquisite carbohydrate separation, even for isomers co-eluted by other separation techniques (Ropartz et al. 2019). Coupled with very sensitive and specific tandem MS methods, (complex) carbohydrate structure can now be completely determined (Ropartz et al. 2019; Tsai et al. 2019). Overall, great improvements in these methodologies allowed a faster separation, handling of more samples (with data acquisition in few minutes for one sample), while consuming less amounts of compounds (a few nmol) for the analysis. These methods, before being used in final screening steps for precise product characterization, now offer real possibilities for the screening of larger libraries with the generation of good quality data.

5. Enzyme discovery by functional (meta)genomics

As explained in section 3, the vast majority of the GPs discovered until now are issued from bacteria, which produce these enzymes for glycoside catabolism. More than 10^{30} bacterial and archaeal cells are present on Earth (Flemming and Wuertz 2019). Depending on the habitats, uncultured species make up more than 99.9 % of the ecosystem (Jo Handelsman 2004). For example, the human gut microbiota, which is one of the most studied ecosystem because of its impact on health, counts around 70 % of uncultured species (van de Guchte, Blottière, and Doré 2018). With the rapid development of next generation sequencing (NGS) technologies, the cost of sequencing significantly decreased in the past two decades, opening the way to the massive sequencing of metagenomes, the collective genomes of a microbiota (J. Handelsman et al. 1998). Most of these works are monogenic metagenomic studies, which correspond to the sequencing of 16S (bacteria, archea) and 18S (eukaryotes) rRNA genes to describe the diversity of microorganisms in an environmental sample without culturing them (Pace et al. 1986). Nevertheless, in the past decade, more and more functional metagenomics projects were launched. Functional metagenomics is the sequencing of functional genes to predict and/or to prove their function (Healy et al. 1995). These projects led to the establishment of several catalogs of genes from various microbiota, in particular those of human (J. Li et al. 2014; Tierney et al. 2019), mouse (Xiao et al. 2015; Lesker et al. 2020), pig (Xiao et al. 2016) and cow (J. Li et al. 2020) guts, totalizing nearly 34 millions genes and thousands metagenome-assembled genomes (MAGs). The drastic increase in available metagenomic sequences from several catalogs led to the discovery of hundreds novel enzymes and activities, through activity- or sequence-based analyses. In the following sections, I detail these approaches and their applications to the discovery of GPs.

5.1. Sampling strategies

Many microbial environments were sampled in the past two decades to search for novel enzymes, in particular CAZymes (Ferrer et al. 2005; Tasse et al. 2010). Enzymes that are naturally adapted to the conditions of industrial processes (heat, cold, acidic or basic conditions, presence of toxic compounds etc.), can be searched from extreme environments (Ufarté et al. 2015; Mirete, Morgante, and González-Pastor 2016). Enrichment steps before sampling can also be used to improve the relative abundance of micro-organisms producing the targeted enzymes, which resist to harsh conditions or are involved in the metabolization of a specific substrate. This can be achieved *in vitro*, by changing the physico-chemical conditions of the environment in enrichment reactors (van Elsas et al. 2008), or by adding the substrate to be metabolized in the reactors (DeAngelis et al. 2010), and even *in vivo*, by feeding animals with the targeted substrate (Bastien et al. 2013).

5.2. Activity-based approaches

The typical procedure of activity-based approaches starts with the cloning of metagenomic DNA (or cDNA for eukaryotes), after sampling their metatranscriptome (Bailly et al. 2007) in a vector (plasmids, fosmids or bacterial artificial chromosomes) to express the genes issued from the microbiome in a recombinant host. In activity-based metagenomic studies, the host is often *Escherichia coli*, due to its high transformation efficiency and its ability to efficiently express genes issued from distant taxa (Tasse et al. 2010). The metagenomic libraries have then to be screened with a suitable screening assay, preferably at a high throughput. Activity-based approaches raise issues, for example because of the GC bias in typical *E. coli* clone libraries, which could affect the DNA insert stability in the vector (Lam and Charles 2015), or the fact that the exogenous gene expression can have toxic effects on the cells.

To circumvent these problems, other host strains and vectors have been developed (Taupp, Mewis, and Hallam 2011; Lewin, Lale, and Wentzel 2017; Lam, Martens, and Charles 2018). In activity-based

metagenomics, the positive hit rate vastly depends on the source of the metagenomic DNA, its abundance in the metagenomic sample, the chosen cloning vector and expression host, which results in variable efficiency of heterologous gene expression, and the sensitivity and inherent bias of the screening assay. A typical hit rate ranges from 0.01 ‰ to 10 ‰ (Tauzin et al. 2017). This approach thus requires high-throughput screening facilities to screen a sufficiently large sequence space (typically 10^4 - 10^5 clones, which corresponds to 0.3 to 30.10^6 genes for fosmid libraries). Depending on the format of the assay, the screening throughput varies between 10^4 (automated liquid assays in micro-plates) to 10^5 assays per day (automated assays on solid plates) (Ufarté, Potocki-Veronese, and Laville 2015), and even 10^6 with droplet-microfluidics (Colin et al. 2015).

To date, only three activity-based functional metagenomic studies have been performed to discover GPs. In the first one, a new sucrose-phosphorylase (Genbank accession FJ472846, not listed in the CAZy database, but likely a GH13_18 member) was isolated from a 5,000 clones metagenomic library constructed from a sucrose refinery sample. In this case the screen was a positive selection assay on a solid mineral medium containing sucrose as sole-carbon source (Du et al. 2012). More recently, the group of S. Withers started to use metagenomics to mine various ecosystems from GPs. The first study was based on the screening of 17,168 clones library obtained from a passive mine tailings biochemical reactor system, on the activated substrate 2,4-dinitrophenyl β-D-glucoside (DNPGlc) and inorganic phosphate, yielding to the discovery of the retaining β-glucoside-phosphorylase BglP, one of the rare GPs from the GH3 family (Macdonald et al. 2018). The last study was focused on cellooligosaccharide-degrading GPs, searched in ~23,000 clones sourced from the same passive mine tailings biochemical reactor system as in the previous study, and from beaver feces. In this work, MacDonald and colleagues used a liquid assay based on molybdenum blue formation, and identified seven new GPs from the GH94 and GH149 families (cellobiose phosphorylases, cellodextrin phosphorylases, laminaribiose phosphorylases, and a β-1,3-glucan phosphorylase) (Macdonald et al. 2019). In addition, the authors demonstrated the versatility of the screening method, varying substrate combinations. This should yield to the discovery of other GPs from families GH13, GH65, GH112, and GH130 in the next future.

5.3. Sequence-based approaches

Contrary to activity-based approaches, sequence-based screening is performed using either *in silico* sequence homology-based methods, or *in vitro*, using DNA/DNA or DNA/RNA hydridization-based methods (such as PCR) to detect sequences with known consensus motifs. The latter, which have been only rarely used to discover CAZymes (Cottrell, Yu, and Kirchman 2005) are not well suited for the discovery of new enzyme activities and families, since the sequences that could be retrieved possess the same motifs involved in substrate recognition and catalysis as those of already known enzymes.

The *in silico* sequence homology-based methods are based on the search for sequences which are similar to a reference sequence of which the function has been experimentally verified, or which differs from it for some particular motifs. Two highly similar sequences have indeed good chances to be evolved from a common ancestor, and thus most likely share a similar function. These methods require alignment-search tools such as BLAST (Altschul et al. 1990) to classify proteins into evolutionary families. It is also possible that two protein sequences which do not share significant overall sequence similarities still have the same function owing to structurally conserved active site motifs. This is the case for members of different CAZy families which have the same activity. For instance, β-mannosidases can be found in five different families (GH1, GH2, GH5, GH130 and GH164), some being unrelated, having different folds $(\beta/\alpha)_8$ for the related families GH1, GH2 and GH5, five-bladed beta-propeller for GH130,

unknown for GH164). In other cases, the CAZy families can be only distantly related, sharing less than 20-30 % sequence identity, but high structural similarities, such as, for example, the GH130 and GT108 families which have a same five-bladed beta-propeller fold, and which both contain β-mannoside-phosphorylases (Sernee et al. 2019).

Several databases including COGs, Pfam and KEGG can be used for automated functional annotation of metagenomic sequences. COGs identify orthologous genes which typically share the same general function, for instance carbohydrate transport and metabolism for the G cluster, which includes CAZymes (Tatusov, Koonin, and Lipman 1997). Pfam contains nearly 18,000 protein families (Pfam 32.0, (El-Gebali et al. 2019). Each family is created based on multiple sequence alignments and hidden Markov models (HMMs) (Coggill, Finn, and Bateman 2008). Since Pfam families (like CAZy families) contain members with several activities, the activity of an enzyme cannot be deduced from the family it belongs to. Finally, KEGG can provide information on the involvement of a protein in the metabolism of an organism or an ecosystem (Coggill, Finn, and Bateman 2008). Nevertheless, orphan genes (which do not share significant sequence similarity with known sequences) cannot be annotated using sequence similarity-based methods. Furthermore, for the other sequences, there can be mistakes in the preliminary functional annotation using these databases, due (and participating to) to the amplification of incorrect functional annotations in databases. For example, when searching for the GH130 β-1,4-mannoside-phosphorylase Uhgb_MP (Genbank accession number ADD61463.1) in these databases, the results are the following:

- COG: the protein belongs to COG2152: Predicted glycosyl hydrolase, GH43/DUF377 family [Carbohydrate transport and metabolism]. The biochemical characterization of this enzyme (Ladevèze et al. 2013), indicated that these results are false, this enzyme being a mannoside-phosphorylase of the GH130 family, which does not contain a GH43 module.
- Pfam: the protein contains the domain *Glyco_hydro_130* (PF04041), described as 'beta-1,4-mannooligosaccharide phosphorylase'. This is true, but the same result is found for the the GH130 β-1,2-Mannosidase AAO78885.1 (Cuskin, Baslé, et al. 2015)
- Kegg: no result

The IMG-M and MG-RAST servers (Markowitz et al. 2008; Meyer et al. 2008) and the commercially available QIAGEN CLC Microbial Genomics Module are convenient tools to integrate the data resulting from the interrogation of these different databases, and make it possible to quantify and compare the abundance of the main functional families in the target ecosystems.

For more accurate function prediction, in particular regarding the substrate specificity of an enzyme or its mechanism, sequence analysis can be complemented with analysis of the genomic neighbourhood, as performed to discover the function of Uhgb_MP (for 'unknown human gut bacterium_mannoside-phosphorylase') the first GP issued from an uncultured bacterium (Ladevèze et al. 2013), thanks to the cloning of entire metagenomic PULs in fosmids (Tasse et al. 2010). More recently, this strategy was also used to elucidate the function of a novel sucrose-6^F-phosphate-phosphorylase from the cultured human gut bacterium *Ruminococcus gnavus* E1 (Tauzin et al. 2019). Nevertheless, such a strategy cannot be used with highly fragmented loci sequences, such as in datasets obtained by massive and random sequencing of metagenomes.

Mining unexplored clades of phylogenetic trees targeting GP-containing CAZy families or subfamilies is also an efficient strategy to discover new GP specificities, as shown for the GH13_18 subfamily (Franceus and Desmet 2020), which led to the discovery of glucosylglycerate phosphorylases (Franceus,

Pinel, and Desmet 2017; Franceus et al. 2018) and of a sucrose 6'-phosphate phosphorylases (Verhaeghe et al. 2014), and for the GH130 family, with the discovery of β-1,3- (Awad et al. 2017) and β-1,2-mannoside-phosphorylases (Chiku et al. 2014) from genomes of cultured strains.

Structure prediction and docking experiments can also be very useful (Lobb and Doxey 2016). Examination of the active site and of the conservation of the catalytic residues is indeed probably the best way to predict if a CAZyme is a GH or a GP. When I started my PhD, it had indeed previously been showed that inverting GH130 enzymes can either be GHs when they possess the two canonical carboxylic amino acids required to catalyze hydrolysis reactions, or GPs when they possess only the proton donor and phosphate-binding residues (Ladevèze et al. 2013; Cuskin, Baslé, et al. 2015). Regarding retaining enzymes, some GH3 members were shown to act as phosphorylases rather than hydrolases because they use a His/Asp dyad as an acid/base catalyst rather than the standard Glu or Asp residues. This allows an anionic phosphate to enter in the active site, which is greatly disfavored when the active site bears a deprotonated carboxylate base directly adjacent to the anomeric center of the glycosyl enzyme (Macdonald, Blaukopf, and Withers 2015). In addition, more recently, it was shown that GH84 retaining hydrolases can be converted to efficient phosphorylases by a single point mutation of the acid base, from Asp to Asn (Teze et al. 2020). However, in 2016, such in depth-sequence analysis or structure prediction had never been attempted for GP discovery from metagenomes, which requires specific expertise and facilities to deal with the comparison and alignment of thousands sequences.

To analyse such high numbers of sequences more easily than by using multiple sequence alignments and phyologenetic tree analyses, and to visualize sequence relationships in protein families, the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) was developed in 2015 (http://efi.igb.illinois.edu/efi-est/ (Gerlt et al. 2015). This web tool, dedicated to all members of the biological/biomedical community, including non-bioinformaticians, automatically generates sequence similarity networks (SSNs) (Atkinson et al. 2009), based on pairwise sequence comparison. In an SSN (Figure 8), each sequence is represented by a node (symbol) and is connected with an edge (line) to the nodes for all other sequence that share a sequence similarity greater than a user-specified value. This approach allows highlighting unexplored sequence spaces, guiding biochemists to discover novel functions. SSNs have been used to split a glycoyl radical enzyme superfamily into different iso-functional clusters, and to target the biochemical characterization efforts towards unexplored clusters of metagenomic sequences highly abundant in the human gut microbiome. This led to the discovery of a novel enzyme involved in the biosynthesis of L-proline, a key mediator of healthy microbiota-host symbioses (Levin et al. 2017). The SSNs-based approach had never been tested to analyze the diversity of CAZymes until very recently, when SSNs have been used to subdivide the multispecific GH16 family into 23 robust subfamilies (Viborg et al. 2019), and to unravel an subgroup of family AA5 containing an aryl alcohol oxidase with weak activity on carbohydrates (Mathieu et al. 2020). However, the CAZy database and the publically available metagenomic sequence datasets represent gold mines to explore with such approaches, to discover novel GPs.

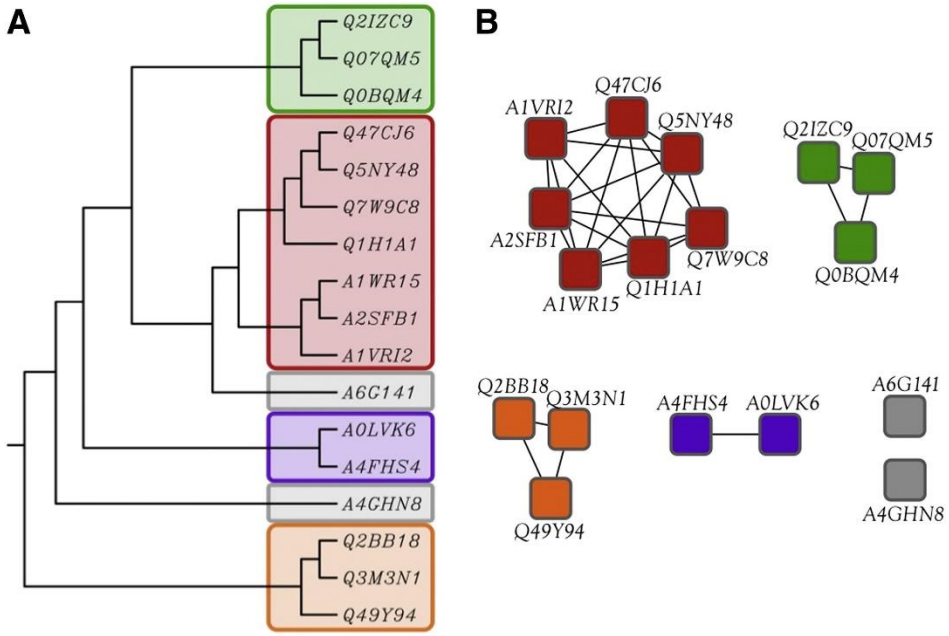


Figure 8: comparison of trees and sequence similarity networks. Panel A, a rooted phylogenetic tree created with ClustalW; panel B, the sequence similarity network using the same sequence set as shown in Panel A. (Gerlt et al. 2015)

Bibliography

Abbott, D. Wade, Eric C. Martens, Harry J. Gilbert, Fiona Cuskin, and Elisabeth C. Lowe. 2015. "Coevolution of Yeast Mannan Digestion: Convergence of the Civilized Human Diet, Distal Gut Microbiome, and Host Immunity." *Gut Microbes* 6 (5): 334–39. https://doi.org/10.1080/19490976.2015.1091913.

Albert, Martin, Werner Repetschnigg, Jörg Ortner, Joseph Gomes, Bernhard J. Paul, Carina Illaszewicz, Hansjörg Weber, Walter Steiner, and Karl Dax. 2000. "Simultaneous Detection of Different Glycosidase Activities by 19F NMR Spectroscopy." *Carbohydrate Research* 327 (4): 395–400. https://doi.org/10.1016/S0008-6215(00)00060-4.

Alexander, James K. 1961. "Characteristics of Cellobiose Phosphorylase." *Journal of Bacteriology* 81 (6): 903–10. https://doi.org/10.1128/JB.81.6.903-910.1961.

Allen, Howard J., and Edward C. Kisailus. 1992. *Glycoconjugates: Composition: Structure, and Function*. CRC Press.

Alonso-Casajús, Nora, David Dauvillée, Alejandro Miguel Viale, Francisco José Muñoz, Edurne Baroja-Fernández, María Teresa Morán-Zorzano, Gustavo Eydallin, Steven Ball, and Javier Pozueta-Romero. 2006. "Glycogen Phosphorylase, the Product of the GlgP Gene, Catalyzes Glycogen Breakdown by Removing Glucose Units from the Nonreducing Ends in Escherichia Coli." *Journal of Bacteriology* 188 (14): 5266–72. https://doi.org/10.1128/JB.01566-05.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anderka, Oliver, Petra Loenze, Thomas Klabunde, Matthias K. Dreyer, Elisabeth Defossa, K. Ulrich Wendt, and Dieter Schmoll. 2008. "Thermodynamic Characterization of Allosteric Glycogen Phosphorylase Inhibitors." *Biochemistry* 47 (16): 4683–91. https://doi.org/10.1021/bi702397d.

Andersen, Joakim Mark, Rodolphe Barrangou, Maher Abou Hachem, Sampo J. Lahtinen, Yong-Jun Goh, Birte Svensson, and Todd R. Klaenhammer. 2012. "Transcriptional Analysis of Prebiotic Uptake and Catabolism by Lactobacillus Acidophilus NCFM." *PLoS ONE* 7 (9). https://doi.org/10.1371/journal.pone.0044409.

Anderson, Lisa A., M. Ahsanul Islam, and Kristala L. J. Prather. 2018. "Synthetic Biology Strategies for Improving Microbial Synthesis of 'Green' Biopolymers." *Journal of Biological Chemistry* 293 (14): 5053–61. https://doi.org/10.1074/jbc.TM117.000368.

Andersson, Ulrika, Fredrik Levander, and Peter Rådström. 2001. "Trehalose-6-Phosphate Phosphorylase Is Part of a Novel Metabolic Pathway for Trehalose Utilization in Lactococcus Lactis." *Journal of Biological Chemistry* 276 (46): 42707–13. https://doi.org/10.1074/jbc.M108279200.

André, Isabelle, Gabrielle Potocki-Véronèse, Sophie Barbe, Claire Moulis, and Magali Remaud-Siméon. 2014. "CAZyme Discovery and Design for Sweet Dreams." *Current Opinion in Chemical Biology*, Biocatalysis and biotransformation Bioinorganic chemistry, 19 (April): 17–24. https://doi.org/10.1016/j.cbpa.2013.11.014.

Aspeborg, Henrik, Pedro M. Coutinho, Yang Wang, Harry Brumer, and Bernard Henrissat. 2012. "Evolution, Substrate Specificity and Subfamily Classification of Glycoside Hydrolase Family 5 (GH5)." *BMC Evolutionary Biology* 12 (1): 186. https://doi.org/10.1186/1471-2148-12-186.

Atkinson, Holly J., John H. Morris, Thomas E. Ferrin, and Patricia C. Babbitt. 2009. "Using Sequence Similarity Networks for Visualization of Relationships across Diverse Protein Superfamilies." *PLOS ONE* 4 (2): e4345. https://doi.org/10.1371/journal.pone.0004345.

Awad, Faisal Nureldin, Pedro Laborda, Meng Wang, Ai Min Lu, Qian Li, Zhi Peng Cai, Li Liu, and Josef Voglmeir. 2017. "Discovery and Biochemical Characterization of a Mannose Phosphorylase Catalyzing the Synthesis of Novel β-1,3-Mannosides." *Biochimica Et Biophysica Acta. General Subjects* 1861 (12): 3231–37. https://doi.org/10.1016/j.bbagen.2017.09.013.

Azad, Kalam Abul, and Mohammad Rasul. 2019. *Advanced Biofuels: Applications, Technologies and Environmental Sustainability*. Woodhead Publishing.

Bailly, Julie, Laurence Fraissinet-Tachet, Marie-Christine Verner, Jean-Claude Debaud, Marc Lemaire, Micheline Wésolowski-Louvel, and Roland Marmeisse. 2007. "Soil Eukaryotic Functional Diversity, a Metatranscriptomic Approach." *The ISME Journal* 1 (7): 632–42. https://doi.org/10.1038/ismej.2007.68.

Barcelos, Mayara C. S., Kele A. C. Vespermann, Franciele M. Pelissari, and Gustavo Molina. 2020. "Current Status of Biotechnological Production and Applications of Microbial Exopolysaccharides." *Critical Reviews in Food Science and Nutrition* 60 (9): 1475–95. https://doi.org/10.1080/10408398.2019.1575791.

Barrangou, Rodolphe, Eric Altermann, Robert Hutkins, Raul Cano, and Todd R. Klaenhammer. 2003. "Functional and Comparative Genomic Analyses of an Operon Involved in Fructooligosaccharide Utilization by Lactobacillus Acidophilus." *Proceedings of the National Academy of Sciences* 100 (15): 8957–62. https://doi.org/10.1073/pnas.1332765100.

Bastien, Géraldine, Grégory Arnal, Sophie Bozonnet, Sandrine Laguerre, Fernando Ferreira, Régis Fauré, Bernard Henrissat, et al. 2013. "Mining for Hemicellulases in the Fungus-Growing Termite Pseudacanthotermes Militaris Using Functional Metagenomics." *Biotechnology for Biofuels* 6 (May): 78. https://doi.org/10.1186/1754-6834-6-78.

Bell, Andrew, Jason Brunt, Emmanuelle Crost, Laura Vaux, Ridvan Nepravishta, C. David Owen, Dimitrios Latousakis, et al. 2019. "Elucidation of a Sialic Acid Metabolism Pathway in Mucus-Foraging Ruminococcus Gnavus Unravels Mechanisms of Bacterial Adaptation to the Gut." *Nature Microbiology* 4 (12): 2393–2404. https://doi.org/10.1038/s41564-019-0590-7.

Benkoulouche, Mounir, Régis Fauré, Magali Remaud-Siméon, Claire Moulis, and Isabelle André. 2019. "Harnessing Glycoenzyme Engineering for Synthesis of Bioactive Oligosaccharides." *Interface Focus* 9 (2): 20180069. https://doi.org/10.1098/rsfs.2018.0069.

Benson, Dennis A., Ilene Karsch-Mizrachi, Karen Clark, David J. Lipman, James Ostell, and Eric W. Sayers. 2012. "GenBank." *Nucleic Acids Research* 40 (Database issue): D48-53. https://doi.org/10.1093/nar/gkr1202.

Berman, Helen M., T. N. Bhat, Philip E. Bourne, Zukang Feng, Gary Gilliland, Helge Weissig, and John Westbrook. 2000. "The Protein Data Bank and the Challenge of Structural Genomics." *Nature Structural Biology* 7 (11): 957–59. https://doi.org/10.1038/80734.

Bhaumik, Prasenjit, and Paresh Laxmikant Dhepe. 2015. "Conversion of Biomass into Sugars." In *Biomass Sugars for Non-Fuel Applications*, 1–53. https://doi.org/10.1039/9781782622079-00001.

Bhuiyan, Shakhawat Hossain, Ahmed Abu Rus'd, Motomitsu Kitaoka, and Kiyoshi Hayashi. 2003. "Characterization of a Hyperthermostable Glycogen Phosphorylase from Aquifex Aeolicus Expressed in Escherichia Coli." *Journal of Molecular Catalysis B: Enzymatic* 22 (3): 173–80. https://doi.org/10.1016/S1381-1177(03)00029-8.

Bianchetti, Christopher M., Nathaniel L. Elsen, Brian G. Fox, and George N. Phillips. 2011. "Structure of Cellobiose Phosphorylase from Clostridium Thermocellum in Complex with Phosphate." *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* 67 (Pt 11): 1345–49. https://doi.org/10.1107/S1744309111032660.

Bibel, Miriam, Cornelia Brettl, Uwe Gosslar, Gernot Kriegshäuser, and Wolfgang Liebl. 1998. "Isolation and Analysis of Genes for Amylolytic Enzymes of the Hyperthermophilic Bacterium Thermotoga Maritima." *FEMS Microbiology Letters* 158 (1): 9–15. https://doi.org/10.1111/j.1574-6968.1998.tb12793.x.

Biely, Peter. 2012. "Microbial Carbohydrate Esterases Deacetylating Plant Polysaccharides." *Biotechnology Advances*, Special issue on ACB 2011, 30 (6): 1575–88. https://doi.org/10.1016/j.biotechadv.2012.04.010.

Bissaro, Bastien, Åsmund K. Røhr, Gerdt Müller, Piotr Chylenski, Morten Skaugen, Zarah Forsberg, Svein J. Horn, Gustav Vaaje-Kolstad, and Vincent G. H. Eijsink. 2017. "Oxidative Cleavage of Polysaccharides by Monocopper Enzymes Depends on H2O2." *Nature Chemical Biology* 13 (10): 1123–28. https://doi.org/10.1038/nchembio.2470.

Blackburn, Neil T., and Anthony J. Clarke. 2001. "Identification of Four Families of Peptidoglycan Lytic Transglycosylases." *Journal of Molecular Evolution* 52 (1): 78–84. https://doi.org/10.1007/s002390010136.

Blanco, Noelia, Ana B. Sanz, Jose M. Rodríguez-Peña, César Nombela, Vladimír Farkaš, Ramón Hurtado-Guerrero, and Javier Arroyo. 2015. "Structural and Functional Analysis of Yeast Crh1 and Crh2 Transglycosylases." *The FEBS Journal* 282 (4): 715–31. https://doi.org/10.1111/febs.13176.

Blow, Nathan. 2009. “Glycobiology: A Spoonful of Sugar.” *Nature* 457 (7229): 617–20. https://doi.org/10.1038/457617a.

Boraston, Alisdair B., David N. Bolam, Harry J. Gilbert, and Gideon J. Davies. 2004. “Carbohydrate-Binding Modules: Fine-Tuning Polysaccharide Recognition.” *Biochemical Journal* 382 (3): 769–81. https://doi.org/10.1042/BJ20040892.

Bourlieu, Claire, Thierry Astruc, Sophie Barbe, Jean-Guy Berrin, Estelle Bonnin, Rachel Boutrou, Virginie Hugouvieux, Steven Le Feunteun, and Gabriel Paës. 2020. “Enzymes to Unravel Bioproducts Architecture.” *Biotechnology Advances*, April, 107546. https://doi.org/10.1016/j.biotechadv.2020.107546.

Boutet, Emmanuel, Damien Lieberherr, Michael Tognolli, Michel Schneider, Parit Bansal, Alan J. Bridge, Sylvain Poux, Lydie Bougueleret, and Ioannis Xenarios. 2016. “UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View.” In *Plant Bioinformatics: Methods and Protocols*, edited by David Edwards, 23–54. Methods in Molecular Biology. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3167-5_2.

Briliūtė, Justina, Paulina A. Urbanowicz, Ana S. Luis, Arnaud Baslé, Neil Paterson, Osmond Rebello, Jenifer Hendel, et al. 2019. “Complex N-Glycan Breakdown by Gut Bacteroides Involves an Extensive Enzymatic Apparatus Encoded by Multiple Co-Regulated Genetic Loci.” *Nature Microbiology* 4 (9): 1571–81. https://doi.org/10.1038/s41564-019-0466-x.

Britton, Joshua, Sudipta Majumdar, and Gregory A. Weiss. 2018. “Continuous Flow Biocatalysis.” *Chemical Society Reviews* 47 (15): 5891–5918. https://doi.org/10.1039/C7CS00906B.

Buchner, Peter, Ljudmilla Borisjuk, and Ulrich Wobus. 1996. “Glucan Phosphorylases in Vicia Faba L.: Cloning, Structural Analysis and Expression Patterns of Cytosolic and Plastidic Forms in Relation to Starch.” *Planta* 199 (1): 64–73. https://doi.org/10.1007/BF00196882.

Burke, J., P. Hwang, L. Anderson, R. Lebo, F. Gorin, and R. Fletterick. 1987. “Intron/Exon Structure of the Human Gene for the Muscle Isozyme of Glycogen Phosphorylase.” *Proteins* 2 (3): 177–87. https://doi.org/10.1002/prot.340020303.

Cameron, Elizabeth A., Kurt J. Kwiatkowski, Byung-Hoo Lee, Bruce R. Hamaker, Nicole M. Koropatkin, and Eric C. Martens. 2014. “Multifunctional Nutrient-Binding Proteins Adapt Human Symbiotic Bacteria for Glycan Competition in the Gut by Separately Promoting Enhanced Sensing and Catalysis.” *MBio* 5 (5): e01441-01414. https://doi.org/10.1128/mBio.01441-14.

Campagnolo, Mara, Cristiana Campa, Rita De Zorzi, Jochen Wuerges, and Silvano Geremia. 2008. “X-Ray Studies on Ternary Complexes of Maltodextrin Phosphorylase.” *Archives of Biochemistry and Biophysics* 471 (1): 11–19. https://doi.org/10.1016/j.abb.2007.11.023.

Cate, James H. Doudna, Yong-Su Jin, Jonathan M. Galazka, and Suk-Jin Ha. 2014. Enhanced cellodextrin metabolism. United States US20140057323A1, filed February 7, 2012, and issued February 27, 2014. https://patents.google.com/patent/US20140057323A1/en.

Chaen, Hiroto, Tomoyuki Nishimoto, Takuo Yamamoto, Tetsuya Nakada, Shigeharu Fukuda, Toshiyuki Sugimoto, Masashi Kurimoto, and Yoshio Tsujisaka. 1999. “Formation of a Nonreducing Trisaccharide, Selaginose, from Trehalose by a Cell-Free System of Thermoanaerobium Brockii.” *Journal of Applied Glycoscience* 46 (2): 129–34. https://doi.org/10.5458/jag.46.129.

Chao, Chen, Soetaert Wim, and Desmet Tom. 2011. “Characterization of β-Galactoside Phosphorylases with Diverging Acceptor Specificities.” *Enzyme and Microbial Technology* 49 (1): 59–65. https://doi.org/10.1016/j.enzmictec.2011.03.010.

Chekan, Jonathan R., In Hyuk Kwon, Vinayak Agarwal, Dylan Dodd, Vanessa Revindran, Roderick I. Mackie, Isaac Cann, and Satish K. Nair. 2014. “Structural and Biochemical Basis for Mannan Utilization by Caldanaerobius Polysaccharolyticus Strain ATCC BAA-17.” *Journal of Biological Chemistry* 289 (50): 34965–77. https://doi.org/10.1074/jbc.M114.579904.

Chen, Rachel. 2018. “Enzyme and Microbial Technology for Synthesis of Bioactive Oligosaccharides: An Update.” *Applied Microbiology and Biotechnology* 102 (7): 3017–26. https://doi.org/10.1007/s00253-018-8839-2.

Chen, Shiqiong, Jingfang Liu, Huadong Pei, Jie Li, Jian Zhou, and Hua Xiang. 2007. “Molecular Investigation of a Novel Thermostable Glucan Phosphorylase from Thermoanaerobacter

Tengcongensis." *Enzyme and Microbial Technology* 41 (3): 390–96. https://doi.org/10.1016/j.enzmictec.2007.03.008.
Chiku, Kazuhiro, Takanori Nihira, Erika Suzuki, Mamoru Nishimoto, Motomitsu Kitaoka, Ken'ichi Ohtsubo, and Hiroyuki Nakai. 2014. "Discovery of Two β-1,2-Mannoside Phosphorylases Showing Different Chain-Length Specificities from Thermoanaerobacter Sp. X-514." *PloS One* 9 (12): e114882. https://doi.org/10.1371/journal.pone.0114882.
Ciucanu, Ionel. 2006. "Per-O-Methylation Reaction for Structural Analysis of Carbohydrates by Mass Spectrometry." *Analytica Chimica Acta* 576 (2): 147–55. https://doi.org/10.1016/j.aca.2006.06.009.
Cobucci-Ponzano, Beatrice, Fiorella Conte, Emiliano Bedini, Maria Michela Corsaro, Michelangelo Parrilli, Gerlind Sulzenbacher, Alexandra Lipski, et al. 2009. "Beta-Glycosyl Azides as Substrates for Alpha-Glycosynthases: Preparation of Efficient Alpha-L-Fucosynthases." *Chemistry & Biology* 16 (10): 1097–1108. https://doi.org/10.1016/j.chembiol.2009.09.013.
Cobucci-Ponzano, Beatrice, and Marco Moracci. 2012. "Glycosynthases as Tools for the Production of Glycan Analogs of Natural Products." *Natural Product Reports* 29 (6): 697–709. https://doi.org/10.1039/C2NP20032E.
Cockburn, Darrell W., and Nicole M. Koropatkin. 2016. "Polysaccharide Degradation by the Intestinal Microbiota and Its Influence on Human Health and Disease." *Journal of Molecular Biology* 428 (16): 3230–52. https://doi.org/10.1016/j.jmb.2016.06.021.
Cogan, Elizabeth B., G. Bruce Birrell, and O. Hayes Griffith. 1999. "A Robotics-Based Automated Assay for Inorganic and Organic Phosphates." *Analytical Biochemistry* 271 (1): 29–35. https://doi.org/10.1006/abio.1999.4100.
Coggill, Penny, Robert D. Finn, and Alex Bateman. 2008. "Identifying Protein Domains with the Pfam Database." *Current Protocols in Bioinformatics* 23 (1): 2.5.1-2.5.17. https://doi.org/10.1002/0471250953.bi0205s23.
Colin, Pierre-Yves, Balint Kintses, Fabrice Gielen, Charlotte M. Miton, Gerhard Fischer, Mark F. Mohamed, Marko Hyvönen, Diego P. Morgavi, Dick B. Janssen, and Florian Hollfelder. 2015. "Ultrahigh-Throughput Discovery of Promiscuous Enzymes by Picodroplet Functional Metagenomics." *Nature Communications* 6 (December): 10008. https://doi.org/10.1038/ncomms10008.
Cori, C. F., and G. T. Cori. 1936. "Mechanism of Formation of Hexosemonophosphate in Muscle and Isolation of a New Phosphate Ester." *Proceedings of the Society for Experimental Biology and Medicine* 34 (5): 702–5. https://doi.org/10.3181/00379727-34-8759P.
Cori, Gerty T., and Carl F. Cori. 1943. "Crystalline Muscle Phosphorylase Iv. Formation of Glycogen." *Journal of Biological Chemistry* 151 (1): 57–63.
Corradini, Claudio, Antonella Cavazza, and Chiara Bignardi. 2012. "High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection as a Powerful Tool to Evaluate Carbohydrates of Food Interest: Principles and Applications." Review Article. International Journal of Carbohydrate Chemistry. Hindawi. 2012. https://doi.org/10.1155/2012/487564.
Cottrell, Matthew T., Liying Yu, and David L. Kirchman. 2005. "Sequence and Expression Analyses of Cytophaga-like Hydrolases in a Western Arctic Metagenomic Library and the Sargasso Sea." *Applied and Environmental Microbiology* 71 (12): 8506–13. https://doi.org/10.1128/AEM.71.12.8506-8513.2005.
Coutinho, Pedro M., Emeline Deleury, Gideon J. Davies, and Bernard Henrissat. 2003. "An Evolving Hierarchical Family Classification for Glycosyltransferases." *Journal of Molecular Biology* 328 (2): 307–17. https://doi.org/10.1016/s0022-2836(03)00307-3.
Cuesta-Seijo, Jose A., Christian Ruzanski, Katarzyna Krucewicz, Sebastian Meier, Per Hägglund, Birte Svensson, and Monica M. Palcic. 2017. "Functional and Structural Characterization of Plastidic Starch Phosphorylase during Barley Endosperm Development." *PLoS ONE* 12 (4). https://doi.org/10.1371/journal.pone.0175488.
Cui, Steve W. 2005. *Food Carbohydrates: Chemistry, Physical Properties, and Applications*. CRC Press.
Cummings, J. H., and A. M. Stephen. 2007. "Carbohydrate Terminology and Classification." *European Journal of Clinical Nutrition* 61 (1): S5–18. https://doi.org/10.1038/sj.ejcn.1602936.

Cuskin, Fiona, Arnaud Baslé, Simon Ladevèze, Alison M. Day, Harry J. Gilbert, Gideon J. Davies, Gabrielle Potocki-Véronèse, and Elisabeth C. Lowe. 2015. "The GH130 Family of Mannoside Phosphorylases Contains Glycoside Hydrolases That Target β-1,2-Mannosidic Linkages in Candida Mannan." *Journal of Biological Chemistry* 290 (41): 25023–33. https://doi.org/10.1074/jbc.M115.681460.

Cuskin, Fiona, Elisabeth C. Lowe, Max J. Temple, Yanping Zhu, Elizabeth A. Cameron, Nicholas A. Pudlo, Nathan T. Porter, et al. 2015. "Human Gut Bacteroidetes Can Utilize Yeast Mannan through a Selfish Mechanism." *Nature* 517 (7533): 165–69. https://doi.org/10.1038/nature13995.

Dakhova, O. N., N. E. Kurepina, V. V. Zverlov, V. A. Svetlichnyi, and G. A. Velikodvorskaya. 1993. "Cloning and Expression in Escherichia Coli of Thermotoga Neapolitana Genes Coding for Enzymes of Carbohydrate Substrate Degradation." *Biochemical and Biophysical Research Communications* 194 (3): 1359–64. https://doi.org/10.1006/bbrc.1993.1974.

Dashtban, Mehdi, Heidi Schraft, and Wensheng Qin. 2009. "Fungal Bioconversion of Lignocellulosic Residues; Opportunities & Perspectives." *International Journal of Biological Sciences* 5 (6): 578–95. https://doi.org/10.7150/ijbs.5.578.

Daudé, David, Isabelle André, Pierre Monsan, and Magali Remaud-Siméon. 2014. "Successes in Engineering Glucansucrases to Enhance Glycodiversification." In *Carbohydrate Chemistry*, 624–45. https://pubs.rsc.org/en/content/chapter/bk9781849739658-00624/978-1-84973-965-8.

Daudé, David, Alizée Vergès, Emmanuelle Cambon, Stéphane Emond, Samuel Tranier, Isabelle André, and Magali Remaud-Siméon. 2019. "Neutral Genetic Drift-Based Engineering of a Sucrose-Utilizing Enzyme toward Glycodiversification." *ACS Catalysis* 9 (2): 1241–52. https://doi.org/10.1021/acscatal.8b03609.

Dauvillée, David, Vincent Chochois, Martin Steup, Sophie Haebel, Nora Eckermann, Gerhard Ritte, Jean-Philippe Ral, et al. 2006. "Plastidial Phosphorylase Is Required for Normal Starch Synthesis in Chlamydomonas Reinhardtii." *The Plant Journal: For Cell and Molecular Biology* 48 (2): 274–85. https://doi.org/10.1111/j.1365-313X.2006.02870.x.

De Bruyn, Frederik, Brecht De Paepe, Jo Maertens, Joeri Beauprez, Pieter De Cocker, Stein Mincke, Christian Stevens, and Marjan De Mey. 2015. "Development of an in Vivo Glucosylation Platform by Coupling Production to Growth: Production of Phenolic Glucosides by a Glycosyltransferase of Vitis Vinifera." *Biotechnology and Bioengineering* 112 (8): 1594–1603. https://doi.org/10.1002/bit.25570.

De Groeve, M. R. M., G. H. Tran, A. Van Hoorebeke, J. Stout, T. Desmet, S. N. Savvides, and W. Soetaert. 2010. "Development and Application of a Screening Assay for Glycoside Phosphorylases." *Analytical Biochemistry* 401 (1): 162–67. https://doi.org/10.1016/j.ab.2010.02.028.

De Groeve, Manu R. M., Miet De Baere, Lieve Hoflack, Tom Desmet, Erick J. Vandamme, and Wim Soetaert. 2009. "Creating Lactose Phosphorylase Enzymes by Directed Evolution of Cellobiose Phosphorylase." *Protein Engineering, Design & Selection: PEDS* 22 (7): 393–99. https://doi.org/10.1093/protein/gzp017.

DeAngelis, Kristen M., Whendee L. Silver, Andrew W. Thompson, and Mary K. Firestone. 2010. "Microbial Communities Acclimate to Recurring Changes in Soil Redox Potential Status." *Environmental Microbiology* 12 (12): 3137–49. https://doi.org/10.1111/j.1462-2920.2010.02286.x.

Déjean, Guillaume, Kazune Tamura, Adriana Cabrera, Namrata Jain, Nicholas A. Pudlo, Gabriel Pereira, Alexander Holm Viborg, Filip Van Petegem, Eric C. Martens, and Harry Brumer. 2020. "Synergy between Cell Surface Glycosidases and Glycan-Binding Proteins Dictates the Utilization of Specific Beta(1,3)-Glucans by Human Gut Bacteroides." *MBio* 11 (2). https://doi.org/10.1128/mBio.00095-20.

Delmer, Deborah P., and Yehudit Amor. 1995. "Cellulose Biosynthesis." *The Plant Cell* 7 (7): 987–1000. https://doi.org/10.1105/tpc.7.7.987.

Den Besten, Gijs, Karen Van Eunen, Albert K. Groen, Koen Venema, Dirk-Jan Reijngoud, and Barbara M. Bakker. 2013. "The Role of Short-Chain Fatty Acids in the Interplay between Diet, Gut Microbiota, and Host Energy Metabolism." *Journal of Lipid Research* 54 (9): 2325–40. https://doi.org/10.1194/jlr.R036012.

Derensy-Dron, Danielle, Frédéric Krzewinski, Colette Brassart, and Stéphane Bouquelet. 1999. “β-1,3-Galactosyl-N-Acetylhexosamine Phosphorylase from Bifidobacterium Bifidum DSM 20082: Characterization, Partial Purification and Relation to Mucin Degradation.” *Biotechnology and Applied Biochemistry* 29 (1): 3–10. https://doi.org/10.1111/j.1470-8744.1999.tb01142.x.

Desmet, Tom, and Wim Soetaert. 2012. “Broadening the Synthetic Potential of Disaccharide Phosphorylases through Enzyme Engineering.” *PROCESS BIOCHEMISTRY* 47 (1): 11–17. http://dx.doi.org/10.1016/j.procbio.2011.10.039.

Drueckes, P., R. Schinzel, and D. Palm. 1995. “Photometric Microtiter Assay of Inorganic Phosphate in the Presence of Acid-Labile Organic Phosphates.” *Analytical Biochemistry* 230 (1): 173–77. https://doi.org/10.1006/abio.1995.1453.

Du, Liqin, Hui Yang, Yunlong Huo, Hang Wei, Yuanjin Xu, Yutuo Wei, and Ribo Huang. 2012. “A Novel Sucrose Phosphorylase from the Metagenomes of Sucrose-Rich Environment: Isolation and Characterization.” *World Journal of Microbiology and Biotechnology* 28 (9): 2871–78. https://doi.org/10.1007/s11274-012-1098-y.

Duwenig, Elke, Martin Steup, and Jens Kossmann. 1997. “Induction of Genes Encoding Plastidic Phosphorylase from Spinach (Spinacia Oleracea L.) and Potato (Solanum Tuberosum L.) by Exogenously Supplied Carbohydrates in Excised Leaf Discs.” *Planta* 203 (1): 111–20. https://doi.org/10.1007/s00050171.

Egloff, Marie-Pierre, Jonas Uppenberg, Lutz Haalck, and Herman van Tilbeurgh. 2001. “Crystal Structure of Maltose Phosphorylase from Lactobacillus Brevis: Unexpected Evolutionary Relationship with Glucoamylases.” *Structure* 9 (8): 689–97. https://doi.org/10.1016/S0969-2126(01)00626-8.

Ehrmann, M. A., and R. F. Vogel. 1998. “Maltose Metabolism of Lactobacillus Sanfranciscensis: Cloning and Heterologous Expression of the Key Enzymes, Maltose Phosphorylase and Phosphoglucomutase.” *FEMS Microbiology Letters* 169 (1): 81–86. https://doi.org/10.1111/j.1574-6968.1998.tb13302.x.

Eklöf, Jens M., and Harry Brumer. 2010. “The XTH Gene Family: An Update on Enzyme Structure, Function, and Phylogeny in Xyloglucan Remodeling.” *Plant Physiology* 153 (2): 456–66. https://doi.org/10.1104/pp.110.156844.

Ekstrom, Jennifer L, Thomas A Pauly, Maynard D Carty, Walter C Soeller, Jeff Culp, Dennis E Danley, Dennis J Hoover, et al. 2002. “Structure-Activity Analysis of the Purine Binding Site of Human Liver Glycogen Phosphorylase.” *Chemistry & Biology* 9 (8): 915–24. https://doi.org/10.1016/S1074-5521(02)00186-2.

El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R. Eddy, Aurélien Luciani, Simon C. Potter, Matloob Qureshi, et al. 2019. “The Pfam Protein Families Database in 2019.” *Nucleic Acids Research* 47 (D1): D427–32. https://doi.org/10.1093/nar/gky995.

Eller, Steffen, Mayeul Collot, Jian Yin, Heung Sik Hahm, and Peter H. Seeberger. 2013. “Automated Solid-Phase Synthesis of Chondroitin Sulfate Glycosaminoglycans.” *Angewandte Chemie (International Ed. in English)* 52 (22): 5858–61. https://doi.org/10.1002/anie.201210132.

Elsas, Jan Dirk van, Rodrigo Costa, Janet Jansson, Sara Sjöling, Mark Bailey, Renaud Nalin, Timothy M. Vogel, and Leo van Overbeek. 2008. “The Metagenomics of Disease-Suppressive Soils - Experiences from the METACONTROL Project.” *Trends in Biotechnology* 26 (11): 591–601. https://doi.org/10.1016/j.tibtech.2008.07.004.

Embuscado, Milda E. 2014. *Functionalizing Carbohydrates for Food Applications: Texturizing and Bioactive/Flavor Delivery Systems*. DEStech Publications, Inc.

Ferjani, Ali, Laszlo Mustardy, Ronan Sulpice, Kay Marin, Iwane Suzuki, Martin Hagemann, and Norio Murata. 2003. “Glucosylglycerol, a Compatible Solute, Sustains Cell Division under Salt Stress.” *Plant Physiology* 131 (4): 1628–37. https://doi.org/10.1104/pp.102.017277.

Ferrer, Manuel, Olga V. Golyshina, Tatyana N. Chernikova, Amit N. Khachane, Vitor A. P. Martins Dos Santos, Michail M. Yakimov, Kenneth N. Timmis, and Peter N. Golyshin. 2005. “Microbial Enzymes Mined from the Urania Deep-Sea Hypersaline Anoxic Basin.” *Chemistry & Biology* 12 (8): 895–904. https://doi.org/10.1016/j.chembiol.2005.05.020.

Fiddes, Ian T., Joel Armstrong, Mark Diekhans, Stefanie Nachtweide, Zev N. Kronenberg, Jason G. Underwood, David Gordon, et al. 2018. “Comparative Annotation Toolkit (CAT)—

Simultaneous Clade and Personal Genome Annotation." *Genome Research* 28 (7): 1029–38. https://doi.org/10.1101/gr.233460.117.

Fierfort, Nicolas, and Eric Samain. 2008. "Genetic Engineering of Escherichia Coli for the Economical Production of Sialylated Oligosaccharides." *Journal of Biotechnology* 134 (3): 261–65. https://doi.org/10.1016/j.jbiotec.2008.02.010.

Fiske, Cyrus H., and Yellapragada Subbarow. 1925. "The Colorimetric Determination of Phosphorus." *Journal of Biological Chemistry* 66 (2): 375–400.

Flemming, Hans-Curt, and Stefan Wuertz. 2019. "Bacteria and Archaea on Earth and Their Abundance in Biofilms." *Nature Reviews Microbiology* 17 (4): 247–60. https://doi.org/10.1038/s41579-019-0158-9.

Foley, Matthew H., Darrell W. Cockburn, and Nicole M. Koropatkin. 2016. "The Sus Operon: A Model System for Starch Uptake by the Human Gut Bacteroidetes." *Cellular and Molecular Life Sciences* 73 (14): 2603–17. https://doi.org/10.1007/s00018-016-2242-x.

Franceus, Jorick, Nikolas Capra, Tom Desmet, and Andy-Mark W. H. Thunnissen. 2019. "Structural Comparison of a Promiscuous and a Highly Specific Sucrose 6F-Phosphate Phosphorylase." *International Journal of Molecular Sciences* 20 (16). https://doi.org/10.3390/ijms20163906.

Franceus, Jorick, Lena Decuyper, Matthias D'hooghe, and Tom Desmet. 2018. "Exploring the Sequence Diversity in Glycoside Hydrolase Family 13_18 Reveals a Novel Glucosylglycerol Phosphorylase." *Applied Microbiology and Biotechnology* 102 (7): 3183–91. https://doi.org/10.1007/s00253-018-8856-1.

Franceus, Jorick, and Tom Desmet. 2020. "Sucrose Phosphorylase and Related Enzymes in Glycoside Hydrolase Family 13: Discovery, Application and Engineering." *International Journal of Molecular Sciences* 21 (7): 2526. https://doi.org/10.3390/ijms21072526.

Franceus, Jorick, Denise Pinel, and Tom Desmet. 2017. "Glucosylglycerate Phosphorylase, an Enzyme with Novel Specificity Involved in Compatible Solute Metabolism." *Applied and Environmental Microbiology* 83 (19). https://doi.org/10.1128/AEM.01434-17.

Fushinobu, Shinya, Masafumi Hidaka, Andressa M. Hayashi, Takayoshi Wakagi, Hirofumi Shoun, and Motomitsu Kitaoka. 2011. "Interactions between Glycoside Hydrolase Family 94 Cellobiose Phosphorylase and Glucosidase Inhibitors." *Journal of Applied Glycoscience* 58 (3): 91–97. https://doi.org/10.5458/jag.jag.JAG-2010_022.

Garron, Marie-Line, and Bernard Henrissat. 2019. "The Continuing Expansion of CAZymes and Their Families." *Current Opinion in Chemical Biology* 53: 82–87. https://doi.org/10.1016/j.cbpa.2019.08.004.

Gawronski, Jeffrey D., and David R. Benson. 2004. "Microtiter Assay for Glutamine Synthetase Biosynthetic Activity Using Inorganic Phosphate Detection." *Analytical Biochemistry* 327 (1): 114–18. https://doi.org/10.1016/j.ab.2003.12.024.

Geremia, S., M. Campagnolo, R. Schinzel, and L. N. Johnson. 2002. "Enzymatic Catalysis in Crystals of Escherichia Coli Maltodextrin Phosphorylase." *Journal of Molecular Biology* 322 (2): 413–23. https://doi.org/10.1016/s0022-2836(02)00771-4.

Gerlt, John A., Jason T. Bouvier, Daniel B. Davidson, Heidi J. Imker, Boris Sadkhin, David R. Slater, and Katie L. Whalen. 2015. "Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1854 (8): 1019–37. https://doi.org/10.1016/j.bbapap.2015.04.015.

Gödl, Christiane, Thornthan Sawangwan, Bernd Nidetzky, and Mario Müller. 2009. Method For Producing 2-O-Glyceryl-Alpha-D-Glucopyranoside. United States US20090318372A1, filed September 21, 2007, and issued December 24, 2009. https://patents.google.com/patent/US20090318372A1/en.

Goedl, Christiane, Richard Griessler, Alexandra Schwarz, and Bernd Nidetzky. 2006. "Structure-Function Relationships for Schizophyllum Commune Trehalose Phosphorylase and Their Implications for the Catalytic Mechanism of Family GT-4 Glycosyltransferases." *The Biochemical Journal* 397 (3): 491–500. https://doi.org/10.1042/BJ20060029.

Goedl, Christiane, and Bernd Nidetzky. 2009. "Sucrose Phosphorylase Harbouring a Redesigned, Glycosyltransferase-Like Active Site Exhibits Retaining Glucosyl Transfer in the Absence of a

Covalent Intermediate." *ChemBioChem* 10 (14): 2333–37. https://doi.org/10.1002/cbic.200900429.

Goedl, Christiane, Thornthan Sawangwan, Patricia Wildberger, and Bernd Nidetzky. 2010. "Sucrose Phosphorylase: A Powerful Transglucosylation Catalyst for Synthesis of α-D-Glucosides as Industrial Fine Chemicals." *Biocatalysis and Biotransformation* 28 (1): 10–21. https://doi.org/10.3109/10242420903411595.

Griessler, Richard, Barbara Psik, Alexandra Schwarz, and Bernd Nidetzky. 2004. "Relationships between Structure, Function and Stability for Pyridoxal 5′-Phosphate-Dependent Starch Phosphorylase from Corynebacterium Callunae as Revealed by Reversible Cofactor Dissociation Studies." *European Journal of Biochemistry* 271 (16): 3319–29. https://doi.org/10.1111/j.1432-1033.2004.04265.x.

Grimaud, Florent, P. Faucard, L. Tarquis, S. Pizzut-Serin, P. Roblin, S. Morel, S. Le Gall, et al. 2018. "Enzymatic Synthesis of Polysaccharide-Based Copolymers." *Green Chemistry* 20 (17): 4012–22. https://doi.org/10.1039/C8GC01251B.

Grimaud, Florent, Sandra Pizzut-Serin, Laurence Tarquis, Simon Ladevèze, Sandrine Morel, Jean-Luc Putaux, and Gabrielle Potocki-Veronese. 2019. "In Vitro Synthesis and Crystallization of β-1,4-Mannan." *Biomacromolecules* 20 (2): 846–53. https://doi.org/10.1021/acs.biomac.8b01457.

Groeve, Manu R. M. De, Laurens Remmery, Annelies Van Hoorebeke, Jan Stout, Tom Desmet, Savvas N. Savvides, and Wim Soetaert. 2010. "Construction of Cellobiose Phosphorylase Variants with Broadened Acceptor Specificity towards Anomerically Substituted Glucosides." *Biotechnology and Bioengineering* 107 (3): 413–20. https://doi.org/10.1002/bit.22818.

Grondin, Julie M., Kazune Tamura, Guillaume Déjean, D. Wade Abbott, and Harry Brumer. 2017. "Polysaccharide Utilization Loci: Fueling Microbial Communities." *Journal of Bacteriology* 199 (15). https://doi.org/10.1128/JB.00860-16.

Guchte, Maarten van de, Hervé M. Blottière, and Joël Doré. 2018. "Humans as Holobionts: Implications for Prevention and Therapy." *Microbiome* 6 (1): 81. https://doi.org/10.1186/s40168-018-0466-8.

Guidolin, L. Soledad, Andrés E. Ciocchini, Nora Iñón de Iannino, and Rodolfo A. Ugalde. 2009. "Functional Mapping of Brucella Abortus Cyclic β-1,2-Glucan Synthase: Identification of the Protein Domain Required for Cyclization." *Journal of Bacteriology* 191 (4): 1230–38. https://doi.org/10.1128/JB.01108-08.

Ha, Suk-Jin, Jonathan M. Galazka, Eun Joong Oh, Vesna Kordić, Heejin Kim, Yong-Su Jin, and Jamie H. D. Cate. 2013. "Energetic Benefits and Rapid Cellobiose Fermentation by Saccharomyces Cerevisiae Expressing Cellobiose Phosphorylase and Mutant Cellodextrin Transporters." *Metabolic Engineering* 15 (January): 134–43. https://doi.org/10.1016/j.ymben.2012.11.005.

Hamaker, Bruce R., and Yunus E. Tuncil. 2014. "A Perspective on the Complexity of Dietary Fiber Structures and Their Potential Effect on the Gut Microbiota." *Journal of Molecular Biology* 426 (23): 3838–50. https://doi.org/10.1016/j.jmb.2014.07.028.

Han, Liang, and Catherine E. Costello. 2013. "Mass Spectrometry of Glycans." *Biochemistry. Biokhimiia* 78 (7): 710–20. https://doi.org/10.1134/S0006297913070031.

Han, Sang-Eun, Hawk-Bin Kwon, Seung-Bum Lee, Bu-Young Yi, Ikuo Murayama, Yutaka Kitamoto, and Myung-Ok Byun. 2003. "Cloning and Characterization of a Gene Encoding Trehalose Phosphorylase (TP) from Pleurotus Sajor-Caju." *Protein Expression and Purification* 30 (2): 194–202. https://doi.org/10.1016/s1046-5928(03)00104-9.

Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman. 1998. "Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products." *Chemistry & Biology* 5 (10): R245-249. https://doi.org/10.1016/s1074-5521(98)90108-9.

Handelsman, Jo. 2004. "Metagenomics: Application of Genomics to Uncultured Microorganisms." *Microbiology and Molecular Biology Reviews* 68 (4): 669–85. https://doi.org/10.1128/MMBR.68.4.669-685.2004.

Harvey, David J. 2011. "Derivatization of Carbohydrates for Analysis by Chromatography; Electrophoresis and Mass Spectrometry." *Journal of Chromatography B*, 879 (17): 1196–1225. https://doi.org/10.1016/j.jchromb.2010.11.010.

Hayes, Marc R., and Jörg Pietruszka. 2017. “Synthesis of Glycosides by Glycosynthases.” *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry* 22 (9). https://doi.org/10.3390/molecules22091434.

Healy, F. G., R. M. Ray, H. C. Aldrich, A. C. Wilkie, L. O. Ingram, and K. T. Shanmugam. 1995. “Direct Isolation of Functional Genes Encoding Cellulases from the Microbial Consortia in a Thermophilic, Anaerobic Digester Maintained on Lignocellulose.” *Applied Microbiology and Biotechnology* 43 (4): 667–74. https://doi.org/10.1007/bf00164771.

Hellerqvist, Carl Gustaf, Bengt Lindberg, Sigfrid Svensson, Tord Holme, and Alf A. Lindberg. 1968. “Structural Studies on the O-Specific Side-Chains of the Cell-Wall Lipopolysaccharide from Salmonella Typhimurium 395 Ms.” *Carbohydrate Research* 8 (1): 43–55. https://doi.org/10.1016/S0008-6215(00)81689-4.

Henrissat, B., and G. Davies. 1997. “Structural and Sequence-Based Classification of Glycoside Hydrolases.” *Current Opinion in Structural Biology* 7 (5): 637–44. https://doi.org/10.1016/s0959-440x(97)80072-3.

Hermann, J., K. Titani, L. H. Ericsson, R. D. Wade, H. Neurath, and K. A. Walsh. 1978. “Amino Acid Sequence of Two Cyanogen Bromide Fragments of Glycogen Phosphorylase.” *Biochemistry* 17 (26): 5672–79. https://doi.org/10.1021/bi00619a013.

Hidaka, Masafumi, Yuji Honda, Motomitsu Kitaoka, Satoru Nirasawa, Kiyoshi Hayashi, Takayoshi Wakagi, Hirofumi Shoun, and Shinya Fushinobu. 2004. “Chitobiose Phosphorylase from Vibrio Proteolyticus, a Member of Glycosyl Transferase Family 36, Has a Clan GH-L-like (Alpha/Alpha)(6) Barrel Fold.” *Structure (London, England: 1993)* 12 (6): 937–47. https://doi.org/10.1016/j.str.2004.03.027.

Hidaka, Masafumi, Motomitsu Kitaoka, Kiyoshi Hayashi, Takayoshi Wakagi, Hirofumi Shoun, and Shinya Fushinobu. 2006. “Structural Dissection of the Reaction Mechanism of Cellobiose Phosphorylase.” *The Biochemical Journal* 398 (1): 37–43. https://doi.org/10.1042/BJ20060274.

Hidaka, Masafumi, Mamoru Nishimoto, Motomitsu Kitaoka, Takayoshi Wakagi, Hirofumi Shoun, and Shinya Fushinobu. 2009. “The Crystal Structure of Galacto-N-Biose/Lacto-N-Biose I Phosphorylase A LARGE DEFORMATION OF A TIM BARREL SCAFFOLD.” *Journal of Biological Chemistry* 284 (11): 7273–83. https://doi.org/10.1074/jbc.M808525200.

Hidaka, Y., Y. Hatada, M. Akita, M. Yoshida, N. Nakamura, M. Takada, T. Nakakuki, S. Ito, and K. Horikoshi. 2005. “Maltose Phosphorylase from a Deep-Sea Paenibacillus Sp.: Enzymatic Properties and Nucleotide and Amino-Acid Sequences.” *Enzyme and Microbial Technology* 37 (2): 185–94. https://doi.org/10.1016/j.enzmictec.2005.02.010.

Higuchi, Yujiro, Yasunari Eshima, Yibo Huang, Takashi Kinoshita, Wataru Sumiyoshi, Shin-Ichi Nakakita, and Kaoru Takegawa. 2017. “Highly Efficient Transglycosylation of Sialo-Complex-Type Oligosaccharide Using Coprinopsis Cinerea Endoglycosidase and Sugar Oxazoline.” *Biotechnology Letters* 39 (1): 157–62. https://doi.org/10.1007/s10529-016-2230-0.

Hiraishi, Masao, Kiyohiko Igarashi, Satoshi Kimura, Masahisa Wada, Motomitsu Kitaoka, and Masahiro Samejima. 2009. “Synthesis of Highly Ordered Cellulose II in Vitro Using Cellodextrin Phosphorylase.” *Carbohydrate Research* 344 (18): 2468–73. https://doi.org/10.1016/j.carres.2009.10.002.

Höltje, J. V., D. Mirelman, N. Sharon, and U. Schwarz. 1975. “Novel Type of Murein Transglycosylase in Escherichia Coli.” *Journal of Bacteriology* 124 (3): 1067–76.

Honda, Yuji, and Motomitsu Kitaoka. 2006. “The First Glycosynthase Derived from an Inverting Glycoside Hydrolase.” *Journal of Biological Chemistry* 281 (3): 1426–31. https://doi.org/10.1074/jbc.M511202200.

Honda, Yuji, Motomitsu Kitaoka, and Kiyoshi Hayashi. 2004. “Reaction Mechanism of Chitobiose Phosphorylase from Vibrio Proteolyticus: Identification of Family 36 Glycosyltransferase in Vibrio.” *The Biochemical Journal* 377 (Pt 1): 225–32. https://doi.org/10.1042/BJ20031171.

Hudson, J. W., K. L. Hefferon, and M. M. Crerar. 1993. “Comparative Analysis of Species-Independent, Isozyme-Specific Amino-Acid Substitutions in Mammalian Muscle, Brain and Liver Glycogen Phosphorylases.” *Biochimica Et Biophysica Acta* 1164 (2): 197–208. https://doi.org/10.1016/0167-4838(93)90248-p.

Hüwel, Stephan, Lutz Haalck, Nikola Conrath, and Friedrich Spener. 1997. “Maltose Phosphorylase from Lactobacillus Brevis: Purification, Characterization, and Application in a Biosensor for

Ortho-Phosphate." *Enzyme and Microbial Technology* 21 (6): 413–20. https://doi.org/10.1016/S0141-0229(97)00014-8.

Hwang, P. K., and R. J. Fletterick. 1986. "Convergent and Divergent Evolution of Regulatory Sites in Eukaryotic Phosphorylases." *Nature* 324 (6092): 80–84. https://doi.org/10.1038/324080a0.

Igarashi, R. Y., and C. R. Meyer. 2000. "Cloning and Sequencing of Glycogen Metabolism Genes from Rhodobacter Sphaeroides 2.4.1. Expression and Characterization of Recombinant ADP-Glucose Pyrophosphorylase." *Archives of Biochemistry and Biophysics* 376 (1): 47–58. https://doi.org/10.1006/abbi.1999.1689.

Iñón de Iannino, Nora, Gabriel Briones, Marcelo Tolmasky, and Rodolfo A. Ugalde. 1998. "Molecular Cloning and Characterization of Cgs, the Brucella Abortus Cyclic β (1-2) Glucan Synthetase Gene: Genetic Complementation of Rhizobium Meliloti NdvB and Agrobacterium Tumefaciens ChvB Mutants." *Journal of Bacteriology* 180 (17): 4392–4400.

Inoue, Yasushi, Keiko Ishii, Tetsuji Tomita, Tsuneya Yatake, and Fumio Fukui. 2002. "Characterization of Trehalose Phosphorylase from *Bacillus Stearothermophilus* SK-1 and Nucleotide Sequence of the Corresponding Gene." *Bioscience, Biotechnology, and Biochemistry* 66 (9): 1835–43. https://doi.org/10.1271/bbb.66.1835.

Inoue, Yasushi, Nozomu Yasutake, Yoshie Oshima, Yoshie Yamamoto, Tetsuji Tomita, Shinsuke Miyoshi, and Tsuneya Yatake. 2002. "Cloning of the Maltose Phosphorylase Gene from Bacillus Sp. Strain RK-1 and Efficient Production of the Cloned Gene and the Trehalose Phosphorylase Gene from Bacillus Stearothermophilus SK-1 in Bacillus Subtilis." *Bioscience, Biotechnology, and Biochemistry* 66 (12): 2594–99. https://doi.org/10.1271/bbb.66.2594.

Irague, Romain, Stéphane Massou, Claire Moulis, Olivier Saurel, Alain Milon, Pierre Monsan, Magali Remaud-Siméon, Jean-Charles Portais, and Gabrielle Potocki-Véronèse. 2011. "NMR-Based Structural Glycomics for High-Throughput Screening of Carbohydrate-Active Enzyme Specificity." *Analytical Chemistry* 83 (4): 1202–6. https://doi.org/10.1021/ac1032148.

Jaito, Nongluck, Wataru Saburi, Rei Odaka, Yusuke Kido, Ken Hamura, Mamoru Nishimoto, Motomitsu Kitaoka, Hirokazu Matsui, and Haruhide Mori. 2014. "Characterization of a Thermophilic 4-O-β-d-Mannosyl-d-Glucose Phosphorylase from Rhodothermus Marinus." *Bioscience, Biotechnology, and Biochemistry* 78 (2): 263–70. https://doi.org/10.1080/09168451.2014.882760.

Jarrell, Ken F., Yan Ding, Benjamin H. Meyer, Sonja-Verena Albers, Lina Kaminski, and Jerry Eichler. 2014. "N-Linked Glycosylation in Archaea: A Structural, Functional, and Genetic Analysis." *Microbiology and Molecular Biology Reviews : MMBR* 78 (2): 304–41. https://doi.org/10.1128/MMBR.00052-13.

Jha, A. B, and R. S Dubey. 2004. "Carbohydrate Metabolism in Growing Rice Seedlings under Arsenic Toxicity." *Journal of Plant Physiology* 161 (7): 867–72. https://doi.org/10.1016/j.jplph.2004.01.004.

John, Franz J. St, Javier M. González, and Edwin Pozharski. 2010. "Consolidation of Glycosyl Hydrolase Family 30: A Dual Domain 4/7 Hydrolase Family Consisting of Two Structurally Distinct Groups." *FEBS Letters* 584 (21): 4435–41. https://doi.org/10.1016/j.febslet.2010.09.051.

Jones, T. H., and B. E. Wright. 1970. "Partial Purification and Characterization of Glycogen Phosphorylase from Dictyostelium Discoideum." *Journal of Bacteriology* 104 (2): 754–61.

Juers, Douglas H., Brian W. Matthews, and Reuben E. Huber. 2012. "LacZ β-Galactosidase: Structure and Function of an Enzyme of Historical and Molecular Biological Importance: LacZ β-Galactosidase." *Protein Science* 21 (12): 1792–1807. https://doi.org/10.1002/pro.2165.

Kaoutari, Abdessamad El, Fabrice Armougom, Jeffrey I. Gordon, Didier Raoult, and Bernard Henrissat. 2013. "The Abundance and Variety of Carbohydrate-Active Enzymes in the Human Gut Microbiota." *Nature Reviews Microbiology* 11 (7): 497–504. https://doi.org/10.1038/nrmicro3050.

Kaper, Thijs, Boguslawa Talik, Thijs J. Ettema, Herman Bos, Marc J. E. C. van der Maarel, and Lubbert Dijkhuizen. 2005. "Amylomaltase of Pyrobaculum Aerophilum IM2 Produces Thermoreversible Starch Gels." *Applied and Environmental Microbiology* 71 (9): 5098–5106. https://doi.org/10.1128/AEM.71.9.5098-5106.2005.

Karunatilaka, Krishanthi S., Elizabeth A. Cameron, Eric C. Martens, Nicole M. Koropatkin, and Julie S. Biteen. 2014. "Superresolution Imaging Captures Carbohydrate Utilization Dynamics in Human Gut Symbionts." *MBio* 5 (6): e02172. https://doi.org/10.1128/mBio.02172-14.

Kawaguchi, Takashi, Yasuo Ikeuchi, Noriko Tsutsumi, Akihiko Kan, Jun-Ichi Sumitani, and Motoo Arai. 1998. "Cloning, Nucleotide Sequence, and Expression of the Clostridium Thermocellum Cellodextrin Phosphorylase Gene and Its Application to Synthesis of Cellulase Inhibitors." *Journal of Fermentation and Bioengineering* 85 (2): 144–49. https://doi.org/10.1016/S0922-338X(97)86758-X.

Kawahara, Ryosuke, Wataru Saburi, Rei Odaka, Hidenori Taguchi, Shigeaki Ito, Haruhide Mori, and Hirokazu Matsui. 2012. "Metabolic Mechanism of Mannan in a Ruminal Bacterium, Ruminococcus Albus, Involving Two Mannoside Phosphorylases and Cellobiose 2-Epimerase: Discovery of a New Carbohydrate Phosphorylase, β-1,4-Mannooligosaccharide Phosphorylase." *Journal of Biological Chemistry* 287 (50): 42389–99. https://doi.org/10.1074/jbc.M112.390336.

Kawasaki, H., N. Nakamura, M. Ohmori, and T. Sakai. 1996. "Cloning and Expression in Escherichia Coli of Sucrose Phosphorylase Gene from Leuconostoc Mesenteroides No. 165." *Bioscience, Biotechnology, and Biochemistry* 60 (2): 322–24. https://doi.org/10.1271/bbb.60.322.

Keyhani, Nemat O., Xi-Bing Li, and Saul Roseman. 2000. "Chitin Catabolism in the Marine Bacterium Vibrio Furnissii Identification and Molecular Cloning of a Chitoporin." *Journal of Biological Chemistry* 275 (42): 33068–76. https://doi.org/10.1074/jbc.M001041200.

Kiel, J. A., J. M. Boels, G. Beldman, and G. Venema. 1994. "Glycogen in Bacillus Subtilis: Molecular Characterization of an Operon Encoding Enzymes Involved in Glycogen Biosynthesis and Degradation." *Molecular Microbiology* 11 (1): 203–18. https://doi.org/10.1111/j.1365-2958.1994.tb00301.x.

Kim, Mijin, Taeyeon Kwon, Hyong Joo Lee, Kyoung Heon Kim, Dae Kyun Chung, Geun Eog Ji, Eui-Seok Byeon, and Jong-Hoon Lee. 2003. "Cloning and Expression of Sucrose Phosphorylase Gene from Bifidobacterium Longum in E. Coli and Characterization of the Recombinant Enzyme." *Biotechnology Letters* 25 (15): 1211–17. https://doi.org/10.1023/A:1025035320983.

Kim, Se-Kwon, ed. 2014. *Marine Carbohydrates: Fundamentals and Applications, Part B*. Academic Press.

Kim, Sun-Ki, Michael E. Himmel, Yannick J. Bomble, and Janet Westpheling. 2018. "Expression of a Cellobiose Phosphorylase from Thermotoga Maritima in Caldicellulosiruptor Bescii Improves the Phosphorolytic Pathway and Results in a Dramatic Increase in Cellulolytic Activity." *Applied and Environmental Microbiology* 84 (3). https://doi.org/10.1128/AEM.02348-17.

Kim, Yeon-Kye, Motomitsu Kitaoka, Manem Krishnareddy, Yutaka Mori, and Kiyoshi Hayashi. 2002. "Kinetic Studies of a Recombinant Cellobiose Phosphorylase (CBP) of the Clostridium Thermocellum YM4 Strain Expressed in Escherichia Coli." *The Journal of Biochemistry* 132 (2): 197–203. https://doi.org/10.1093/oxfordjournals.jbchem.a003210.

Kiss, John Z., Aurea C. Vasconcelos, and Richard E. Triemer. 1987. "Structure of the Euglenoid Storage Carbohydrate, Paramylon." *American Journal of Botany* 74 (6): 877–82. https://doi.org/10.1002/j.1537-2197.1987.tb08691.x.

Kitamoto, Yutaka, Hajime Akashi, Hisashi Tanaka, and Nobuhiro Mori. 1988. "α-Glucose-1-Phosphate Formation by a Novel Trehalose Phosphorylase from Flammulina Velutipes." *FEMS Microbiology Letters* 55 (2): 147–50. https://doi.org/10.1111/j.1574-6968.1988.tb13924.x.

Kitaoka, Motomitsu. 2015. "Diversity of Phosphorylases in Glycoside Hydrolase Families." *Applied Microbiology and Biotechnology* 99 (20): 8377–90. https://doi.org/10.1007/s00253-015-6927-0.

Kitaoka, Motomitsu, Yasuyuki Matsuoka, Kiyotaka Mori, Mamoru Nishimoto, and Kiyoshi Hayashi. 2012. "Characterization of a Bacterial Laminaribiose Phosphorylase." *Bioscience, Biotechnology, and Biochemistry* 76 (2): 343–48. https://doi.org/10.1271/bbb.110772.

Kitaoka, Motomitsu, Jiesheng Tian, and Mamoru Nishimoto. 2005. "Novel Putative Galactose Operon Involving Lacto-N-Biose Phosphorylase in Bifidobacterium Longum." *Applied and Environmental Microbiology* 71 (6): 3158–62. https://doi.org/10.1128/AEM.71.6.3158-3162.2005.

Klabunde, Thomas, K. Ulrich Wendt, Dieter Kadereit, Volker Brachvogel, Hans-Jörg Burger, Andreas W. Herling, Nikos G. Oikonomakos, et al. 2005. "Acyl Ureas as Human Liver Glycogen

Phosphorylase Inhibitors for the Treatment of Type 2 Diabetes." *Journal of Medicinal Chemistry* 48 (20): 6178–93. https://doi.org/10.1021/jm049034y.

Koga, T., K. Nakamura, Y. Shirokane, K. Mizusawa, S. Kitao, and M. Kikuchi. 1991. "Purification and Some Properties of Sucrose Phosphorylase from Leuconostoc Mesenteroides." *Agricultural and Biological Chemistry* 55 (7): 1805–10.

Koide, A., K. Titani, L. H. Ericsson, S. Kumar, H. Neurath, and K. A. Walsh. 1978. "Sequence of the Amino-Terminal 349 Residues of Rabbit Muscle Glycogen Phosphorylase Including the Sites of Covalent and Allosteric Control." *Biochemistry* 17 (26): 5657–72. https://doi.org/10.1021/bi00619a012.

Koksharova, O. A., U. Brandt, and R. Cerff. 2004. "[The gap1 operon of the cyanobacterium Synechococcus PCC 7942 carries a gene encoding glycogen phosphorylase and is induced under anaerobic conditions]." *Mikrobiologiia* 73 (3): 388–92.

Koropatkin, Nicole M., Eric C. Martens, Jeffrey I. Gordon, and Thomas J. Smith. 2008. "Starch Catabolism by a Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices." *Structure (London, England: 1993)* 16 (7): 1105–15. https://doi.org/10.1016/j.str.2008.03.017.

Koshland, D. E. 1953. "Stereochemistry and the Mechanism of Enzymatic Reactions." *Biological Reviews* 28 (4): 416–36. https://doi.org/10.1111/j.1469-185X.1953.tb01386.x.

Koyama, Yoshiyuki, Masafumi Hidaka, Mamoru Nishimoto, and Motomitsu Kitaoka. 2013. "Directed Evolution to Enhance Thermostability of Galacto-N-Biose/Lacto-N-Biose I Phosphorylase." *Protein Engineering, Design and Selection* 26 (11): 755–61. https://doi.org/10.1093/protein/gzt049.

Kraus, Michael, Clemens Grimm, and Jürgen Seibel. 2016. "Redesign of the Active Site of Sucrose Phosphorylase through a Clash-Induced Cascade of Loop Shifts." *Chembiochem: A European Journal of Chemical Biology* 17 (1): 33–36. https://doi.org/10.1002/cbic.201500514.

———. 2017. "Switching Enzyme Specificity from Phosphate to Resveratrol Glucosylation." *Chemical Communications* 53 (90): 12181–84. https://doi.org/10.1039/C7CC05993K.

Kuhaudomlarp, Sakonwan, Nicola J. Patron, Bernard Henrissat, Martin Rejzek, Gerhard Saalbach, and Robert A. Field. 2018. "Identification of Euglena Gracilis β-1,3-Glucan Phosphorylase and Establishment of a New Glycoside Hydrolase (GH) Family GH149." *Journal of Biological Chemistry*, January, jbc.RA117.000936. https://doi.org/10.1074/jbc.RA117.000936.

Kuhaudomlarp, Sakonwan, Giulia Pergolizzi, Nicola J. Patron, Bernard Henrissat, and Robert A. Field. 2019. "Unraveling the Subtleties of β-(1→3)-Glucan Phosphorylase Specificity in the GH94, GH149, and GH161 Glycoside Hydrolase Families." *Journal of Biological Chemistry* 294 (16): 6483–93. https://doi.org/10.1074/jbc.RA119.007712.

Kuhaudomlarp, Sakonwan, Clare E. M. Stevenson, David M. Lawson, and Robert A. Field. 2019. "The Structure of a GH149 β-(1 → 3) Glucan Phosphorylase Reveals a New Surface Oligosaccharide Binding Site and Additional Domains That Are Absent in the Disaccharide-Specific GH94 Glucose-β-(1 → 3)-Glucose (Laminaribiose) Phosphorylase." *Proteins: Structure, Function, and Bioinformatics* 87 (10): 885–92. https://doi.org/10.1002/prot.25745.

Kuhaudomlarp, Sakonwan, Samuel Walpole, Clare E. M. Stevenson, Sergey A. Nepogodiev, David M. Lawson, Jesus Angulo, and Robert A. Field. 2019. "Unravelling the Specificity of Laminaribiose Phosphorylase from Paenibacillus Sp. YM-1 towards Donor Substrates Glucose/Mannose 1-Phosphate by Using X-Ray Crystallography and Saturation Transfer Difference NMR Spectroscopy." *Chembiochem: A European Journal of Chemical Biology* 20 (2): 181–92. https://doi.org/10.1002/cbic.201800260.

La Rosa, Sabina Leanti, Maria Louise Leth, Leszek Michalak, Morten Ejby Hansen, Nicholas A. Pudlo, Robert Glowacki, Gabriel Pereira, et al. 2019. "The Human Gut Firmicute Roseburia Intestinalis Is a Primary Degrader of Dietary β-Mannans." *Nature Communications* 10 (1): 905. https://doi.org/10.1038/s41467-019-08812-y.

Ladevèze, Simon. 2015. "Functional and Structural Insights into Glycoside Hydrolase Family 130 Enzymes : Implications in Carbohydrate Foraging by Human Gut Bacteria." Phdthesis, INSA de Toulouse. https://tel.archives-ouvertes.fr/tel-01309164.

Ladevèze, Simon, Gianluca Cioci, Pierre Roblin, Lionel Mourey, Samuel Tranier, and Gabrielle Potocki-Véronèse. 2015. "Structural Bases for N-Glycan Processing by Mannoside

Phosphorylase." *Acta Crystallographica Section D: Biological Crystallography* 71 (6): 1335–46. https://doi.org/10.1107/S1399004715006604.

Ladevèze, Simon, Laurence Tarquis, Davide A. Cecchini, Juliette Bercovici, Isabelle André, Christopher M. Topham, Sandrine Morel, et al. 2013. "Role of Glycoside Phosphorylases in Mannose Foraging by Human Gut Bacteria." *The Journal of Biological Chemistry* 288 (45): 32370–83. https://doi.org/10.1074/jbc.M113.483628.

Laine, R. A. 1994. "A Calculation of All Possible Oligosaccharide Isomers Both Branched and Linear Yields 1.05 x 10(12) Structures for a Reducing Hexasaccharide: The Isomer Barrier to Development of Single-Method Saccharide Sequencing or Synthesis Systems." *Glycobiology* 4 (6): 759–67. https://doi.org/10.1093/glycob/4.6.759.

Laine, Roger A. 2008. "The Information-Storing Potential of the Sugar Code." In *Glycosciences*, 1–14. John Wiley & Sons, Ltd. https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527614738.ch1.

Lairson, L.L., B. Henrissat, G.J. Davies, and S.G. Withers. 2008. "Glycosyltransferases: Structures, Functions, and Mechanisms." *Annual Review of Biochemistry* 77 (1): 521–55. https://doi.org/10.1146/annurev.biochem.76.061005.092322.

Lal, Manju A. 2018. "Metabolism of Storage Carbohydrates." In *Plant Physiology, Development and Metabolism*, edited by Satish C Bhatla and Manju A. Lal, 339–77. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-2023-1_9.

Lam, Kathy N., and Trevor C. Charles. 2015. "Strong Spurious Transcription Likely Contributes to DNA Insert Bias in Typical Metagenomic Clone Libraries." *Microbiome* 3: 22. https://doi.org/10.1186/s40168-015-0086-5.

Lam, Kathy N., Eric C. Martens, and Trevor C. Charles. 2018. "Developing a Bacteroides System for Function-Based Screening of DNA from the Human Gut Microbiome." *MSystems* 3 (3). https://doi.org/10.1128/mSystems.00195-17.

Lapébie, Pascal, Vincent Lombard, Elodie Drula, Nicolas Terrapon, and Bernard Henrissat. 2019. "Bacteroidetes Use Thousands of Enzyme Combinations to Break down Glycans." *Nature Communications* 10 (1): 1–7. https://doi.org/10.1038/s41467-019-10068-5.

Larsbrink, Johan, Atsushi Izumi, Glyn R. Hemsworth, Gideon J. Davies, and Harry Brumer. 2012. "Structural Enzymology of Cellvibrio Japonicus Agd31B Protein Reveals α-Transglucosylase Activity in Glycoside Hydrolase Family 31." *Journal of Biological Chemistry* 287 (52): 43288–99. https://doi.org/10.1074/jbc.M112.416511.

Larsbrink, Johan, Theresa E. Rogers, Glyn R. Hemsworth, Lauren S. McKee, Alexandra S. Tauzin, Oliver Spadiut, Stefan Klinter, et al. 2014. "A Discrete Genetic Locus Confers Xyloglucan Metabolism in Select Human Gut Bacteroidetes." *Nature* 506 (7489): 498–502. https://doi.org/10.1038/nature12907.

Lee, Han-Seung, Keith R. Shockley, Gerrit J. Schut, Shannon B. Conners, Clemente I. Montero, Matthew R. Johnson, Chung-Jung Chou, et al. 2006. "Transcriptional and Biochemical Analysis of Starch Metabolism in the Hyperthermophilic Archaeon Pyrococcus Furiosus." *Journal of Bacteriology* 188 (6): 2115–25. https://doi.org/10.1128/JB.188.6.2115-2125.2006.

Lee, Jin-Ha, Young-Hwan Moon, Nahyun Kim, Young-Min Kim, Hee-Kyoung Kang, Ji-Yeon Jung, Emad Abada, Seong-Soo Kang, and Doman Kim. 2008. "Cloning and Expression of the Sucrose Phosphorylase Gene from Leuconostoc Mesenteroides in Escherichia Coli." *Biotechnology Letters* 30 (4): 749–54. https://doi.org/10.1007/s10529-007-9608-y.

Lee, Jin-Ha, Seung-Heon Yoon, Seung-Hee Nam, Young-Hwan Moon, You-Youn Moon, and Doman Kim. 2006. "Molecular Cloning of a Gene Encoding the Sucrose Phosphorylase from Leuconostoc Mesenteroides B-1149 and the Expression in Escherichia Coli." *Enzyme and Microbial Technology* 39 (4): 612–20. https://doi.org/10.1016/j.enzmictec.2005.11.008.

Lee, Seung Seo, Sung You Hong, James C. Errey, Atsushi Izumi, Gideon J. Davies, and Benjamin G. Davis. 2011. "Mechanistic Evidence for a Front-Side, S N i-Type Reaction in a Retaining Glycosyltransferase." *Nature Chemical Biology* 7 (9): 631–38. https://doi.org/10.1038/nchembio.628.

Lee, Y. P. 1960. "Potato Phosphorylase. I. Purification, Physicochemical Properties and Catalytic Activity." *Biochimica Et Biophysica Acta* 43 (September): 18–24. https://doi.org/10.1016/0006-3002(60)90401-7.

Leemhuis, Hans, Ronan M. Kelly, and Lubbert Dijkhuizen. 2010. “Engineering of Cyclodextrin Glucanotransferases and the Impact for Biotechnological Applications.” *Applied Microbiology and Biotechnology* 85 (4): 823–35. https://doi.org/10.1007/s00253-009-2221-3.

Lesker, Till R., Abilash C. Durairaj, Eric J. C. Gálvez, Ilias Lagkouvardos, John F. Baines, Thomas Clavel, Alexander Sczyrba, Alice C. McHardy, and Till Strowig. 2020. “An Integrated Metagenome Catalog Reveals New Insights into the Murine Gut Microbiome.” *Cell Reports* 30 (9): 2909-2922.e6. https://doi.org/10.1016/j.celrep.2020.02.036.

Levasseur, Anthony, Elodie Drula, Vincent Lombard, Pedro M. Coutinho, and Bernard Henrissat. 2013. “Expansion of the Enzymatic Repertoire of the CAZy Database to Integrate Auxiliary Redox Enzymes.” *Biotechnology for Biofuels* 6 (1): 41. https://doi.org/10.1186/1754-6834-6-41.

Levin, B. J., Y. Y. Huang, S. C. Peck, Y. Wei, A. Martínez-del Campo, J. A. Marks, E. A. Franzosa, C. Huttenhower, and E. P. Balskus. 2017. “A Prominent Glycyl Radical Enzyme in Human Gut Microbiomes Metabolizes Trans-4-Hydroxy-l-Proline.” *Science* 355 (6325): eaai8386. https://doi.org/10.1126/science.aai8386.

Lewin, A., R. Lale, and A. Wentzel. 2017. “Expression Platforms for Functional Metagenomics: Emerging Technology Options Beyond Escherichia Coli.” In *Functional Metagenomics: Tools and Applications*, edited by Trevor C. Charles, Mark R. Liles, and Angela Sessitsch. Springer International Publishing. https://doi.org/10.1007/978-3-319-61510-3.

Ley, Ruth E., Peter J. Turnbaugh, Samuel Klein, and Jeffrey I. Gordon. 2006. “Human Gut Microbes Associated with Obesity.” *Nature* 444 (7122): 1022–23. https://doi.org/10.1038/4441022a.

Li, Junhua, Huijue Jia, Xianghang Cai, Huanzi Zhong, Qiang Feng, Shinichi Sunagawa, Manimozhiyan Arumugam, et al. 2014. “An Integrated Catalog of Reference Genes in the Human Gut Microbiome.” *Nature Biotechnology* 32 (8): 834–41. https://doi.org/10.1038/nbt.2942.

Li, Junhua, Huanzi Zhong, Yuliaxis Ramayo-Caldas, Nicolas Terrapon, Vincent Lombard, Gabrielle Potocki-Veronese, Jordi Estellé, et al. 2020. “A Catalog of Microbial Genes from the Bovine Rumen Unveils a Specialized and Diverse Biomass-Degrading Environment.” *GigaScience Database*, May. https://doi.org/10.1101/272690.

Li, Lei, Yunpeng Liu, Cheng Ma, Jingyao Qu, Angie D. Calderon, Baolin Wu, Na Wei, et al. 2015. “Efficient Chemoenzymatic Synthesis of an N-Glycan Isomer Library.” *Chemical Science* 6 (10): 5652–61. https://doi.org/10.1039/C5SC02025E.

Li, Wanqing, John B. McArthur, and Xi Chen. 2019. “Strategies for Chemoenzymatic Synthesis of Carbohydrates.” *Carbohydrate Research* 472 (January): 86–97. https://doi.org/10.1016/j.carres.2018.11.014.

Lin, C. T., K. W. Yeh, P. D. Lee, and J. C. Su. 1991. “Primary Structure of Sweet Potato Starch Phosphorylase Deduced from Its CDNA Sequence.” *Plant Physiology* 95 (4): 1250–53. https://doi.org/10.1104/pp.95.4.1250.

Lin, Kai, Virginia L. Rath, Shirleko C. Dai, Robert J. Fletterick, and Peter K. Hwang. 1996. “A Protein Phosphorylation Switch at the Conserved Allosteric Site in GP.” *Science* 273 (5281): 1539–41. https://doi.org/10.1126/science.273.5281.1539.

Liu, Aimin, Hidehiro Tomita, Hebiao Li, Hiroshi Miyaki, Chika Aoyagi, Satoshi Kaneko, and Kiyoshi Hayashi. 1998. “Cloning, Sequencing and Expression of the Cellobiose Phosphorylase Gene of Cellvibrio Gilvus.” *Journal of Fermentation and Bioengineering* 85 (5): 511–13. https://doi.org/10.1016/S0922-338X(98)80071-8.

Lobb, Briallen, and Andrew C. Doxey. 2016. “Novel Function Discovery through Sequence and Structural Data Mining.” *Current Opinion in Structural Biology* 38: 53–61. https://doi.org/10.1016/j.sbi.2016.05.017.

Lombard, Vincent, Thomas Bernard, Corinne Rancurel, Harry Brumer, Pedro M. Coutinho, and Bernard Henrissat. 2010. “A Hierarchical Classification of Polysaccharide Lyases for Glycogenomics.” *The Biochemical Journal* 432 (3): 437–44. https://doi.org/10.1042/BJ20101185.

Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. 2014. “The Carbohydrate-Active Enzymes Database (CAZy) in 2013.” *Nucleic Acids Research* 42 (D1): D490–95. https://doi.org/10.1093/nar/gkt1178.

Lu, Yan, Jon M. Steichen, Jian Yao, and Thomas D. Sharkey. 2006. “The Role of Cytosolic Alpha-Glucan Phosphorylase in Maltose Metabolism and the Comparison of Amylomaltase in

Arabidopsis and Escherichia Coli." *Plant Physiology* 142 (3): 878–89. https://doi.org/10.1104/pp.106.086850.

Lukacs, Christine M., Nikos G. Oikonomakos, Robert L. Crowther, Li-Na Hong, R. Ursula Kammlott, Wayne Levin, Shirley Li, et al. 2006. "The Crystal Structure of Human Muscle Glycogen Phosphorylase a with Bound Glucose and AMP: An Intermediate Conformation with T-State and R-State Features." *Proteins* 63 (4): 1123–26. https://doi.org/10.1002/prot.20939.

Macdonald, Spencer S., Zachary Armstrong, Connor Morgan-Lang, Magdalena Osowiecka, Kyle Robinson, Steven J. Hallam, and Stephen G. Withers. 2019. "Development and Application of a High-Throughput Functional Metagenomic Screen for Glycoside Phosphorylases." *Cell Chemical Biology* 26 (7): 1001-1012.e5. https://doi.org/10.1016/j.chembiol.2019.03.017.

Macdonald, Spencer S., Markus Blaukopf, and Stephen G. Withers. 2015. "N-Acetylglucosaminidases from CAZy Family GH3 Are Really Glycoside Phosphorylases, Thereby Explaining Their Use of Histidine as an Acid/Base Catalyst in Place of Glutamic Acid." *The Journal of Biological Chemistry* 290 (8): 4887–95. https://doi.org/10.1074/jbc.M114.621110.

Macdonald, Spencer S., Ankoor Patel, Veronica L. C. Larmour, Connor Morgan-Lang, Steven J. Hallam, Brian L. Mark, and Stephen G. Withers. 2018. "Structural and Mechanistic Analysis of a β-Glycoside Phosphorylase Identified by Screening a Metagenomic Library." *The Journal of Biological Chemistry* 293 (9): 3451–67. https://doi.org/10.1074/jbc.RA117.000948.

Mackenzie, Lloyd F., Qingping Wang, R. Antony J. Warren, and Stephen G. Withers. 1998. "Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis." *Journal of the American Chemical Society* 120 (22): 5583–84. https://doi.org/10.1021/ja980833d.

Malbert, Yannick, Claire Moulis, Yoann Brison, Sandrine Morel, Isabelle André, and Magali Remaud-Simeon. 2018. "Engineering a Branching Sucrase for Flavonoid Glucoside Diversification." *Scientific Reports* 8 (1): 1–12. https://doi.org/10.1038/s41598-018-33394-y.

Malet, C., and A. Planas. 1998. "From Beta-Glucanase to Beta-Glucansynthase: Glycosyl Transfer to Alpha-Glycosyl Fluorides Catalyzed by a Mutant Endoglucanase Lacking Its Catalytic Nucleophile." *FEBS Letters* 440 (1–2): 208–12. https://doi.org/10.1016/s0014-5793(98)01448-3.

Mann, Alexander J., Richard L. Hahnke, Sixing Huang, Johannes Werner, Peng Xing, Tristan Barbeyron, Bruno Huettel, et al. 2013. "The Genome of the Alga-Associated Marine Flavobacterium Formosa Agariphila KMM 3901T Reveals a Broad Potential for Degradation of Algal Polysaccharides." *Applied and Environmental Microbiology* 79 (21): 6813–22. https://doi.org/10.1128/AEM.01937-13.

Markowitz, Victor M., Natalia N. Ivanova, Ernest Szeto, Krishna Palaniappan, Ken Chu, Daniel Dalevi, I-Min A. Chen, et al. 2008. "IMG/M: A Data Management and Analysis System for Metagenomes." *Nucleic Acids Research* 36 (Database issue): D534–38. https://doi.org/10.1093/nar/gkm869.

Martens, Eric C., Amelia G. Kelly, Alexandra S. Tauzin, and Harry Brumer. 2014. "The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes." *Journal of Molecular Biology* 426 (23): 3851–65. https://doi.org/10.1016/j.jmb.2014.06.022.

Maruta, Kazuhiko, Kazuhisa Mukai, Hiroshi Yamashita, Michio Kubota, Hiroto Chaen, Shigeharu Fukuda, and Masashi Kurimoto. 2002. "Gene Encoding a Trehalose Phosphorylase from Thermoanaerobacter Brockii ATCC 35047." *Bioscience, Biotechnology, and Biochemistry* 66 (9): 1976–80. https://doi.org/10.1271/bbb.66.1976.

Mathieu, Cécile, Ines Li de la Sierra-Gallay, Romain Duval, Ximing Xu, Angélique Cocaign, Thibaut Léger, Gary Woffendin, et al. 2016. "Insights into Brain Glycogen Metabolism: THE STRUCTURE OF HUMAN BRAIN GLYCOGEN PHOSPHORYLASE." *The Journal of Biological Chemistry* 291 (35): 18072–83. https://doi.org/10.1074/jbc.M116.738898.

Mathieu, Yann, Wendy A. Offen, Stephanie M. Forget, Luisa Ciano, Alexander Holm Viborg, Elena Blagova, Bernard Henrissat, Paul H. Walton, Gideon J. Davies, and Harry Brumer. 2020. "Discovery of a Fungal Copper Radical Oxidase with High Catalytic Efficiency toward 5-Hydroxymethylfurfural and Benzyl Alcohols for Bioprocessing." *ACS Catalysis* 10 (5): 3042–58. https://doi.org/10.1021/acscatal.9b04727.

Matsuzawa, Tomohiko. 2019. “The Metagenome Approach: A New Resource for Glycosidases.” *Trends in Glycoscience and Glycotechnology* 31 (178): E15–20. https://doi.org/10.4052/tigg.1763.1E.

Mayer, Christoph, David J. Vocadlo, Melanie Mah, Karen Rupitz, Dominik Stoll, R. a. J. Warren, and Stephen G. Withers. 2006. “Characterization of a β-N-Acetylhexosaminidase and a β-N-Acetylglucosaminidase/β-Glucosidase from Cellulomonas Fimi.” *The FEBS Journal* 273 (13): 2929–41. https://doi.org/10.1111/j.1742-4658.2006.05308.x.

McArthur, John B., and Xi Chen. 2016. “Glycosyltransferase Engineering for Carbohydrate Synthesis.” *Biochemical Society Transactions* 44 (1): 129–42. https://doi.org/10.1042/BST20150200.

McInerney, Melissa, Geidy Serrano Rodriguez, Wojciech Pawlina, Christopher B. Hurt, Bradley S. Fletcher, Philip J. Laipis, and Susan C. Frost. 2002. “Glycogen Phosphorylase Is Activated in Response to Glucose Deprivation but Is Not Responsible for Enhanced Glucose Transport Activity in 3T3-L1 Adipocytes.” *Biochimica Et Biophysica Acta* 1570 (1): 53–62. https://doi.org/10.1016/s0304-4165(02)00154-x.

Meng, Xiangfeng, Joana Gangoiti, Yuxiang Bai, Tjaard Pijning, Sander S. Van Leeuwen, and Lubbert Dijkhuizen. 2016. “Structure–Function Relationships of Family GH70 Glucansucrase and 4,6-α-Glucanotransferase Enzymes, and Their Evolutionary Relationships with Family GH13 Enzymes.” *Cellular and Molecular Life Sciences* 73: 2681–2706. https://doi.org/10.1007/s00018-016-2245-7.

Mewis, Keith, Nicolas Lenfant, Vincent Lombard, and Bernard Henrissat. 2016. “Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: A Motivation for Detailed Enzyme Characterization.” *Applied and Environmental Microbiology* 82 (6): 1686–92. https://doi.org/10.1128/AEM.03453-15.

Meyer, F., D. Paarmann, M. D’Souza, R. Olson, EM Glass, M. Kubal, T. Paczian, et al. 2008. “The Metagenomics RAST Server – a Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes.” *BMC Bioinformatics* 9 (1): 386. https://doi.org/10.1186/1471-2105-9-386.

Mirete, Salvador, Verónica Morgante, and José Eduardo González-Pastor. 2016. “Functional Metagenomics of Extreme Environments.” *Current Opinion in Biotechnology* 38 (April): 143–49. https://doi.org/10.1016/j.copbio.2016.01.017.

Mirza, Osman, Lars K. Skov, Desiree Sprogøe, Lambertus A. M. van den Broek, Gerrit Beldman, Jette S. Kastrup, and Michael Gajhede. 2006. “Structural Rearrangements of Sucrose Phosphorylase from Bifidobacterium Adolescentis during Sucrose Conversion.” *Journal of Biological Chemistry* 281 (46): 35576–84. https://doi.org/10.1074/jbc.M605611200.

Mizanur, Rahman M., Amanda K. K. Griffin, and Nicola L. Pohl. 2008. “Recombinant Production and Biochemical Characterization of a Hyperthermostable Alpha-Glucan/Maltodextrin Phosphorylase from Pyrococcus Furiosus.” *Archaea (Vancouver, B.C.)* 2 (3): 169–76. https://doi.org/10.1155/2008/549759.

Monchois, Vincent, René-Marc Willemot, and Pierre Monsan. 1999. “Glucansucrases: Mechanism of Action and Structure–Function Relationships.” *FEMS Microbiology Reviews* 23 (2): 131–51. https://doi.org/10.1111/j.1574-6976.1999.tb00394.x.

Mori, H., K. Tanizawa, and T. Fukui. 1991. “Potato Tuber Type H Phosphorylase Isozyme. Molecular Cloning, Nucleotide Sequence, and Expression of a Full-Length CDNA in Escherichia Coli.” *The Journal of Biological Chemistry* 266 (28): 18446–53.

Moulis, Claire, Isabelle André, and Magali Remaud-Simeon. 2016. “GH13 Amylosucrases and GH70 Branching Sucrases, Atypical Enzymes in Their Respective Families.” *Cellular and Molecular Life Sciences: CMLS* 73 (14): 2661–79. https://doi.org/10.1007/s00018-016-2244-8.

Mukai, Kazuhisa, Kazuhiko Maruta, Kazuhiro Satouchi, Michio Kubota, Shigeharu Fukuda, Masashi Kurimoto, and Yoshio Tsujisaka. 2004. “Cyclic Tetrasaccharide-Synthesizing Enzymes from Arthrobacter Globiformis A19.” *Bioscience, Biotechnology, and Biochemistry* 68 (12): 2529–40. https://doi.org/10.1271/bbb.68.2529.

Mukherjee, Keya, Tamari Narindoshvili, and Frank M. Raushel. 2018. “Discovery of a Kojibiose Phosphorylase in Escherichia Coli K-12.” *Biochemistry* 57 (19): 2857–67. https://doi.org/10.1021/acs.biochem.8b00392.

Muzzarelli, Riccardo A. A. 2013. *Chitin*. Elsevier.

Nakae, Setsu, Shigeaki Ito, Mariko Higa, Takeshi Senoura, Jun Wasaki, Atsushi Hijikata, Masafumi Shionyu, Susumu Ito, and Tsuyoshi Shirai. 2013. “Structure of Novel Enzyme in Mannan Biodegradation Process 4-o-β-d-Mannosyl-d-Glucose Phosphorylase MGP.” *Journal of Molecular Biology* 425 (22): 4468–78. https://doi.org/10.1016/j.jmb.2013.08.002.

Nakai, Hiroyuki, Maher Abou Hachem, Bent O. Petersen, Yvonne Westphal, Karin Mannerstedt, Martin J. Baumann, Adiphol Dilokpimol, Henk A. Schols, Jens Ø Duus, and Birte Svensson. 2010. “Efficient Chemoenzymatic Oligosaccharide Synthesis by Reverse Phosphorolysis Using Cellobiose Phosphorylase and Cellodextrin Phosphorylase from Clostridium Thermocellum.” *Biochimie* 92 (12): 1818–26. https://doi.org/10.1016/j.biochi.2010.07.013.

Nakajima, Masahiro, and Motomitsu Kitaoka. 2008. “Identification of Lacto-N-Biose I Phosphorylase from Vibrio Vulnificus CMCP6.” *Applied and Environmental Microbiology* 74 (20): 6333–37. https://doi.org/10.1128/AEM.02846-07.

Nakajima, Masahiro, Takanori Nihira, Mamoru Nishimoto, and Motomitsu Kitaoka. 2008. “Identification of Galacto-N-Biose Phosphorylase from Clostridium Perfringens ATCC13124.” *Applied Microbiology and Biotechnology* 78 (3): 465–71. https://doi.org/10.1007/s00253-007-1319-8.

Nakajima, Masahiro, Mamoru Nishimoto, and Motomitsu Kitaoka. 2009a. “Characterization of Beta-1,3-Galactosyl-N-Acetylhexosamine Phosphorylase from Propionibacterium Acnes.” *Applied Microbiology and Biotechnology* 83 (1): 109–15. https://doi.org/10.1007/s00253-008-1838-y.

———. 2009b. “Characterization of Three β-Galactoside Phosphorylases from Clostridium Phytofermentans Discovery of d-Dalactosyl-β1→4-L-Rhamnose Phosphorylase.” *Journal of Biological Chemistry* 284 (29): 19220–27. https://doi.org/10.1074/jbc.M109.007666.

———. 2010. “Characterization of D-Galactosyl-β1→4-l-Rhamnose Phosphorylase from Opitutus Terrae.” *Enzyme and Microbial Technology* 46 (3): 315–19. https://doi.org/10.1016/j.enzmictec.2009.12.007.

Nakajima, Masahiro, Nobukiyo Tanaka, Nayuta Furukawa, Takanori Nihira, Yuki Kodutsumi, Yuta Takahashi, Naohisa Sugimoto, et al. 2017. “Mechanistic Insight into the Substrate Specificity of 1,2-β-Oligoglucan Phosphorylase from Lachnoclostridium Phytofermentans.” *Scientific Reports* 7: 42671. https://doi.org/10.1038/srep42671.

Nakajima, Masahiro, Hiroyuki Toyoizumi, Koichi Abe, Hiroyuki Nakai, Hayao Taguchi, and Motomitsu Kitaoka. 2014. “1,2-β-Oligoglucan Phosphorylase from Listeria Innocua.” *PloS One* 9 (3): e92353. https://doi.org/10.1371/journal.pone.0092353.

Nam, Young-Woo, Takanori Nihira, Takatoshi Arakawa, Yuka Saito, Motomitsu Kitaoka, Hiroyuki Nakai, and Shinya Fushinobu. 2015. “Crystal Structure and Substrate Recognition of Cellobionic Acid Phosphorylase, Which Plays a Key Role in Oxidative Cellulose Degradation by Microbes.” *The Journal of Biological Chemistry* 290 (30): 18281–92. https://doi.org/10.1074/jbc.M115.664664.

Ndeh, Didier, and Harry J Gilbert. 2018. “Biochemistry of Complex Glycan Depolymerisation by the Human Gut Microbiota.” *FEMS Microbiology Reviews* 42 (2): 146–64. https://doi.org/10.1093/femsre/fuy002.

Newgard, C. B., D. R. Littman, C. van Genderen, M. Smith, and R. J. Fletterick. 1988. “Human Brain Glycogen Phosphorylase. Cloning, Sequence Analysis, Chromosomal Mapping, Tissue Expression, and Comparison with the Human Liver and Muscle Isozymes.” *The Journal of Biological Chemistry* 263 (8): 3850–57.

Newgard, C. B., K. Nakano, P. K. Hwang, and R. J. Fletterick. 1986. “Sequence Analysis of the CDNA Encoding Human Liver Glycogen Phosphorylase Reveals Tissue-Specific Codon Usage.” *Proceedings of the National Academy of Sciences of the United States of America* 83 (21): 8132–36. https://doi.org/10.1073/pnas.83.21.8132.

Nidetzky, Bernd, Richard Griessler, Alexandra Schwarz, and Barbara Splechtna. 2004. “Cellobiose Phosphorylase from Cellulomonas Uda: Gene Cloning and Expression in Escherichia Coli, and Application of the Recombinant Enzyme in a ‘Glycosynthase-Type’ Reaction.” *Journal of Molecular Catalysis B: Enzymatic*, Proceedings of the 6th International Symposium on Biocatalysis and Biotransformations - BIOTRANS’03, 29 (1): 241–48. https://doi.org/10.1016/j.molcatb.2003.11.014.

Nielsen, Michael Martin, and Christian Marcus Pedersen. 2018. “Catalytic Glycosylations in Oligosaccharide Synthesis.” *Chemical Reviews* 118 (17): 8285–8358. https://doi.org/10.1021/acs.chemrev.8b00144.

Nihira, Takanori, Hiroyuki Nakai, Kazuhiro Chiku, and Motomitsu Kitaoka. 2012. “Discovery of Nigerose Phosphorylase from Clostridium Phytofermentans.” *Applied Microbiology and Biotechnology* 93 (4): 1513–22. https://doi.org/10.1007/s00253-011-3515-9.

Nihira, Takanori, Hiroyuki Nakai, and Motomitsu Kitaoka. 2012. “3-O-α-D-Glucopyranosyl-L-Rhamnose Phosphorylase from Clostridium Phytofermentans.” *Carbohydrate Research* 350 (March): 94–97. https://doi.org/10.1016/j.carres.2011.12.019.

Nihira, Takanori, Mamoru Nishimoto, Hiroyuki Nakai, Ken’ichi Ohtsubo, and Motomitsu Kitaoka. 2014. “Characterization of Two α-1,3-Glucoside Phosphorylases from Clostridium Phytofermentans.” *Journal of Applied Glycoscience* advpub. https://doi.org/10.5458/jag.jag.JAG-2013_013.

Nihira, Takanori, Yuka Saito, Motomitsu Kitaoka, Mamoru Nishimoto, Ken’ichi Otsubo, and Hiroyuki Nakai. 2012. “Characterization of a Laminaribiose Phosphorylase from Acholeplasma Laidlawii PG-8A and Production of 1,3-β-D-Glucosyl Disaccharides.” *Carbohydrate Research* 361 (November): 49–54. https://doi.org/10.1016/j.carres.2012.08.006.

Nihira, Takanori, Yuka Saito, Motomitsu Kitaoka, Ken’ichi Otsubo, and Hiroyuki Nakai. 2012. “Identification of Bacillus Selenitireducens MLS10 Maltose Phosphorylase Possessing Synthetic Ability for Branched α-d-Glucosyl Trisaccharides.” *Carbohydrate Research* 360 (October): 25–30. https://doi.org/10.1016/j.carres.2012.07.014.

Nihira, Takanori, Yuka Saito, Mamoru Nishimoto, Motomitsu Kitaoka, Kiyohiko Igarashi, Ken’ichi Ohtsubo, and Hiroyuki Nakai. 2013. “Discovery of Cellobionic Acid Phosphorylase in Cellulolytic Bacteria and Fungi.” *FEBS Letters* 587 (21): 3556–61. https://doi.org/10.1016/j.febslet.2013.09.014.

Nihira, Takanori, Yuka Saito, Ken’ichi Ohtsubo, Hiroyuki Nakai, and Motomitsu Kitaoka. 2014. “2-O-α-D-Glucosylglycerol Phosphorylase from Bacillus Selenitireducens MLS10 Possessing Hydrolytic Activity on β-D-Glucose 1-Phosphate.” *PloS One* 9 (1): e86548. https://doi.org/10.1371/journal.pone.0086548.

Nihira, Takanori, Erika Suzuki, Motomitsu Kitaoka, Mamoru Nishimoto, Ken’ichi Ohtsubo, and Hiroyuki Nakai. 2013. “Discovery of β-1,4-D-Mannosyl-N-Acetyl-D-Glucosamine Phosphorylase Involved in the Metabolism of N-Glycans.” *The Journal of Biological Chemistry* 288 (38): 27366–74. https://doi.org/10.1074/jbc.M113.469080.

Nishi, Aiko, Yasunori Nakamura, Naoki Tanaka, and Hikaru Satoh. 2001. “Biochemical and Genetic Analysis of the Effects OfAmylose-Extender Mutation in Rice Endosperm.” *Plant Physiology* 127 (2): 459–72. https://doi.org/10.1104/pp.010127.

Nishimoto, Mamoru. 2020. “Large Scale Production of Lacto-N-Biose I, a Building Block of Type I Human Milk Oligosaccharides, Using Sugar Phosphorylases.” *Bioscience, Biotechnology, and Biochemistry* 84 (1): 17–24. https://doi.org/10.1080/09168451.2019.1670047.

Nishimoto, Mamoru, and Motomitsu Kitaoka. 2007. “Practical Preparation of Lacto-N-Biose I, a Candidate for the Bifidus Factor in Human Milk.” *Bioscience, Biotechnology, and Biochemistry* 71 (8): 2101–4. https://doi.org/10.1271/bbb.70320.

Nishimoto, Tomoyuki, Tetsuya Nakada, Hiroto Chaen, Shigeharu Fukuda, Toshiyuki Sugimoto, Masashi Kurimoto, and Yoshio Tsujisaka. 1996. “Purification and Characterization of a Thermostable Trehalose Synthase from Thermus Aquaticus.” *Bioscience, Biotechnology, and Biochemistry* 60 (5): 835–39. https://doi.org/10.1271/bbb.60.835.

O. Sheridan, Paul, Jennifer C. Martin, Trevor D. Lawley, Hilary P. Browne, Hugh M. B. Harris, Annick Bernalier-Donadille, Sylvia H. Duncan, Paul W. O’Toole, Karen P. Scott, and Harry J. Flint. 2016. “Polysaccharide Utilization Loci and Nutritional Specialization in a Dominant Group of Butyrate-Producing Human Colonic Firmicutes.” *Microbial Genomics* 2 (2). https://doi.org/10.1099/mgen.0.000043.

Okada, Satoshi, Takuo Yamamoto, Hikaru Watanabe, Tomoyuki Nishimoto, Hiroto Chaen, Shigeharu Fukuda, Takayoshi Wakagi, and Shinya Fushinobu. 2014. “Structural and Mutational Analysis

of Substrate Recognition in Kojibiose Phosphorylase." *The FEBS Journal* 281 (3): 778–86. https://doi.org/10.1111/febs.12622.

Okuyama, Masayuki, Haruhide Mori, Kotomi Watanabe, Atsuo Kimura, and Seiya Chiba. 2002. "α-Glucosidase Mutant Catalyzes 'α-Glycosynthase'-Type Reaction." *Bioscience, Biotechnology, and Biochemistry* 66 (4): 928–33. https://doi.org/10.1271/bbb.66.928.

Onda, Kenichi, Takayuki Suzuki, Ryota Shiraki, Yasuhiro Yonetoku, Kenji Negoro, Kazuhiro Momose, Naoko Katayama, et al. 2008. "Synthesis of 5-Chloro-N-Aryl-1H-Indole-2-Carboxamide Derivatives as Inhibitors of Human Liver Glycogen Phosphorylase a." *Bioorganic & Medicinal Chemistry* 16 (10): 5452–64. https://doi.org/10.1016/j.bmc.2008.04.010.

O'Neill, Ellis C., and Robert A. Field. 2015. "Enzymatic Synthesis Using Glycoside Phosphorylases." *Carbohydrate Research* 403 (February): 23–37. https://doi.org/10.1016/j.carres.2014.06.010.

O'Neill, Ellis C., Giulia Pergolizzi, Clare E. M. Stevenson, David M. Lawson, Sergey A. Nepogodiev, and Robert A. Field. 2017. "Cellodextrin Phosphorylase from Ruminiclostridium Thermocellum: X-Ray Crystal Structure and Substrate Specificity Analysis." *Carbohydrate Research* 451 (November): 118–32. https://doi.org/10.1016/j.carres.2017.07.005.

O'Neill, Ellis, Abdul M. Rashid, Clare E. M. Stevenson, Anne-Claire Hetru, A. Patrick Gunning, Martin Rejzek, Sergey A. Nepogodiev, Stephen Bornemann, David M. Lawson, and Robert A. Field. 2014. "Sugar-Coated Sensor Chip and Nanoparticle Surfaces for the in Vitro Enzymatic Synthesis of Starch-like Materials." *Chemical Science* 5 (1): 341–50. https://doi.org/10.1039/C3SC51829A.

O'Reilly, M., K. A. Watson, and L. N. Johnson. 1999. "The Crystal Structure of the Escherichia Coli Maltodextrin Phosphorylase-Acarbose Complex." *Biochemistry* 38 (17): 5337–45. https://doi.org/10.1021/bi9828573.

O'Reilly, M., K. A. Watson, R. Schinzel, D. Palm, and L. N. Johnson. 1997. "Oligosaccharide Substrate Binding in Escherichia Coli Maltodextrin Phosphorylase." *Nature Structural Biology* 4 (5): 405–12. https://doi.org/10.1038/nsb0597-405.

Overkleeft, Hermen S., and Peter H. Seeberger. 2015. "Chemoenzymatic Synthesis of Glycans and Glycoconjugates." In *Essentials of Glycobiology*, edited by Ajit Varki, Richard D. Cummings, Jeffrey D. Esko, Pamela Stanley, Gerald W. Hart, Markus Aebi, Alan G. Darvill, et al., 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. http://www.ncbi.nlm.nih.gov/books/NBK453031/.

Pace, Norman R., David A. Stahl, David J. Lane, and Gary J. Olsen. 1986. "The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences." In *Advances in Microbial Ecology*, edited by K. C. Marshall, 1–55. Advances in Microbial Ecology. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4757-0611-6_1.

Palm, D., R. Goerl, and K. J. Burger. 1985. "Evolution of Catalytic and Regulatory Sites in Phosphorylases." *Nature* 313 (6002): 500–502. https://doi.org/10.1038/313500a0.

Panza, Matteo, Salvatore G. Pistorio, Keith J. Stine, and Alexei V. Demchenko. 2018. "Automated Chemical Oligosaccharide Synthesis: Novel Approach to Traditional Challenges." *Chemical Reviews* 118 (17): 8105–50. https://doi.org/10.1021/acs.chemrev.8b00051.

Park, Byung H., Tatiana V. Karpinets, Mustafa H. Syed, Michael R. Leuze, and Edward C. Uberbacher. 2010. "CAZymes Analysis Toolkit (CAT): Web Service for Searching and Analyzing Carbohydrate-Active Enzymes in a Newly Sequenced Organism Using CAZy Database." *Glycobiology* 20 (12): 1574–84. https://doi.org/10.1093/glycob/cwq106.

Park, Jae Kweon, Nemat O. Keyhani, and Saul Roseman. 2000. "Chitin Catabolism In The Marine Bacterium Vibrio Furnissii Identification, Molecular Cloning, And Characterization Of An,N′-Diacetylchitobiose Phosphorylase." *Journal of Biological Chemistry* 275 (42): 33077–83. https://doi.org/10.1074/jbc.M001042200.

Pergolizzi, Giulia, Sakonwan Kuhaudomlarp, Eeshan Kalita, and Robert A. Field. 2017. "Glycan Phosphorylases in Multi-Enzyme Synthetic Processes." *Protein and Peptide Letters* 24 (8): 696–709. https://doi.org/10.2174/0929866524666170811125109.

Persson, K., H. D. Ly, M. Dieckelmann, W. W. Wakarchuk, S. G. Withers, and N. C. Strynadka. 2001. "Crystal Structure of the Retaining Galactosyltransferase LgtC from Neisseria Meningitidis in Complex with Donor and Acceptor Sugar Analogs." *Nature Structural Biology* 8 (2): 166–75. https://doi.org/10.1038/84168.

Philips, Katherine B., Metin Kurtoglu, Howard J. Leung, Huaping Liu, Ningguo Gao, Mark A. Lehrman, Timothy G. Murray, and Theodore J. Lampidis. 2014. “Increased Sensitivity to Glucose Starvation Correlates with Downregulation of Glycogen Phosphorylase Isoform PYGB in Tumor Cell Lines Resistant to 2-Deoxy-D-Glucose.” *Cancer Chemotherapy and Pharmacology* 73 (2): 349–61. https://doi.org/10.1007/s00280-013-2358-8.

Plou, Francisco J., Aránzazu Gómez de Segura, and Antonio Ballesteros. 2007. “Application of Glycosidases and Transglycosidases in the Synthesis of Oligosaccharides.” In *Industrial Enzymes: Structure, Function and Applications*, edited by Julio Polaina and Andrew P. MacCabe, 141–57. Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-5377-0_9.

Puchart, Vladimír. 2015. “Glycoside Phosphorylases: Structure, Catalytic Properties and Biotechnological Potential.” *Biotechnology Advances* 33 (2): 261–76. https://doi.org/10.1016/j.biotechadv.2015.02.002.

Qin, Zhen, Shaoqing Yang, Liming Zhao, Xin You, Qiaojuan Yan, and Zhengqiang Jiang. 2017. “Catalytic Mechanism of a Novel Glycoside Hydrolase Family 16 ‘Elongating’ β-Transglycosylase.” *Journal of Biological Chemistry* 292 (5): 1666–78. https://doi.org/10.1074/jbc.M116.762419.

Qureshi, Muhammad Nasimullah, Guenther Stecher, Tahira Sultana, Gudrun Abel, Michael Popp, and Guenther K. Bonn. 2011. “Determination of Carbohydrates in Medicinal Plants-Comparison between TLC, Mf-MELDI-MS and GC-MS.” *Phytochemical Analysis* 22 (4): 296–302. https://doi.org/10.1002/pca.1280.

Raimo, Alén. 2018. *Carbohydrate Chemistry: Fundamentals and Applications*. World Scientific Publishing Company.

Rajan, Shyamala S., Xiaojing Yang, Frank Collart, Vivian L. Y. Yip, Stephen G. Withers, Annabelle Varrot, John Thompson, Gideon J. Davies, and Wayne F. Anderson. 2004. “Novel Catalytic Mechanism of Glycoside Hydrolysis Based on the Structure of an NAD+/Mn2+ -Dependent Phospho-Alpha-Glucosidase from Bacillus Subtilis.” *Structure (London, England: 1993)* 12 (9): 1619–29. https://doi.org/10.1016/j.str.2004.06.020.

Rakotoarivonina, Harivony, Pierre-Vincent Revol, Nathalie Aubry, and Caroline Rémond. 2016. “The Use of Thermostable Bacterial Hemicellulases Improves the Conversion of Lignocellulosic Biomass to Valuable Molecules.” *Applied Microbiology and Biotechnology* 100 (17): 7577–90. https://doi.org/10.1007/s00253-016-7562-0.

Rao, V. S. R. 2019. *Conformation of Carbohydrates*. CRC Press.

Rath, Virginia L, Mark Ammirati, Dennis E Danley, Jennifer L Ekstrom, E Michael Gibbs, Thomas R Hynes, Alan M Mathiowetz, et al. 2000. “Human Liver Glycogen Phosphorylase Inhibitors Bind at a New Allosteric Site.” *Chemistry & Biology* 7 (9): 677–82. https://doi.org/10.1016/S1074-5521(00)00004-1.

Rath, Virginia L., Mark Ammirati, Peter K. LeMotte, Kimberly F. Fennell, Mahmoud N. Mansour, Dennis E. Danley, Thomas R. Hynes, Gayle K. Schulte, David J. Wasilko, and Jayvardhan Pandit. 2000. “Activation of Human Liver Glycogen Phosphorylase by Alteration of the Secondary Structure and Packing of the Catalytic Core.” *Molecular Cell* 6 (1): 139–48. https://doi.org/10.1016/S1097-2765(05)00006-7.

Reichenbecher, Marisa, Friedrich Lottspeich, and Karin Bronnenmeier. 1997. “Purification and Properties of a Cellobiose Phosphorylase (CepA) and a Cellodextrin Phosphorylase (CepB) from the Cellulolytic Thermophile Clostridium Stercorarium.” *European Journal of Biochemistry* 247 (1): 262–67. https://doi.org/10.1111/j.1432-1033.1997.00262.x.

Reith, Jan, and Christoph Mayer. 2011. “Peptidoglycan Turnover and Recycling in Gram-Positive Bacteria.” *Applied Microbiology and Biotechnology* 92 (1): 1–11. https://doi.org/10.1007/s00253-011-3486-x.

Renzi, Francesco, Pablo Manfredi, Mélanie Dol, Jian Fu, Stéphane Vincent, and Guy Richard Cornelis. 2015. “Glycan-Foraging Systems Reveal the Adaptation of Capnocytophaga Canimorsus to the Dog Mouth.” *MBio* 6 (2): e02507. https://doi.org/10.1128/mBio.02507-14.

Roach, Peter. 2002. “Glycogen and Its Metabolism.” *Current Molecular Medicine* 2 (2): 101–20. https://doi.org/10.2174/1566524024605761.

Rogers, P. V., S. Luo, J. F. Sucic, and C. L. Rutherford. 1992. “Characterization and Cloning of Glycogen Phosphorylase 1 from Dictyostelium Discoideum.” *Biochimica Et Biophysica Acta* 1129 (3): 262–72. https://doi.org/10.1016/0167-4781(92)90502-q.

Ropartz, David, Mathieu Fanuel, Jakub Ujma, Martin Palmer, Kevin Giles, and Hélène Rogniaux. 2019. “Structure Determination of Large Isomeric Oligosaccharides of Natural Origin through Multipass and Multistage Cyclic Traveling-Wave Ion Mobility Mass Spectrometry.” *Analytical Chemistry* 91 (18): 12030–37. https://doi.org/10.1021/acs.analchem.9b03036.

Rudroff, Florian, Marko D. Mihovilovic, Harald Gröger, Radka Snajdrova, Hans Iding, and Uwe T. Bornscheuer. 2018. “Opportunities and Challenges for Combining Chemo- and Biocatalysis.” *Nature Catalysis* 1 (1): 12–22. https://doi.org/10.1038/s41929-017-0010-4.

Russell, R. R., H. Mukasa, A. Shimamura, and J. J. Ferretti. 1988. “Streptococcus Mutans GtfA Gene Specifies Sucrose Phosphorylase.” *Infection and Immunity* 56 (10): 2763–65.

Rybak, Konstantin Vyacheslavovich, Ekaterina Aleksandrovna Slivinskaya, Elvira Borisovna Voroshilova, and Yury Ivanovich Kozlov. 2008. Method for producing an l-amino acid using a bacterium of the enterobacteriaceae family having a pathway of glycogen biosynthesis disrupted. United States US7422880B2, filed January 17, 2006, and issued September 9, 2008. https://patents.google.com/patent/US7422880B2/en.

Saito, K., H. Yamazaki, Y. Ohnishi, S. Fujimoto, E. Takahashi, and S. Horinouchi. 1998. “Production of Trehalose Synthase from a Basidiomycete, Grifola Frondosa, in Escherichia Coli.” *Applied Microbiology and Biotechnology* 50 (2): 193–98. https://doi.org/10.1007/s002530051276.

Saito, Koki, Toshiya Kase, Eiichi Takahashi, Eisaku Takahashi, and Sueharu Horinouchi. 1998. “Purification and Characterization of a Trehalose Synthase from the Basidiomycete Grifola Frondosa.” *Applied and Environmental Microbiology* 64 (11): 4340–45.

Salamone, Stéphane, Catherine Guerreiro, Emmanuelle Cambon, Isabelle André, Magali Remaud-Siméon, and Laurence A. Mulard. 2015. “Programmed Chemo-Enzymatic Synthesis of the Oligosaccharide Component of a Carbohydrate-Based Antibacterial Vaccine Candidate.” *Chemical Communications* 51 (13): 2581–84. https://doi.org/10.1039/C4CC08805K.

Samain, Eric, Christine Lancelon-Pin, Frédéric Férigo, Vincent Moreau, Henri Chanzy, Alain Heyraud, and Hugues Driguez. 1995. “Phosphorolytic Synthesis of Cellodextrins.” *Carbohydrate Research* 271 (2): 217–26. https://doi.org/10.1016/0008-6215(95)00022-L.

Samain, Eric, and Bernard Priem. 2001. Procede de production d’oligosaccharides. World Intellectual Property Organization WO2001004341A1, filed July 7, 2000, and issued January 18, 2001. https://patents.google.com/patent/WO2001004341A1/fr.

Sato, Yutaka, Kazuko Okamoto-Shibayama, and Toshifumi Azuma. 2013. “The MalQ Gene Is Essential for Starch Metabolism in Streptococcus Mutans.” *Journal of Oral Microbiology* 5 (1): 21285. https://doi.org/10.3402/jom.v5i0.21285.

Sawano, Tatsuya, Wataru Saburi, Ken Hamura, Hirokazu Matsui, and Haruhide Mori. 2013. “Characterization of Ruminococcus Albus Cellodextrin Phosphorylase and Identification of a Key Phenylalanine Residue for Acceptor Specificity and Affinity to the Phosphate Group.” *The FEBS Journal* 280 (18): 4463–73. https://doi.org/10.1111/febs.12408.

Scheurwater, Edie, Chris W. Reid, and Anthony J. Clarke. 2008. “Lytic Transglycosylases: Bacterial Space-Making Autolysins.” *The International Journal of Biochemistry & Cell Biology* 40 (4): 586–91. https://doi.org/10.1016/j.biocel.2007.03.018.

Schiebel, K., E. Pekel, and D. Mayer. 1992. “The Nucleotide Sequence of Rat Liver Glycogen Phosphorylase CDNA.” *Biochimica Et Biophysica Acta* 1130 (3): 349–51. https://doi.org/10.1016/0167-4781(92)90453-7.

Schliselfeld, Louis H., and Moris J. Danon. 2002. “Inverse Relationship of Skeletal Muscle Glycogen from Wild-Type and Genetically Modified Mice to Their Phosphorylase a Activity.” *Biochemical and Biophysical Research Communications* 290 (2): 874–77. https://doi.org/10.1006/bbrc.2001.6292.

Schmölzer, Katharina, Melanie Weingarten, Kai Baldenius, and Bernd Nidetzky. 2019. “Glycosynthase Principle Transformed into Biocatalytic Process Technology: Lacto-N-Triose II Production with Engineered Exo-Hexosaminidase.” *ACS Catalysis* 9 (6): 5503–14. https://doi.org/10.1021/acscatal.9b01288.

Schupp, Nicole, and Paul Ziegler. 2004. “The Relation of Starch Phosphorylases to Starch Metabolism in Wheat.” *Plant and Cell Physiology* 45 (10): 1471–84. https://doi.org/10.1093/pcp/pch170.

Seeberger, Peter H., and Christoph Rademacher. 2014. *Carbohydrates as Drugs*. Springer.

Seibold, Gerd M., Martin Wurst, and Bernhard J. Eikmanns. 2009. “Roles of Maltodextrin and Glycogen Phosphorylases in Maltose Utilization and Glycogen Metabolism in Corynebacterium Glutamicum.” *Microbiology,* 155 (2): 347–58. https://doi.org/10.1099/mic.0.023614-0.

Senoura, Takeshi, Shigeaki Ito, Hidenori Taguchi, Mariko Higa, Shigeki Hamada, Hirokazu Matsui, Tadahiro Ozawa, et al. 2011. “New Microbial Mannan Catabolic Pathway That Involves a Novel Mannosylglucose Phosphorylase.” *Biochemical and Biophysical Research Communications* 408 (4): 701–6. https://doi.org/10.1016/j.bbrc.2011.04.095.

Sernee, M. Fleur, Julie E. Ralton, Tracy L. Nero, Lukasz F. Sobala, Joachim Kloehn, Marcel A. Vieira-Lara, Simon A. Cobbold, et al. 2019. “A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in Leishmania Parasites.” *Cell Host & Microbe* 26 (3): 385-399.e9. https://doi.org/10.1016/j.chom.2019.08.009.

Serrano, Aurelio, and Wolfgang Löffelhardt. 1994. “Identification of Two Different Glyceraldehyde-3-Phosphate Dehydrogenases (Phosphorylating) in the Photosynthetic Protist Cyanophora Paradoxa.” *Archives of Microbiology* 162 (1–2): 14–19. https://doi.org/10.1007/BF00264367.

Shemesh, Moshe., Avshalom. Tam, and Doron. Steinberg. 2007. “Differential Gene Expression Profiling of Streptococcus Mutans Cultured under Biofilm and Planktonic Conditions.” *Microbiology,* 153 (5): 1307–17. https://doi.org/10.1099/mic.0.2006/002030-0.

Skov, Lars K., Osman Mirza, Anette Henriksen, Gabrielle Potocki De Montalk, Magali Remaud-Simeon, Patricia Sarçabal, Rene-Marc Willemot, Pierre Monsan, and Michael Gajhede. 2001. “Amylosucrase, a Glucan-Synthesizing Enzyme from the α-Amylase Family.” *Journal of Biological Chemistry* 276 (27): 25273–78. https://doi.org/10.1074/jbc.M010998200.

Sonnewald, U., A. Basner, B. Greve, and M. Steup. 1995. “A Second L-Type Isozyme of Potato Glucan Phosphorylase: Cloning, Antisense Inhibition and Expression Analysis.” *Plant Molecular Biology* 27 (3): 567–76. https://doi.org/10.1007/bf00019322.

Sprogøe, Desiree, Lambertus A. M. van den Broek, Osman Mirza, Jette S. Kastrup, Alphons G. J. Voragen, Michael Gajhede, and Lars K. Skov. 2004. “Crystal Structure of Sucrose Phosphorylase from Bifidobacterium Adolescentis.” *Biochemistry* 43 (5): 1156–62. https://doi.org/10.1021/bi0356395.

Stam, Mark R., Etienne G. J. Danchin, Corinne Rancurel, Pedro M. Coutinho, and Bernard Henrissat. 2006. “Dividing the Large Glycoside Hydrolase Family 13 into Subfamilies: Towards Improved Functional Annotations of α-Amylase-Related Proteins.” *Protein Engineering, Design and Selection* 19 (12): 555–62. https://doi.org/10.1093/protein/gzl044.

Stick, Robert V., and Spencer Williams. 2010. *Carbohydrates: The Essential Molecules of Life*. Elsevier.

Suárez, Antonio S. Guillén, Alessandra Stefan, Silvia Lemma, Emanuele Conte, and Alejandro Hochkoeppler. 2012. “Continuous Enzyme-Coupled Assay of Phosphate- or Pyrophosphate-Releasing Enzymes.” *BioTechniques* 53 (2): 99–103. https://doi.org/10.2144/000113905.

Takaha, T. (Ezaki Glico Co Ltd, M. Yanase, H. Takata, and S. Okada. 2001. “Structure and Properties of Thermus Aquaticus Alpha-Glucan Phosphorylase Expressed in Escherichia Coli.” *Journal of Applied Glycoscience (Japan)*. http://agris.fao.org/agris-search/search.do?recordID=JP2001002957.

Takaha, T., M. Yanase, S. Okada, and S. M. Smith. 1993. “Disproportionating Enzyme (4-Alpha-Glucanotransferase; EC 2.4.1.25) of Potato. Purification, Molecular Cloning, and Potential Role in Starch Metabolism.” *The Journal of Biological Chemistry* 268 (2): 1391–96.

Takata, H., T. Takaha, S. Okada, M. Takagi, and T. Imanaka. 1997. “Characterization of a Gene Cluster for Glycogen Biosynthesis and a Heterotetrameric ADP-Glucose Pyrophosphorylase from Bacillus Stearothermophilus.” *Journal of Bacteriology* 179 (15): 4689–98. https://doi.org/10.1128/jb.179.15.4689-4698.1997.

Takeyasu, Kunio, Tsubasa Kawase, and Shige H. Yoshimura. 2003. “Intermolecular Interaction between Na+/K+-ATPase Alpha Subunit and Glycogen Phosphorylase.” *Annals of the New York Academy of Sciences* 986 (April): 522–24. https://doi.org/10.1111/j.1749-6632.2003.tb07239.x.

Tan, P., J. G. Allen, S. D. Wilton, P. A. Akkari, C. R. Huxtable, and N. G. Laing. 1997. "A Splice-Site Mutation Causing Ovine McArdle's Disease." *Neuromuscular Disorders: NMD* 7 (5): 336–42. https://doi.org/10.1016/s0960-8966(97)00062-x.

Tancula, E, M J Feldhaus, L A Bedzyk, and A A Salyers. 1992. "Location and Characterization of Genes Involved in Binding of Starch to the Surface of Bacteroides Thetaiotaomicron." *Journal of Bacteriology* 174 (17): 5609–16. https://doi.org/10.1128/JB.174.17.5609-5616.1992.

Tasse, Lena, Juliette Bercovici, Sandra Pizzut-Serin, Patrick Robe, Julien Tap, Christophe Klopp, Brandi L. Cantarel, et al. 2010. "Functional Metagenomics to Mine the Human Gut Microbiome for Dietary Fiber Catabolic Enzymes." *Genome Research* 20 (11): 1605–12. https://doi.org/10.1101/gr.108332.110.

Tatusov, Roman L., Eugene V. Koonin, and David J. Lipman. 1997. "A Genomic Perspective on Protein Families." *Science* 278 (5338): 631–37. https://doi.org/10.1126/science.278.5338.631.

Taupp, Marcus, Keith Mewis, and Steven J. Hallam. 2011. "The Art and Design of Functional Metagenomic Screens." *Current Opinion in Biotechnology* 22 (3): 465–72. https://doi.org/10.1016/j.copbio.2011.02.010.

Tauzin, Alexandra S., Laetitia Bruel, Elisabeth Laville, Cendrine Nicoletti, David Navarro, Bernard Henrissat, Josette Perrier, Gabrielle Potocki-Veronese, Thierry Giardina, and Mickael Lafond. 2019. "Sucrose 6F-Phosphate Phosphorylase: A Novel Insight in the Human Gut Microbiome." *Microbial Genomics*, 5 (4): e000253. https://doi.org/10.1099/mgen.0.000253.

Tauzin, Alexandra S., Elisabeth Laville, Davide Cecchini, Hervé M. Blottière, Marion Leclerc, Joël Doré, and Gabrielle Potocki-Veronese. 2017. "Human Gut Metagenomics: Success and Limits of the Activity-Based Approaches." In *Functional Metagenomics: Tools and Applications*, edited by Trevor C. Charles, Mark R. Liles, and Angela Sessitsch, 161–78. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-61510-3_10.

Tedokon, M., K. Suzuki, Y. Kayamori, S. Fujita, and Y. Katayama. 1992. "Enzymatic Assay of Inorganic Phosphate with Use of Sucrose Phosphorylase and Phosphoglucomutase." *Clinical Chemistry* 38 (4): 512–15.

Teixeira, Januana S., Reihaneh Abdi, Marcia Shu-Wei Su, Clarissa Schwab, and Michael G. Gänzle. 2013. "Functional Characterization of Sucrose Phosphorylase and ScrR, a Regulator of Sucrose Metabolism in Lactobacillus Reuteri." *Food Microbiology* 36 (2): 432–39. https://doi.org/10.1016/j.fm.2013.07.011.

Terada, Y., K. Fujii, T. Takaha, and S. Okada. 1999. "Thermus Aquaticus ATCC 33923 Amylomaltase Gene Cloning and Expression and Enzyme Characterization: Production of Cycloamylose." *Applied and Environmental Microbiology* 65 (3): 910–15.

Terrapon, Nicolas, Vincent Lombard, Élodie Drula, Pascal Lapébie, Saad Al-Masaudi, Harry J Gilbert, and Bernard Henrissat. 2018. "PULDB: The Expanded Database of Polysaccharide Utilization Loci." *Nucleic Acids Research* 46 (D1): D677–83. https://doi.org/10.1093/nar/gkx1022.

Terwisscha van Scheltinga, Anke C., Sylvie Armand, Kor H. Kalk, Akira Isogai, Bernard Henrissat, and Bauke W. Dijkstra. 1995. "Stereochemistry of Chitin Hydrolysis by a Plant Chitinase/Lysozyme and x-Ray Structure of a Complex with Allosamidin Evidence for Substrate Assisted Catalysis." *Biochemistry* 34 (48): 15619–23. https://doi.org/10.1021/bi00048a003.

Teze, David, Joan Coines, Lluís Raich, Valentina Kalichuk, Claude Solleux, Charles Tellier, Corinne André-Miral, Birte Svensson, and Carme Rovira. 2020. "A Single Point Mutation Converts GH84 *O* -GlcNAc Hydrolases into Phosphorylases: Experimental and Theoretical Evidence." *Journal of the American Chemical Society* 142 (5): 2120–24. https://doi.org/10.1021/jacs.9b09655.

Teze, David, Johann Hendrickx, Mirjam Czjzek, David Ropartz, Yves-Henri Sanejouand, Vinh Tran, Charles Tellier, and Michel Dion. 2014. "Semi-Rational Approach for Converting a GH1 β-Glycosidase into a β-Transglycosidase." *Protein Engineering, Design and Selection* 27 (1): 13–19. https://doi.org/10.1093/protein/gzt057.

Thomson, Stephen A., Pierette Banker, D. Mark Bickett, Joyce A. Boucheron, H. Luke Carter, Daphne C. Clancy, Joel P. Cooper, et al. 2009. "Anthranilimide Based Glycogen Phosphorylase Inhibitors for the Treatment of Type 2 Diabetes. Part 3: X-Ray Crystallographic Characterization, Core and Urea Optimization and in Vivo Efficacy." *Bioorganic & Medicinal Chemistry Letters* 19 (4): 1177–82. https://doi.org/10.1016/j.bmcl.2008.12.085.

Tick, Gabriella, Imre Cserpán, Viktor Dombrádi, Bernard M. Mechler, István Török, and István Kiss. 1999. “Structural and Functional Characterization of the Drosophila Glycogen Phosphorylase Gene.” *Biochemical and Biophysical Research Communications* 257 (1): 34–43. https://doi.org/10.1006/bbrc.1999.0396.

Tierney, Braden T., Zhen Yang, Jacob M. Luber, Marc Beaudin, Marsha C. Wibowo, Christina Baek, Eleanor Mehlenbacher, Chirag J. Patel, and Aleksandar D. Kostic. 2019. “The Landscape of Genetic Content in the Gut and Oral Human Microbiome.” *Cell Host & Microbe* 26 (2): 283-295.e8. https://doi.org/10.1016/j.chom.2019.07.008.

Tingirikari, Jagan Mohan Rao. 2018. “Microbiota-Accessible Pectic Poly- and Oligosaccharides in Gut Health.” *Food & Function* 9 (10): 5059–73. https://doi.org/10.1039/C8FO01296B.

Tiwari, Vinod K. 2020. *Carbohydrates in Drug Discovery and Development: Synthesis and Application*. ELSEVIER.

Touhara, Kouki K., Takanori Nihira, Motomitsu Kitaoka, Hiroyuki Nakai, and Shinya Fushinobu. 2014. “Structural Basis for Reversible Phosphorolysis and Hydrolysis Reactions of 2-O-α-Glucosylglycerol Phosphorylase.” *The Journal of Biological Chemistry* 289 (26): 18067–75. https://doi.org/10.1074/jbc.M114.573212.

Tran, Linh T., Vincent Blay, Sukanya Luang, Chatchakorn Eurtivong, Sunaree Choknud, Humbert González-Díaz, and James R. Ketudat Cairns. 2019. “Engineering Faster Transglycosidases and Their Acceptor Specificity.” *Green Chemistry* 21 (10): 2823–36. https://doi.org/10.1039/C9GC00621D.

Trethewey, R. N., A. R. Fernie, A. Bachmann, H. Fleischer-Notter, P. Geigenberger, and L. Willmitzer. 2001. “Expression of a Bacterial Sucrose Phosphorylase in Potato Tubers Results in a Glucose-Independent Induction of Glycolysis.” *Plant, Cell & Environment* 24 (3): 357–65. https://doi.org/10.1046/j.1365-3040.2001.00679.x.

Tsai, Shang-Ting, Chia Yen Liew, Chen Hsu, Shih-Pei Huang, Wei-Chien Weng, Yu-Hsiang Kuo, and Chi-Kung Ni. 2019. “Automatic Full Glycan Structural Determination through Logically Derived Sequence Tandem Mass Spectrometry.” *ChemBioChem* 20 (18): 2351–59. https://doi.org/10.1002/cbic.201900228.

Tsuda, Tomohiro, Takanori Nihira, Kazuhiro Chiku, Erika Suzuki, Takatoshi Arakawa, Mamoru Nishimoto, Motomitsu Kitaoka, Hiroyuki Nakai, and Shinya Fushinobu. 2015. “Characterization and Crystal Structure Determination of β-1,2-Mannobiose Phosphorylase from Listeria Innocua.” *FEBS Letters* 589 (24PartB): 3816–21. https://doi.org/10.1016/j.febslet.2015.11.034.

Tsujino, Seiichi, Sara Shanske, Stephanie J. Valberg, George H. Cardinet, Bradford P. Smith, and Salvatore DiMauro. 1996. “Cloning of Bovine Muscle Glycogen Phosphorylase CDNA and Identification of a Mutation in Cattle with Myophosphorylase Deficiency, an Animal Model for McArdle’s Disease.” *Neuromuscular Disorders: NMD* 6 (1): 19–26. https://doi.org/10.1016/0960-8966(95)00014-3.

Ufarté, Lisa, Élisabeth Laville, Sophie Duquesne, and Gabrielle Potocki-Veronese. 2015. “Metagenomics for the Discovery of Pollutant Degrading Enzymes.” *Biotechnology Advances* 33 (8): 1845–54. https://doi.org/10.1016/j.biotechadv.2015.10.009.

Ufarté, Lisa, Gabrielle Potocki-Veronese, and Élisabeth Laville. 2015. “Discovery of New Protein Families and Functions: New Challenges in Functional Metagenomics for Biotechnologies and Microbial Ecology.” *Frontiers in Microbiology* 6 (June). https://doi.org/10.3389/fmicb.2015.00563.

Ugalde, Juan E, Viviana Lepek, Antonio Uttaro, Julia Estrella, Alberto Iglesias, and Rodolfo A Ugalde. 1998. “Gene Organization and Transcription Analysis of the Agrobacterium Tumefaciens Glycogen (Glg) Operon: Two Transcripts for the Single Phosphoglucomutase Gene.” *Journal of Bacteriology* 180 (24): 6557–64.

Ursinus, A., and J. V. Höltje. 1994. “Purification and Properties of a Membrane-Bound Lytic Transglycosylase from Escherichia Coli.” *Journal of Bacteriology* 176 (2): 338–43. https://doi.org/10.1128/jb.176.2.338-343.1994.

Vaaje-Kolstad, Gustav, Bjørge Westereng, Svein J. Horn, Zhanliang Liu, Hong Zhai, Morten Sørlie, and Vincent G. H. Eijsink. 2010. “An Oxidative Enzyme Boosting the Enzymatic Conversion

of Recalcitrant Polysaccharides." *Science* 330 (6001): 219–22. https://doi.org/10.1126/science.1192231.

Van den Broek, L. A. M., E. L. Van Boxtel, R. P. Kievit, R. Verhoef, G. Beldman, and A. G. J. Voragen. 2004. "Physico-Chemical and Transglucosylation Properties of Recombinant Sucrose Phosphorylase from Bifidobacterium Adolescentis DSM20083." *Applied Microbiology and Biotechnology* 65 (2): 219–27. https://doi.org/10.1007/s00253-003-1534-x.

Van Hoorebeke, Annelies, Jan Stout, Ruben Van der Meeren, John Kyndt, Jozef Van Beeumen, and Savvas N. Savvides. 2010. "Crystallization and X-Ray Diffraction Studies of Inverting Trehalose Phosphorylase from Thermoanaerobacter Sp." *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 66 (Pt 4): 442–47. https://doi.org/10.1107/S1744309110005749.

Verhaeghe, Tom, Dirk Aerts, Margo Diricks, Wim Soetaert, and Tom Desmet. 2014. "The Quest for a Thermostable Sucrose Phosphorylase Reveals Sucrose 6′-Phosphate Phosphorylase as a Novel Specificity." *Applied Microbiology and Biotechnology* 98 (16): 7027–37. https://doi.org/10.1007/s00253-014-5621-y.

Viborg, Alexander Holm, Nicolas Terrapon, Vincent Lombard, Gurvan Michel, Mirjam Czjzek, Bernard Henrissat, and Harry Brumer. 2019. "A Subfamily Roadmap for Functional Glycogenomics of the Evolutionarily Diverse Glycoside Hydrolase Family 16 (GH16)." *Journal of Biological Chemistry*, September, jbc.RA119.010619. https://doi.org/10.1074/jbc.RA119.010619.

Waites, Michael J., Neil L. Morgan, John S. Rockey, and Gary Higton. 2009. *Industrial Microbiology: An Introduction*. John Wiley & Sons.

Wallenfels, Kurt. 1951. "Enzymatische Synthese von Oligosacchariden aus Disacchariden." *Naturwissenschaften* 38 (13): 306–7. https://doi.org/10.1007/BF00636782.

Walvoort, Marthe T. C., Hans van den Elst, Obadiah J. Plante, Lenz Kröck, Peter H. Seeberger, Herman S. Overkleeft, Gijsbert A. van der Marel, and Jeroen D. C. Codée. 2012. "Automated Solid-Phase Synthesis of β-Mannuronic Acid Alginates." *Angewandte Chemie (International Ed. in English)* 51 (18): 4393–96. https://doi.org/10.1002/anie.201108744.

Wang, Zhen, Zoeisha S. Chinoy, Shailesh G. Ambre, Wenjie Peng, Ryan McBride, Robert P. de Vries, John Glushka, James C. Paulson, and Geert-Jan Boons. 2013. "A General Strategy for the Chemoenzymatic Synthesis of Asymmetrically Branched N-Glycans." *Science (New York, N.Y.)* 341 (6144): 379–83. https://doi.org/10.1126/science.1236231.

Watson, Kimberly A, Clare McCleverty, Silvano Geremia, Sylvain Cottaz, Hugues Driguez, and Louise Driguez Johnson. 1999. "Phosphorylase Recognition and Phosphorolysis of Its Oligosaccharide Substrate: Answers to a Long Outstanding Question." *The EMBO Journal* 18 (17): 4619–32. https://doi.org/10.1093/emboj/18.17.4619.

Whittamore, Paul R. O., Matthew S. Addie, Stuart N. L. Bennett, Alan M. Birch, Michael Butters, Linda Godfrey, Peter W. Kenny, et al. 2006. "Novel Thienopyrrole Glycogen Phosphorylase Inhibitors: Synthesis, in Vitro SAR and Crystallographic Studies." *Bioorganic & Medicinal Chemistry Letters* 16 (21): 5567–71. https://doi.org/10.1016/j.bmcl.2006.08.047.

Wilson, Charlotte M., Miguel Rodriguez, Courtney M. Johnson, Stanton L. Martin, Tzu Ming Chu, Russ D. Wolfinger, Loren J. Hauser, et al. 2013. "Global Transcriptome Analysis of Clostridium Thermocellum ATCC 27405 during Growth on Dilute Acid Pretreated Populus and Switchgrass." *Biotechnology for Biofuels* 6 (1): 179. https://doi.org/10.1186/1754-6834-6-179.

Wong, Chi-Huey. 1996. "Practical Synthesis Of Oligosaccharides Based On Glycosyltransferases And Glycosylphosphites." In *Modern Methods in Carbohydrate Synthesis*, edited by Shaheer H. Khan, 467–91. CRC Press.

Wright, Stephen W., Virginia L. Rath, Paul E. Genereux, David L. Hageman, Carolyn B. Levy, Lester D. McClure, Scott C. McCoid, et al. 2005. "5-Chloroindoloyl Glycine Amide Inhibitors of Glycogen Phosphorylase: Synthesis, in Vitro, in Vivo, and X-Ray Crystallographic Characterization." *Bioorganic & Medicinal Chemistry Letters* 15 (2): 459–65. https://doi.org/10.1016/j.bmcl.2004.10.048.

Wu, Gang, and Miklös Müller. 2003. "Glycogen Phosphorylase Sequences from the Amitochondriate Protists, Trichomonas Vaginalis, Mastigamoeba Balamuthi, Entamoeba Histolytica and Giardia Intestinalis1." *Journal of Eukaryotic Microbiology* 50 (5): 366–72. https://doi.org/10.1111/j.1550-7408.2003.tb00151.x.

Xavier, Karina B., Ralf Peist, Marina Kossmann, Winfried Boos, and Helena Santos. 1999. “Maltose Metabolism in the Hyperthermophilic Archaeon Thermococcus Litoralis: Purification and Characterization of Key Enzymes.” *Journal of Bacteriology* 181 (11): 3358–67.

Xiao, Liang, Jordi Estellé, Pia Kiilerich, Yuliaxis Ramayo-Caldas, Zhongkui Xia, Qiang Feng, Suisha Liang, et al. 2016. “A Reference Gene Catalogue of the Pig Gut Microbiome.” *Nature Microbiology* 1 (September): 16161. https://doi.org/10.1038/nmicrobiol.2016.161.

Xiao, Liang, Qiang Feng, Suisha Liang, Si Brask Sonne, Zhongkui Xia, Xinmin Qiu, Xiaoping Li, et al. 2015. “A Catalog of the Mouse Gut Metagenome.” *Nature Biotechnology* 33 (10): 1103–8. https://doi.org/10.1038/nbt.3353.

Xie, Gary, David C. Bruce, Jean F. Challacombe, Olga Chertkov, John C. Detter, Paul Gilna, Cliff S. Han, et al. 2007. “Genome Sequence of the Cellulolytic Gliding Bacterium Cytophaga Hutchinsonii.” *Applied and Environmental Microbiology* 73 (11): 3536–46. https://doi.org/10.1128/AEM.00225-07.

Xu, Chenggang, Ranran Huang, Lin Teng, Dongmei Wang, Christopher L. Hemme, Ilya Borovok, Qiang He, et al. 2013. “Structure and Regulation of the Cellulose Degradome in Clostridium Cellulolyticum.” *Biotechnology for Biofuels* 6 (1): 73. https://doi.org/10.1186/1754-6834-6-73.

Yamamoto, Takuo, Kazuhiko Maruta, Kazuhisa Mukai, Hiroshi Yamashita, Tomoyuki Nishimoto, Michio Kubota, Shigeharu Fukuda, Masashi Kurimoto, and Yoshio Tsujisaka. 2004. “Cloning and Sequencing of Kojibiose Phosphorylase Gene from Thermoanaerobacter Brockii ATCC35047.” *Journal of Bioscience and Bioengineering* 98 (2): 99–106. https://doi.org/10.1016/S1389-1723(04)70249-2.

Yamamoto, Takuo, Mio Nishio-Kosaka, Seisuke Izawa, Hajime Aga, Tomoyuki Nishimoto, Hiroto Chaen, and Shigeharu Fukuda. 2011. “Enzymatic Properties of Recombinant Kojibiose Phosphorylase from Caldicellulosiruptor Saccharolyticus ATCC43494.” *Bioscience, Biotechnology, and Biochemistry* 75 (6): 1208–10. https://doi.org/10.1271/bbb.110116.

Ye, Yuxin, Wataru Saburi, Rei Odaka, Koji Kato, Naofumi Sakurai, Keisuke Komoda, Mamoru Nishimoto, Motomitsu Kitaoka, Haruhide Mori, and Min Yao. 2016. “Structural Insights into the Difference in Substrate Recognition of Two Mannoside Phosphorylases from Two GH130 Subfamilies.” *FEBS Letters* 590 (6): 828–37. https://doi.org/10.1002/1873-3468.12105.

Yernool, Dinesh A., James K. McCarthy, Douglas E. Eveleigh, and Jin-Duck Bok. 2000. “Cloning and Characterization of the Glucooligosaccharide Catabolic Pathway β-Glucan Glucohydrolase and Cellobiose Phosphorylase in the Marine HyperthermophileThermotoga Neapolitana.” *Journal of Bacteriology* 182 (18): 5172–79. https://doi.org/10.1128/JB.182.18.5172-5179.2000.

Yu, Hai, and Xi Chen. 2016. “One-Pot Multienzyme (OPME) Systems for Chemoenzymatic Synthesis of Carbohydrates.” *Organic & Biomolecular Chemistry* 14 (10): 2809–18. https://doi.org/10.1039/c6ob00058d.

Zeuner, Birgitte, David Teze, Jan Muschiol, and Anne S. Meyer. 2019. “Synthesis of Human Milk Oligosaccharides: Protein Engineering Strategies for Improved Enzymatic Transglycosylation.” *Molecules* 24 (11): 2033. https://doi.org/10.3390/molecules24112033.

Zhang, Han, Tanner Yohe, Le Huang, Sarah Entwistle, Peizhi Wu, Zhenglu Yang, Peter K. Busk, Ying Xu, and Yanbin Yin. 2018. “DbCAN2: A Meta Server for Automated Carbohydrate-Active Enzyme Annotation.” *Nucleic Acids Research* 46 (W1): W95–101. https://doi.org/10.1093/nar/gky418.

Zhang, Hui, Xiao Sun, Wenjie Li, Tuoping Li, Suhong Li, and Motomitsu Kitaoka. 2018. “Expression and Characterization of Recombinant Sucrose Phosphorylase.” *The Protein Journal* 37 (1): 93–100. https://doi.org/10.1007/s10930-017-9754-6.

Zhong, Chao, Božidar Duić, Juan M. Bolivar, and Bernd Nidetzky. 2020. “Three-Enzyme Phosphorylase Cascade Immobilized on Solid Support for Biocatalytic Synthesis of Cello−oligosaccharides.” *ChemCatChem* 12 (5): 1350–58. https://doi.org/10.1002/cctc.201901964.

Thesis objectives

As introduced in the previous sections, GPs are widespread in many living organisms, in particular bacteria, in which they ensure, for most of them, the final steps of glycoside catabolism. *In vitro*, these enzymes are also highly interesting biocatalytic tools for glycoside synthesis. Despite this interesting feature, the diversity of known GPs is much lower than that of others carbohydrate active enzymes. Indeed, their prevalence in all living organisms, their role in carbohydrate metabolism and their biotechnological potential are probably underestimated, as they are often misannotated as glycoside hydrolases and glycosyltransferases in genomic and metagenomic datasets because of the structural and mechanistic similarities they share with them.

When this project started, in November 2016, there were eight CAZy families containing GPs (GH3, GH13_Sub18, GH65, GH94, GH112, GH130, GT4 and GT35), against eleven today. In only three years, three families were indeed created after the biochemical characterization of new enzymes with GP activity (GH149, GH161 and GT108), highlighting the increasing interest that these enzymes currently generate. However, there was no mature approach to mine large sequence spaces for GP discovery.

In this context, the main goal of my PhD project was to develop a novel a sequence-based approach of metagenome mining for GPs, in order to identify novel functions of carbohydrate breakdown in microbial ecosystems, and to design novel enzymatic routes for the synthesis of oligosaccharides.

At the beginning of this project, several questions arised:

- How to easily handle thousands CAZy sequences to highlight those that could code for putative GPs?
- How to predict their linkage specificity?
- What is the best strategy to rapidly validate GP activity and screen for substrate specificity?
- What is the diversity, abundance and prevalence of GPs in gut microbiomes?
- What are their physiological roles in these ecosystems?
- Would some GPs display novel activities that are relevant for biotechnological applications, in particular for human health?

The work carried out to address these questions is presented in the three following chapters.

Results

Chapter 1

As explained in the previous chapter, GPs are widespread in bacteria, even though some of them are found in archaea and in eukaryotes. When I started this project, it had been shown that GPs can be found in prominent human and bovine gut bacteria. However, only one GP had been identified from uncultured bacteria, the Uhgb_MP enzyme from the GH130 family, discovered in 2013 by Ladeveze et al. in the DiscOmics group where I performed my PhD work. We thus decided to develop an approach to mine large genomic and metagenomic datasets for new GPs, in order to better understand their physiological roles and their sequence/function relationships. I established to proof of concept of this strategy by targeting the GH130 CAZy family. This family indeed contains both phosphorylases and hydrolases, and its biochemically characterized members target mannosides of various structures and origins. In addition, GH130 enzymes are involved in several interactions between eukaryotes and prokaryotes in the human gut. Furthermore, they are biocatalytic tools of interest for the synthesis of high-added value mannosides for the food and health industries.

In this chapter, I present the work carried out on this CAZy family, to analyze its functional and sequence diversity and to discover new GPs. To this aim, I tested the EFI-EST tool presented in the previous chapter to create sequence similarity networks. In the Enzymology pole of TBI, I was the first one to test this approach, which had never been used to analyze CAZyme diversity, at least in published works.

In this study, Vincent Lombard, Nicolas Terrapon and Bernard Henrissat, from the CAZy team in Marseille (France), annotated the GH130 enzymes in the targeted metagenomes. Diego Morgavi, from INRAE Saint-Genès-Champanelle (France), provided us with the bovine rumen metagenomic dataset. Elisabeth Laville provided me with the sequence abundance and prevalence data in the human gut microbiome, and I performed their analysis. I miniaturized the molybdate blue assay to screen for GP activity in micro-plates, and with Laurence Tarquis I characterized the enzymes identified thanks to the sequence analysis that I performed. Simon Charnock and Darren Cook, from Prozomix in Northumberland (UK), provided me with some of the enzymes characterized in this study. David Guieysse performed the NMR analyses and David Ropartz, from the BIBS platform at INRAE of Nantes (France), the mass spectrometry ones. Julien Durand provided me with β-mannoside standards and GPs with various specificities. Jeremy Esque helped me to manage large sequence datasets with the EFI-EST and Cytoscape tools, and supervised the phylogenetic analyses. This results presented here are described in a publication submitted to Microbial Genomics on February 27th, 2020.

Analysis of the diversity of the Glycoside Hydrolase Family 130 in mammal gut microbiomes reveals a novel mannoside-phosphorylase function

Data summary

The GH130 amino acid sequences were collected from the manually curated CAZy database and from four public gut microbial gene datasets from human gut (GigaDB, DOI: 10.5524/100064, (Li et al. 2014)), mouse gut (GigaDB, DOI: 10.5524/100114, (Xiao et al. 2015)), pig gut (EBI, PRJEB11755, (Xiao et al. 2016)) and bovine rumen (GigaScience Database, DOI: 10.5524/100391(Li et al. 2020)) automatically annotated with human supervision. The sequence accession numbers used in this study are listed in Table 1 (available in the online version of this article).

Introduction

Mannosides occur in many living organisms, from microbes to plants and animals, in the form of homo- or hetero-polymers or glycoconjugates. They play a critical role in several cellular processes, such as cell wall structuring in plants (Brett and Waldron 1996), cell signaling, as key components of proteoglycans (Sharma, Ichikawa, and Freeze 2014), lipopolysaccharides and capsular polysaccharides in bacteria, yeasts, mammals and plants (Zeleznick et al. 1965), and also as a carbohydrate storage polymer in certain parasites (Zhang and Beverley 2019). Mannoside synthesis is mostly carried out by glycoside transferases (GTs) from mannose diphosphonucleotides, whereas mannoside degradation is carried out by glycoside hydrolases (GHs). These enzymes are very diverse in terms of structures, mechanisms and specificities, as indicated by their respective membership to 13 and 15 different GT and GH families, respectively, in the classification of carbohydrate-active enzymes (CAZy database, http://www.cazy.org/, (Lombard et al. 2014)). In addition, mannoside synthesis and degradation can also be performed by mannoside phosphorylases (GPs). The latter enzymes, which are classified in related CAZy families GH130 (Senoura et al. 2011) and GT108 (Sernee et al. 2019), breakdown β-mannosides in the presence of inorganic phosphate to generate α-mannose-1-phosphate (αMan1P), through an inverting phosphorolytic mechanism. They can also act as β-mannoside-synthases from αMan1P and glycosyl acceptors, in the so-called reverse-phosphorolysis reaction. The newly created GT108 family contains only six characterized members, all acting on mannogen, a linear β-1,2-polymannoside. In contrast, the GH130 family is much more polyspecific, with 18 characterized members, of which 3 are β-1,2-mannosidases (Cuskin et al. 2015; Nihira et al. 2015) and 14 are glycoside phosphorylases acting on a large variety of substrates such as β-1,4-mannooligosaccharides (Kawahara et al. 2012; Chekan et al. 2014; Ye et al. 2016; La Rosa et al. 2019; Grimaud et al. 2019), β-1,4-mannosyl-N-acetyl-glucosamine (Nihira et al. 2013), β-1,4-mannosyl-N,N'-diacetylchitobiose (Ladevèze et al. 2013), β-1,2-mannobiose (Chiku et al. 2014; Tsuda et al. 2015), β-1,2-oligomannans (Chiku et al. 2014) and β-1,3-mannobiose (Awad et al. 2017). Their large diversity of specificities, the tolerance of some enzymes towards acceptors, and their ability to generate αMan1P through phosphorolysis of cheap substrates for the synthesis of various heteromannosides, makes GH130 glycoside phosphorylases attractive enzymes for the *in vitro* synthesis of β-mannosidic linkages, which are considered to be some of the most challenging glycosidic linkages in synthetic carbohydrate chemistry (Tang and Pohl 2015). However, the functional and sequence diversity of GH130 enzymes, and their role in the functioning of ecosystems, are underinvestigated. Less than 1% of the 2,142 GH130 enzymes listed in the CAZy database (release 08/01/2020) have been biochemically characterized. In order to predict the specificity of enzymes in this family that are still uncharacterized, in 2013 we proposed a classification into two subfamilies, namely GH130_1, strictly acting on β-1,4-mannosyl-glucose, and GH130_2, targeting β-1,4-mannosyl-N,N'-diacetylchitobiose (the core oligosaccharide of N-glycans), or β-1,4-mannooligosaccharides (Ladevèze et al. 2013). This subfamily analysis also revealed one heterogeneous GH130_NC sequence cluster, which was later shown to include enzymes of various specificities and mechanisms, such as β-1,2-mannosidases (Cuskin et al. 2015; Nihira et al. 2015), β-1,2-mannobiose- and β-1,2-oligomannan-phosphorylases (Chiku et al. 2014; Tsuda et al. 2015), and a β-1,3-mannobiose-phosphorylase (Awad et al. 2017). Even though these early subfamilies are still used today (La Rosa et al. 2019), we wished to perform a new analysis based on the greater sequence diversity available today. Indeed, the number of GH130 sequences listed in the CAZy database alone is now almost six times larger than in 2013. Because the CAZy database only lists sequences from daily releases from GenBank, we also wished to include a number of metagenomic GH130 sequences not present in CAZy in our novel analysis, as these may reveal a diversity not yet present in GenBank.

The Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) was developed in 2015 (http://efi.igb.illinois.edu/efi-est/ (Gerlt et al. 2015)) to rationalize the search of sequences of interest in large datasets. This web tool automatically generates sequence similarity networks (SSNs) to visualize sequence relationships in protein families (Atkinson et al. 2009). SSNs have recently been used to search for new functions in the glycyl radical enzyme superfamily, a class of prominent enzymes in the human gut microbiome (Levin et al. 2017), and to subdivide the multispecific GH16 family into 23 robust subfamilies (Viborg et al. 2019).

In the present study, we used SSNs to analyze the diversity of GH130 sequences listed in the CAZy database, and in four metagenome datasets selected from mammalian guts, as ecosystems that are rich in mannosides of various origins and structures (Ladevèze et al. 2017). By segregating the GH130 sequences into iso-functional meta-nodes and analyzing the conservation of the active site residues involved in catalysis and substrate accommodation, we uncover a strategy to distinguish GPs from GHs and to predict their linkage specificity. We tested our predictions by biochemically characterizing members of the unexplored meta-nodes. This study allowed us to reveal a new function of heteromannoside degradation and synthesis.

Methods

GH130 sequence mining in mammal gut metagenomes

The GH130 sequences were identified from the metagenomic sequence datasets following a procedure previously described for other metagenomics analyses (Svartström et al. 2017). The procedure consisted of a BLAST (Altschul et al. 1990) search of the protein sequences against the full-length GH130 sequences included in the CAZy database (http://www.cazy.org) in November 2018, using a cut-off E-value of 10^{-6}. Sequences that aligned over their entire length with a GH130 sequence in the database with >50% identity were directly assigned to the GH130 family. The remaining sequences were subjected to a sequence similarity search using the Hidden Markov models (HMMER v3) built for each CAZy family, allowing identification of the CAZy family they belong to (Lombard et al. 2014).

Sequence similarity networks

The web-based Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (http://efi.igb.illinois.edu/efi-est) (Gerlt et al. 2015) was used to construct sequence similarity networks (SSNs) from the GH130 sequences retrieved from the CAZy database in November 2018 (1,592 sequences) and from the 4 metagenomes (4,714 sequences), plus two sequences of recently characterized GPs (VCV21229.1 and VCV21228.1), with E-value thresholds between 10^{-30} and 10^{-90}. The node networks were visualized using Cytoscape 3.2 (https://cytoscape.org/ (Shannon et al. 2003)).

Analysis of sequence similarity

The CD-HIT program was used to analyze sequence redundancy within and between the datasets (http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi) (Huang et al. 2010). For each GH130 sequence selected for biochemical characterization, the most similar sequence was identified from the CAZy database using BLAST, with the NIH blastp default search parameters (https://blast.ncbi.nlm.nih.gov). The conservation of active site residues was analyzed using multiple sequence alignments performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The

alignments were corrected in the region containing residue K199 (reference sequence AAO78885.1), according to available structural information.

Phylogenetic analysis

First, a multiple structural alignment using mustang 3.2.2 (Konagurthu et al. 2006) was performed with the 11 GH130 sequences associated to a PDB code (1VKD, 3QC2, 3R67, 3TAW, 3WAS (Nakae et al. 2013), 4ONZ, 4UDG (Ladevèze et al. 2015) , 5A7V (Cuskin et al. 2015), 5AYC (Ye et al. 2016), 5AYE (Ye et al. 2016) and 5B0R (Tsuda et al. 2015) . Then, 10 sequences for each of the 15 meta-nodes (150 sequences in total) were chosen, including the 18 characterized sequences and 132 randomly chosen sequences. The multiple structural alignment was used to guide the multiple sequence alignment of the whole dataset (150 sequences) using MAFFT and the G-INS-i strategy (Iterative refinement, using WSP and consistency scores, of pairwise Needleman-Wunsch global alignments) (Katoh and Standley 2013). Finally, a phylogenetic tree was constructed by setting 1000 bootstrap replicates using the default neighbor-joining method (Katoh and Standley 2013; Naruya Saitou and Nei 1987) and was visualized with iTOL (Letunic and Bork 2016).

Signal peptide detection

Signal peptides were predicted using the SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen 2017).

Analysis of sequence prevalence and abundance in the human gut metagenome

The prevalence and abundance of the GH130 genes in the human gut microbiome were calculated from the abundance table of 9.9 million genes from the human gut metagenomes of 1,267 subjects (http://meta.genomics.cn/meta/dataTools, (Li et al. 2014)).

Chemicals

The reverse phosphorolysis reaction was tested using different acceptors and donors, as follows:

i) Acceptors: D-glucose (Glc) and D-fructose (Fru) (VWR Chemicals BDH), D-mannose (Man), D-galactose (Gal), N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-galactosamine (GalNAc), L-arabinose (Ara), L-rhamnose (Rha), D-xylose (Xyl), D-glucuronic acid (GlcA) (Sigma Aldrich), D-glucosamine (GlcN), N-acetyl-D-mannosamine (ManNAc) and D-galacturonic acid (GalA) (Carbosynth);

ii) Donors: α-D-glucose-1-phosphate (αGlc1P), α-D-mannose-1-phosphate (αMan1P), α-D-galactose-1-phosphate (αGal1P), α-D-N-acetyl-glucosamine-1-phosphate (αGlcNAc1P) (Sigma Aldrich); β-D-glucose-1-phosphate (βGlc1P) (Carbosynth). β-1,4-mannobiose, β-1,4-mannotriose (Megazyme), β-1,3-mannobiose and β-1,2-mannobiose (Carbosynth) were used as standards. The sodium molybdate reagent and L-ascorbic acid were purchased from Sigma Aldrich.

The hydrolysis reaction was assessed using various *p*NP-substrates : 4-nitrophenyl α-L-arabinopyranoside; 4-nitrophenyl β-L-arabinopyranoside; 4-nitrophenyl α-D-mannopyranoside; 4-nitrophenyl β-D-mannopyranoside; 4-nitrophenyl α-D-glucopyranoside; 4-nitrophenyl β-D-glucopyranoside; 4-nitrophenyl α-D-galactopyranoside; 4-nitrophenyl β-D-galactopyranoside; 4-nitrophenyl β-D-galactofuranoside; 4-nitrophenyl α-L-fucopyranoside; 4-nitrophenyl β-L-

fucopyranoside; 4-nitrophenyl β-D-fucopyranoside; 4-nitrophenyl-β-D-glucopyranosiduronic acid; 4-nitrophenyl β-D-xylopyranoside and 4-nitrophenyl α-L-rhamnopyranoside (Megazyme).

Enzyme production

The genes encoding U3, U6 and U7 were synthesized by Biomatik Limited (Cambridge, Ontario, Canada), with optimization of codon usage for expression in *E. coli*, andcloned in the pET-23a(+) vector, in fusion with N- and C-terminal $(His)_6$ tags, and expressed in the *E. coli* BL21-DE3 star strain. The recombinant E. coli cells were cultured overnight at 37°C in auto-inducible medium ZYM5052, supplemented with100 µg/ml ampicillin. The cells were harvested by centrifugation for 5 min at 5,000 rpm, before being resuspended and concentrated in 20 mM Tris-HCl, pH 7.0 to obtain a final $OD_{600\ nm}$ of 80. They were then lysed by sonication. The cell debris were centrifuged at 13,000 rpm for 10 minutes and the cytoplasmic extracts were filtered using a 0.20 µm Minisart RC4 syringe filter.

The U1 and U4 enzymes were produced with an N-terminal $(His)_6$ tag by Prozomix (Northumberland, UK), and provided in the form of cell free extract powder.

Enzyme purification

The recombinant *E. coli* BL21-DE3 star cells were harvested, resuspended in the TALON buffer (20 mM Tris-HCl, pH 7.0, 300 mM NaCl) at $OD_{600\ nm}$ 80, and lysed by sonication.The supernatant was collected after centrifugation at 13,000 rpm for 10 minutes and the cytoplasmic extracts were filtered using a 0.20 µm Minisart RC4 syringe filter. For U1 and U4, the cell free extract powder was solubilized in the TALON buffer, at 10 mg mL^{-1}.

The cytoplasmic extracts in the TALON buffer were subsequently applied to a TALON resin loaded with cobalt (GE Healthcare), washed with 20 mM Tris-HCl, pH 7.0, 300 mM NaCl, then eluted with 20 mM Tris-HCl, pH 7.0, 300 mM NaCl, 20 mM imidazole. The purified protein was eluted with 20 mM Tris-HCl, pH 7.0, 300 mM NaCl, and 200 mM imidazole. Lastly, the purified protein was desalted using a PD-10 column (GE Healthcare) and recovered in 20 mM Tris-HCl, pH 7.0. The protein concentration was determined by spectrometry using a NanoDrop (Thermo Fisher Scientific, Waltham, MA).

Acceptor and donor specificity screening

All enzymatic reactions were carried out at 37 °C. The Molybdenum Blue Activity Assay was used to assess the release of inorganic phosphate during reverse phosphorolysis (Gawronski and Benson 2004; De Groeve et al. 2010; Macdonald et al. 2019). The reaction mixture was composed of 45 µL of purifed enzyme (fnal concentration 0.03 mg ml^{-1} for U1 and U3, and 0.0025 mg ml^{-1} for U7) in 255 µl of a solution containing 20 mM Tris/HCl (pH 7.0), 200 mM sodium molybdate, 10 mM sugar-1-phosphate and 10 mM monosaccharide (fnal concentrations).

At sampling times, 50 µL of the reaction mixture were transferred to a 96-well plate containing, per well, 150 µL of a L-ascorbic acid solution 0.24% (w/v) in 0.1 N HCl. After incubation at room temperature for 5 min, 150 µL of stop solution (2% (v/v) acetic acid and 2% (w/v) sodium citrate tribasic dihydrate) were added to stop the reaction. The absorbance was measured at 655 nm using a microplate reader (InfiniteM200pro; TECAN). One unit of activity corresponds to the amount of enzyme that catalyzes the production of 1 µmol of inorganic phosphate min^{-1} in the assay conditions

Hydrolysis assays

The hydrolytic activity of the purified enzymes was quantified by monitoring the hydrolysis of *p*NP-glycosides (list above). Assays were performed for 1h at 37°C in 96-well microtiter plates. The reaction mixture was composed of 1mM (final concentration) *p*NP-glycosides in 140 µL of 20 mM Tris-HCl buffer, pH 7.0, and of 40 µL of 1 mg mL^{-1} of purified enzyme. *Para*-nitrophenol (0 to 1mM) was used as standard. The reaction was stopped by adding 180 µL of 200 mM Na_2CO_3. At the initial and fnal reaction times, the OD was measured at 405 nm with a microplate reader (InfiniteM200pro; TECAN).

Mannoside synthesis

The synthesis of manno-oligosaccharides by reverse-phosphorolysis was performed over 24h at 37°C in a total reaction volume of 100 µL, with 0.1 mg mL^{-1} purified protein, 10mM αMan1P and 10 mM acceptor in 20 mM Tris-HCl buffer, pH 7.0. The variation of substrate and product amounts was assessed by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD).

HPAEC-PAD analysis

Carbohydrates and αMan1P were separated on a 4×250mm Dionex CarboPack PA100 column. A gradient of sodium acetate (from 0 to 150 mM in 15 min) and an isocratic step of 300 mM sodium acetate in 150 mM NaOH was applied at a flow rate of 1mL min^{-1}. Detection was performed by using a Dionex ED40 module with a gold working electrode and an Ag/AgCl pH reference.

NMR Spectroscopy

All spectra were acquired using a Bruker Advance 500-MHz spectrometer with a 5-mm z-gradient TBI probe at 298 K, and an acquisition frequency of 500.13 MHz for ^{1}H and 125.75 MHz for ^{13}C NMR. The data were processed using TopSpin 3.0 software. The freeze-dried reaction media and standards were re-suspended in deuterated water. Deuterium oxide was used as the solvent and the chemical shift scales were internally referenced to sodium 2,2,3,3-tetradeuterio-3-trimethylsilylpropanoate (d4-TSP) for ^{1}H and to the non-deuterated acetone (singlet at 30.89 ppm) for ^{13}C NMR. The various signals were assigned by comparison with signals obtained from D-mannose, D-glucose, αMan1P, β-1,4-D-mannobiose (Megazyme, Irleland), and β-1,3-mannobiose (Carbosynth, UK). Usual 2D techniques such as TOCSY, HSQC, and HMBC were used.

Mass spectrometry

The reverse-phosphorolysis reaction performed with U1 was analyzed by mass spectrometry (MS), on a Synapt G2 Si HDMS (Waters Corp., Manchester, UK) equipped with a LockSpray Exact Mass Ionization Source (Electrospray ion source, ESI). The instrument was operated in positive polarity in the sensitivity mode. The parameters used for electrospray were as follows: capillary voltage: 2.1 kV; sampling cone: 70; desolvation temperature: 120°C; desolvation gas: 350 L h^{-1}. The tandem MS (MS/MS) measurement was performed in the transfer cell of the triwave setting using an energy of 22. After drying, the sample was solubilized in H_2O/MeOH and infused at a flow rate of 5 µL min^{-1}. MS and MS/MS measurements were recorded for 30 seconds. Data were acquired using MassLynx 4.1 and processed using Mmass 5.5.0 (Niedermeyer and Strohalm 2012). The MS/MS spectrum was annotated according to the nomenclature of Domon and Costello (Domon and Costello 1988).

Results

GH130 sequence mining in mammal gut microbiomes

In order to extend the GH130 diversity from the genomic to the metagenomic sequence space, 4 mammal gut metagenomes, totaling 33,962,473 genes, were searched for sequences containing at least one GH130 module (Figure 1A). Respectively GH130 sequences were retrieved from the human, mouse, pig and bovine gut metagenomes (1,205, 467, 681 and 2,361 sequences respectively, making a total of 4,714) (Table 2). The relative numbers of GH130 genes in these four datasets are of the same order of magnitude, varying from 6×10^{-3} to 26×10^{-3}% of the genes in each metagenome. Te proportion of sequences that lack N- and/or C-terminal ends varies between 31–66%.Logically, the larger is the N50 assembled contig length in the metagenomic dataset and the smaller is the percentage of fragmentary sequences. Most of the metagenomic sequences contain one single catalytic module. Only 19 combine a GH130 module with a GH2, GH43, GH92, GH127, or with a GH142, module (Table 3). This very low number of multi-modular sequences is probably not due to gene fragmentation in the metagenomic datasets, since only six of the 1,592 sequences listed in the CAZy database as of November 2018 are multimodular (GH130-GT81 and GH130-GH130 modular combinations). None of the CAZy or metagenomic sequences contains a CBM domain. Of the 4,714 metagenomic sequences, only 0.7% (31 sequences) are fully identical to sequences already listed in the CAZy database. Among the metagenomic sequences, only 0.5% (24 sequences) are identical. With a threshold of 90% sequence identity, sequence redundancy between each metagenome and the CAZy database does not exceed 10.7%. Between the metagenomes, it reaches 51.5%, highlighting the similarities between mammal gut microbiomes regarding these mannoside-metabolizing enzymes (Table 2).

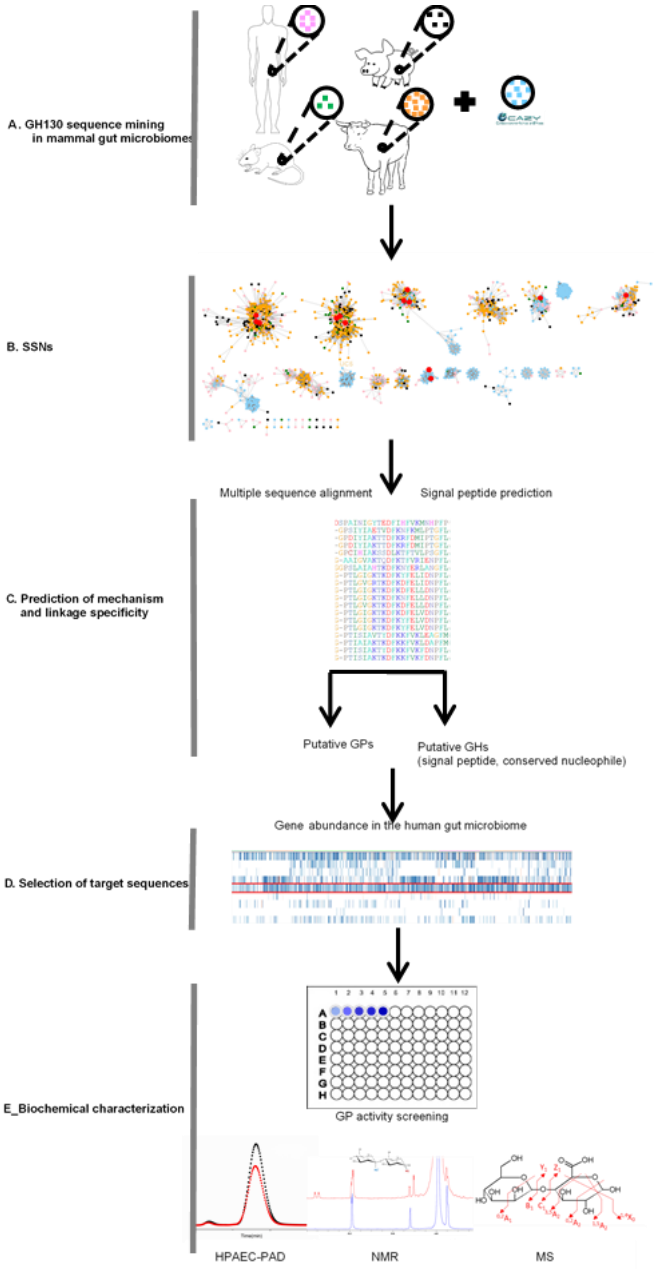


Figure 1: Strategy of glycoside phosphorylase discovery by metagenome mining. (A), the human, bovine, pig and mouse gut metagenomes were mined for GH130 sequences, by comparing the GH130 sequences of the CAZy database and the HMM built for each CAZy family to each metagenome; (B), sequence similarity networks were constructed using the EFI-EST webtool; (C), for each SSN meta-node, the mechanism of glycosidic linkage breakdown (phosphorolysis or hydrolysis) and linkage specificity were predicted based on detection of signal peptides and analysis of the conservation of the residues involved in catalysis and substrate binding; (D), for each uncharacterized SSN meta-node, the most abundant and prevalent sequences of the human gut microbiome were identified for biochemical characterization; (E), biochemical characterization of the new glycoside phosphorylases, by screening acceptor and donor specificity (analysis of the rate of inorganic phosphate release during the reverse-phosphorolysis reaction, using the Molybdenum Blue Activity Assay), and by elucidating the structure of the synthesized mannosides by HPAEC-PAD, NMR and MS.

Table 2: Diversity of metagenomic GH130 sequences in mammal gut metagenomes

Data bases			GH130 sequences					
Metagenome	**Number of genes**	**N50 average contig length (kb)**	**Number of GH130 sequences**	**% of GH130**	**Number of complete sequences**	**Number of multimodular sequences**	**Number of redundant sequences with those of the CAZy database**	
							100 % identity	**≥ 90 % identity**
Human	9,878,647	5.0	1,205	12×10^{-3}	676 (56 %)	3	22	210
Mouse	7,685,872	6.6	467	6×10^{-3}	324 (69 %)	0	7	53
Pig	2,572,074	1.87	681	26×10^{-3}	298 (44 %)	0	1	62
Bovine	13,825,880	0.91	2,361	17×10^{-3}	800 (34 %)	16	1	180
Total	33,962,473	-	4,714	-	2,098	19	31	505
Redundancy between metagenomes							24	2,426

Table 3: List of the multi-modular sequences containing a GH130 module, and their location in the SSN meta-nodes.

Sequence ID	Modularity	SSN Cluster
CAZy DB		
ACY49488.1	GH130-GH130	small meta-node (<20 sequences)
AEN74517.1	GH130-GH130	small meta-node (<20 sequences)
AKM78124.1	GH130-GH130	small meta-node (<20 sequences)
AKM83974.1	GH130-GH130	small meta-node (<20 sequences)
CCP27154.1	GH130-GT81	small meta-node (<20 sequences)
AEE92278.1	GH130-GT81	small meta-node (<20 sequences)
Human metagenome		
DOM025_GL0036051	GH130-GH92	Singletons
MH0378_GL0079806	GH92-GH130	C3
MH0369_GL0053907	GH127-GH130_distantly related	small meta-node (<20 sequences)
Rumen metagenome		
0081_GL1521727	GH130-GH78	C3
3042_GL0231557	GH130-GH78	C3
0081_GL0217456	GH130-GH78	C3
7049_GL1068002	GH2-GH130	C3
554_GL2279888	GH92-GH130	C3
7049_GL1012174	GH92-GH130	C3
2009040_GL1278742	GH43-GH130	C3
555_GL0472243	GH130_dist-GH142	UC4
555_GL1429890	GH130_dist-GH142	UC4
0081_GL0027620	GH130_dist-GH142	UC4
0081_GL1803283	GH130_dist-GH142	UC4
2009040_GL1002494	GH130_dist-GH142	UC4
7049_GL0472254	GH130_dist-GH142	UC4
7049_GL0484382	GH130_dist-GH142	UC4
552_GL1270477	GH130_dist-GH142	UC4
554_GL0058843	GH130_dist-GH142	UC4

Sequence diversity in the GH130 family

To analyze sequence diversity in the GH130 family, we constructed SSNs with the 1,592 sequences from the CAZy database, plus 2 sequences corresponding to recently characterized GPs, the 4,714 metagenomic sequences (Figure 1B). The low number of identical sequences does not affect clusterization, since they appear in the same nodes. We also checked that the presence of the four multi-modular GH130-GH130 sequences in our dataset (Table 3) did not create erroneous edges between otherwise distinct meta-nodes. Taken independently, the sequences of each of the modules indeed gather in the same meta-node. Different E-value thresholds (10^{-30} and 10^{-90}) were tested to construct the SSNs

(Figures 2 and 3). For each SSN, we mapped the known functions of the 18 characterized GH130 sequences on the meta-nodes obtained. We defined isofunctional clusters as meta-nodes containing characterized members with the same osidic linkage and acceptor specificities. The optimal E-value threshold was defined as the highest threshold yielding to the separation of the sequences into the maximum number of isofunctional meta-nodes. Optimal clusterization was obtained by setting an E-value threshold of 10^{-70}, corresponding to 63% sequence identity (Figure 2).

In this optimal configuration, we obtained five isofunctional meta-nodes (meta-nodes C1, and C3 to C6, Figure 2). The characterized GPs targeting β-1,2-mannosides appear in two different meta-nodes (C4 and C6). This result cannot be explained by differences of specificity towards β-1,2-mannosides of various chain lengths, since meta-node C6 contains both a β-1,2-mannobiose phosphorylase and a β-1,2-oligomannan phosphorylase. Conversely, the GPs acting on β-1,4-Man-GlcNAc and on β-1,4-mannosides (sequences belonging to the former GH130_2 subfamily) are grouped in the same meta-node (C2), apart from the those targeting β-1,4-Man-Glc (former GH130_1 subfamily). It was not possible to obtain an SSN representation grouping GPs which act on β-1,2-mannosides into one single meta-node while splitting them out from those acting on β-1,3- mannosides (C5, Figure 3). In addition, even at the lowest E-value tested (10^{-90}), the characterized sequences targeting β-1,4-Man-GlcNAc and β-1,4- mannosides are still grouped in the same meta-node. This may be due to the high promiscuity towards acceptors of these enzymes, as demonstrated for the RaMP2, BT1033 and Uhgb_MP enzymes (Kawahara et al. 2012; Nihira et al. 2013; Ladevèze et al. 2013).

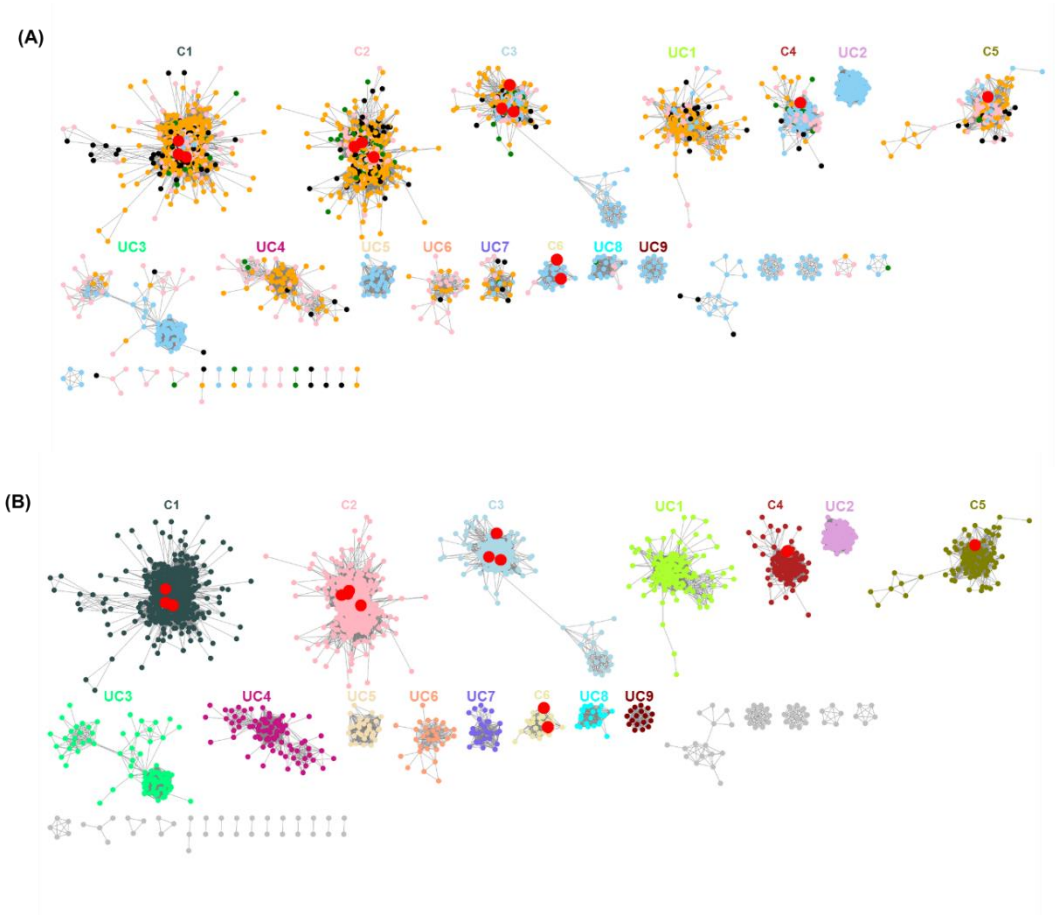
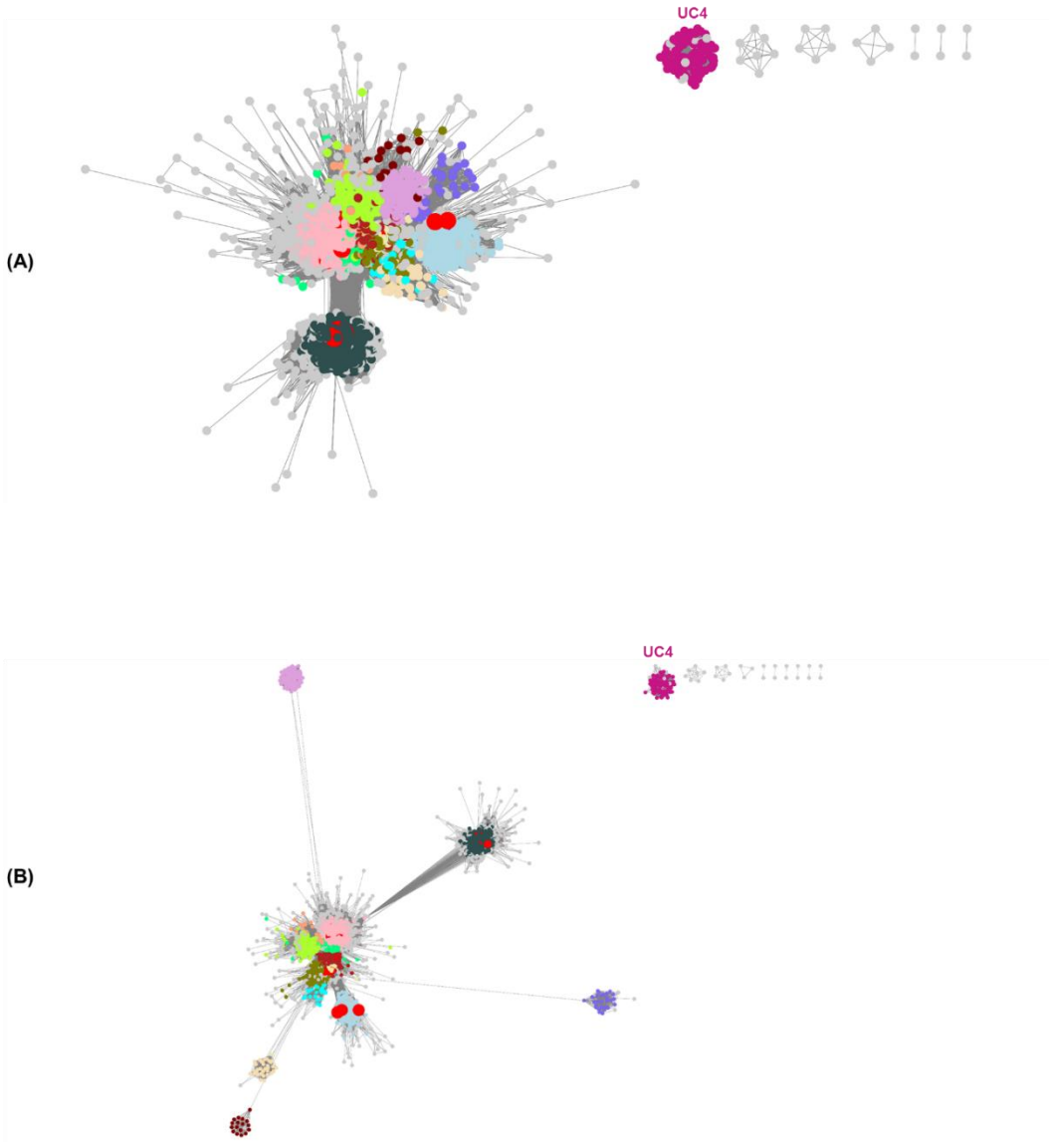
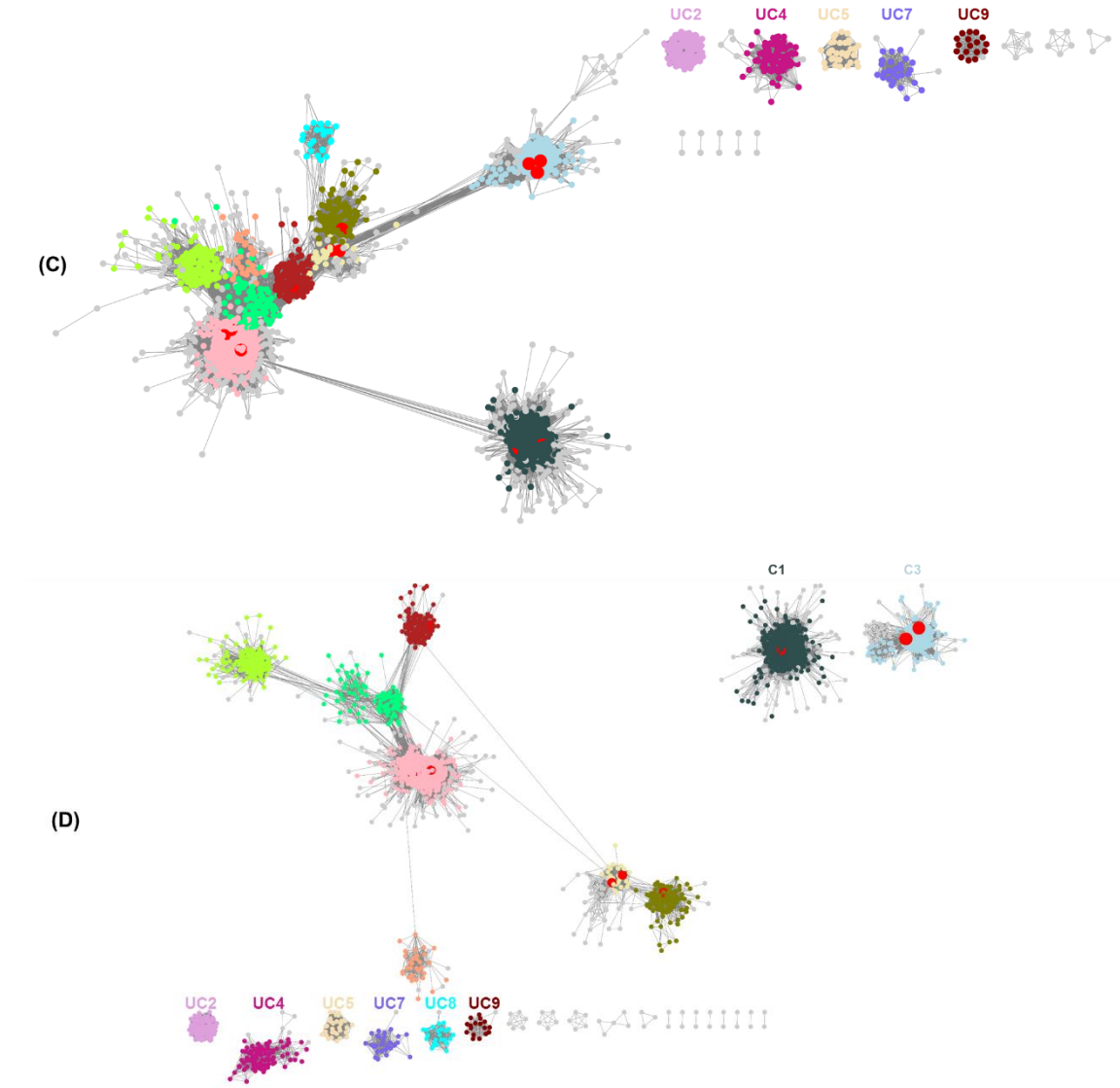


Figure 2: Protein sequence similarity network (SSN) of the 6,308 GH130 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-70}. Each of the 3,637 nodes contains sequences with more than 70% identity. Nodes are connected by an edge when the pairwise sequence identity is over 63%. In this figure, the 694 sequences appearing in 641 singletons have been omitted. Red nodes containing biochemically characterized members, belonging to meta-nodes C1 to C6. The meta-nodes containing at least 20 sequences and with no characterized members are labelled as UC1 to UC9. (A), nodes colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; green, mouse gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome. (B), nodes colored according to the meta-node they belong to.




(C)
UC2
UC4
UC5
UC7
UC9
(D)
C1
C3
UC2
UC4
UC5
UC7
UC8
UC9
(E)
C1
C2
C3
UC1
UC2
C4
C5
UC3
UC4
UC5
UC7
C6
UC6
UC8
UC9

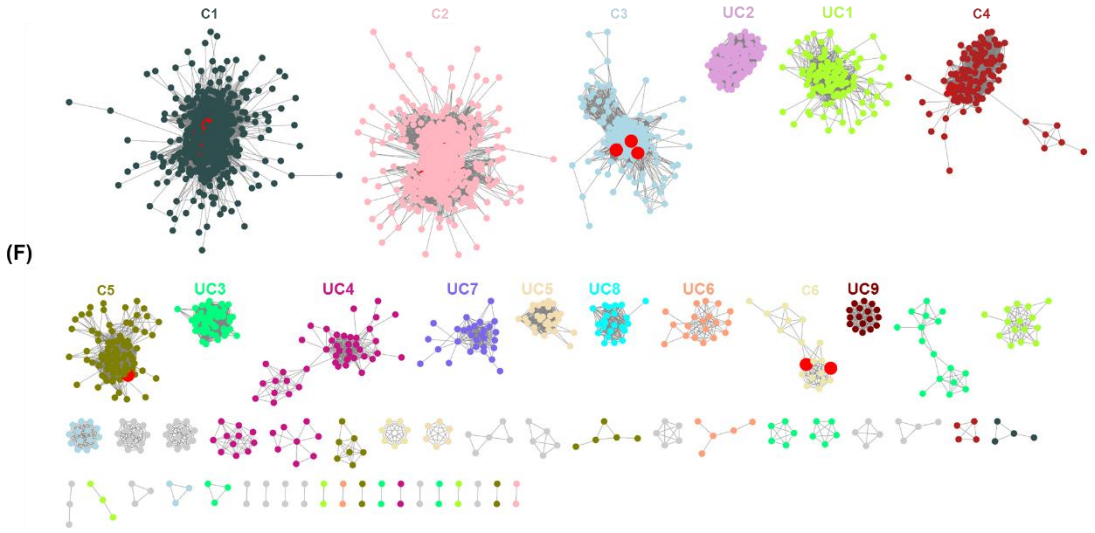


Figure 3: Protein sequence similarity networks (SSNs) of the 6,308 GH130 sequences from the CAZy database and from gut metagenomes, with various E-value thresholds. The 3,637 nodes contain sequences with more than 70% identity. Red nodes correspond to biochemically characterized members of the GH130 family. Other nodes are colored according to the meta-node they belong to in the optimal SSN, obtained with the E-value threshold 10^{-70}. (A), E-value threshold 10^{-30}, corresponding to 32% sequence identity; the 79 sequences appearing in 78 singletons have been omitted from the figure. (B), E-value threshold 10^{-40}, corresponding to 41% sequence identity; the 189 sequences appearing in 177 singletons have been omitted from the figure. (C), E-value threshold 10^{-50}, corresponding to 53% sequence identity; the 333 sequences appearing in 315 singletons have been omitted from the figure. (D), E-value threshold 10^{-60}, corresponding to 61% sequence identity; the 491 sequences appearing in 457 singletons have been omitted from the figure. (E), E-value threshold 10^{-80}, corresponding to 64% sequence identity; the 896 sequences appearing in 826 singletons have been omitted from the figure. (F), E-value threshold 10^{-90}, corresponding to 65% sequence identity; the 1130 sequences appearing in 1027 singletons have been omitted from the figure.

The optimal SSN contains 15 meta-nodes containing more than 20 sequences each, this is 1 of the criteria used to define subfamilies in the CAZy database (Viborg et al. 2019). Six meta-nodes contained characterized enzymes (meta-nodes C1 to C6). The other nine meta-nodes (UC1 to UC9) did not contain characterized sequences and were therefore investigated for their functions. In addition to these meta-nodes, the SSN also contains 20 meta-nodes with fewer than 20 sequences, and 641 singletons sharing less than 63% identity with the sequences in meta-nodes. These small meta-nodes and singletons contain 798 sequences, which correspond to 12.7% of the 6,308 sequences used to construct the SSN. They correspond to small sequences, of less than 292 amino acids in size, which is the length of the shortest full-length GH130 sequence indexed in the CAZy database (CEP78012.1). These short sequences appearing as singletons represent the majority (94%) of the incomplete metagenomic sequences. The longest sequences correspond to the few multi-modular proteins, which are distributed in different meta-nodes, in particular C3 and UC4 (Tables 1 and 3).

Prediction of mechanism

In order to support the functional assignment of meta-nodes containing characterized proteins, and to predict the function of the members of the uncharacterized SSN meta-nodes, we predicted the mechanism of osidic linkage breakdown (phosphorolysis or hydrolysis) for all the meta-nodes, based on three criteria (Figure 1C). The first one was the presence or absence of a signal peptide in the sequences. Because GPs need intracellular phosphate to perform phosphorolysis, we assumed that they were all cytoplasmic enzymes. To test this hypothesis, we predicted the presence or absence of a signal peptide in all the sequences of characterized GPs from the GH3, GH13_18, GH65, GH94, GH112, GH130, GH149, GH161, GT4, GT35 and GT108 families, using the SignalP 4.1 server. No signal peptide was detected. In addition, none of the sequences from the GH130 SSN meta-nodes containing characterized GPs was predicted to have a signal peptide (Table 4). Conversely, the meta-node containing the characterized GHs (β-1,2-mannosidases) contained sequences with signal peptides (for example, two of the three characterized members). Based on this finding, we considered that the meta-nodes containing sequences with a predicted signal peptide do not include GPs.

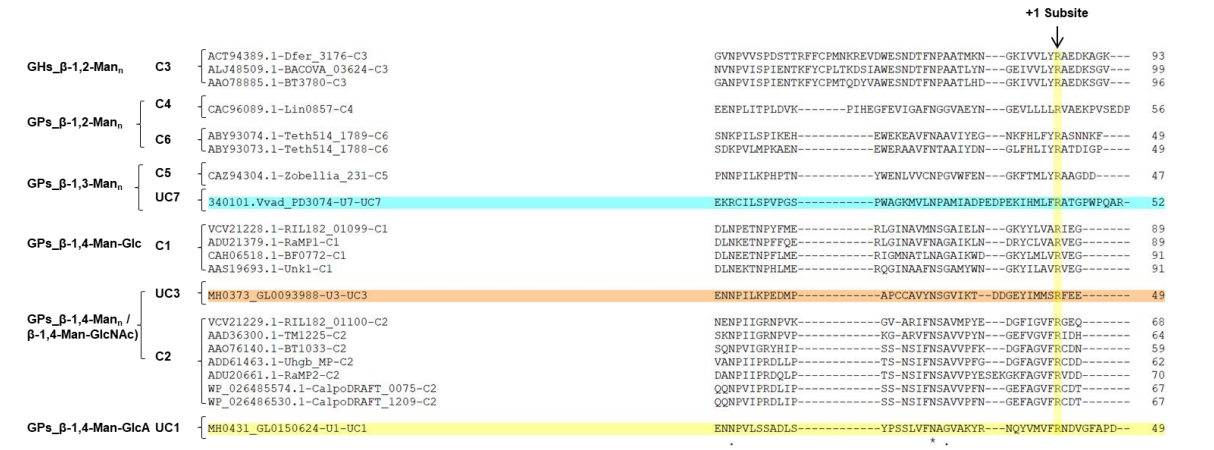
The second criterion was the conservation of the catalytic machinery. As previously shown for GH130 enzymes (Cuskin et al. 2015; Ladevèze et al. 2013), inverting GPs indeed require a single catalytic amino acid to perform the catalytic event, with phosphate playing the role of a catalytic base . Conversely, inverting GHs require the presence in their catalytic site of an additional carboxylic amino acid acting as a catalytic base to catalyze the hydrolytic reaction. We thus examined the conservation of the catalytic proton donor Aspartate (D), and searched for D or Glutamate (E) residues which may act as a catalytic base, by aligning the non-truncated sequences of each meta-node with those of the 18 phosphorylases and hydrolases characterized previously. Sequences ADD61463.1 and AAO78885.1 were chosen as references for amino acid numbering (Table 4). We assigned the GP activity to the meta-nodes where more than 95% (to take into account the rare shifts in the automatic multiple sequence alignments) of sequences included the putative catalytic donor but lacking the catalytic base, and the GH activity to those where more than 95% of sequences contained the two putative catalytic residues.

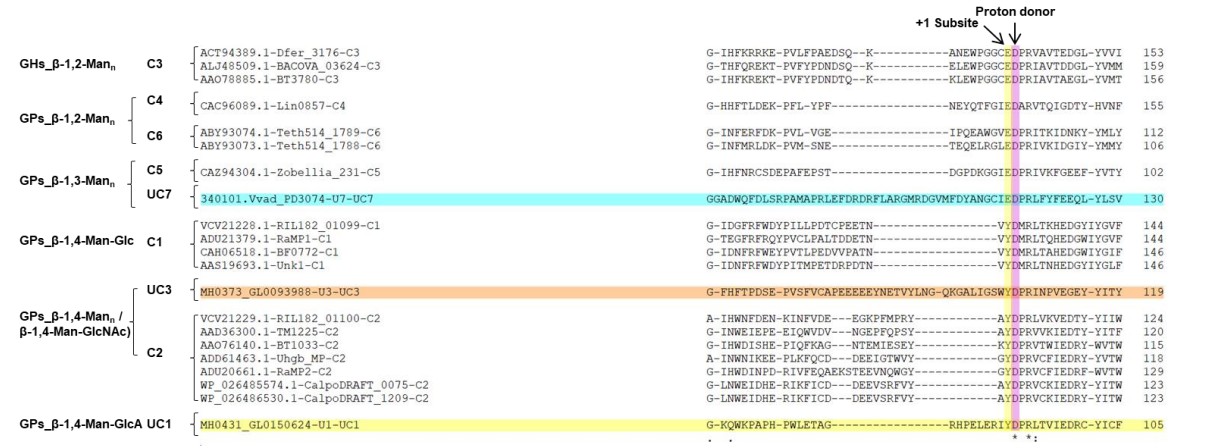
The final criterion was the conservation of the phosphate binding residues in GP-encoding sequences. By examining the multiple alignments in each meta-node, we assigned the GP activity to the meta-nodes where more than 90% of sequences included the arginine/histidine (R/H) pair described as being involved in Pi binding (Ladevèze et al. 2015). Conversely, all the sequences clustering with the characterized GHs lack the phosphate binding residues. An overview of this analysis is given in Figure 4, highlighting key regions in the multiple sequence alignment, as proton donor, putative catalytic base, +1 subsite and Pi binding site.

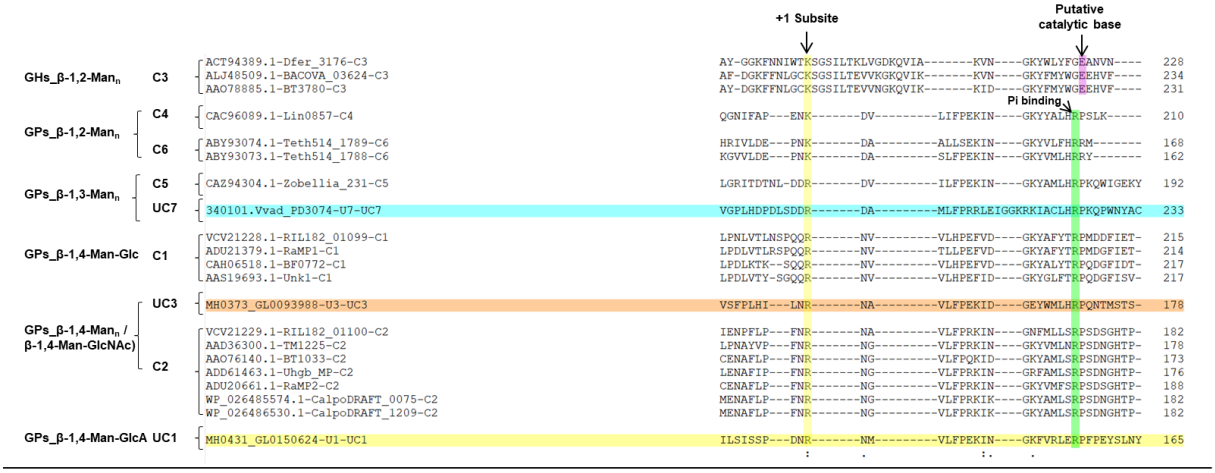
Based on the three criteria, we assigned all UC meta-nodes except UC4 to GPs. It is noted that meta-node UC3 has 15 sequences (< 10%) where the phosphate binding residue H is replaced by Tyrosine (Y). This variation in phosphate binding residues have already been described in other GP-containing families, such as members of family GH13_18 (Wildberger, Todea, and Nidetzky 2012). In the sequences belonging to the meta-node UC4, the catalytic and phosphate binding residues are not conserved, but the presence of a signal peptide in some of these sequences suggests that they could be hydrolases.

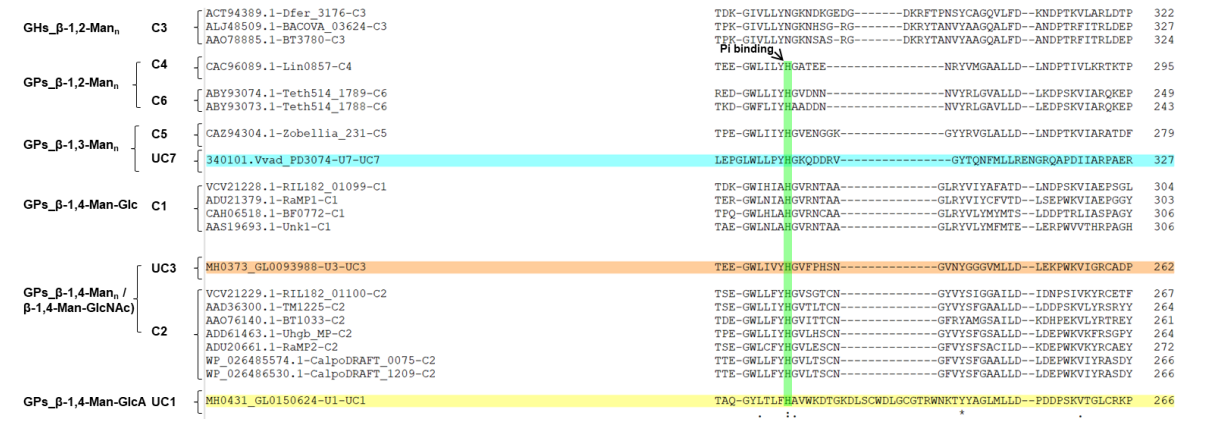
Table 4: Presence of signal peptides and conserved residues in the sequences of each SSN meta-node. The proven activities are mentioned in bold type.

Reference sequences				Catalytic residues		Phosphate binding residues		Subsite +1				Predicted mechanism and linkage specificity
				Proton donor_D	Catalytic base_E/D							
β-1-4_Man_GlcNAc phosphorylase ADD61463.1 *Ladevèze et al., 2013*				**D** 104	–	**R** 168	**H** 231	**R** 59	**Y** 103	**R** 150	**D** 304	
β-1,2-Mannosidase AAO78885.1 *Cuskin et al.,2015*				**D** 142	**E** 227	–	–	**R** 89	**E** 141	**K** 199	**D** 363	
Meta-node	**Numbers of full length sequences in CAZy/metagenomes/Total**		**SignalP**	**Catalytic residues**		**Phosphate binding residues**		**+1 Subsite**				
				Proton donor_D	**Catalytic base_E/D**							
C1	221/836/1057		No	D	_	R	H	R	**Y**	**R**	D	**GPs-β-1-4**
C2	268/757/1025		No	D	_	R	H	R	**Y**	**R**	D	**GPs-β-1-4**
C3	113/152/265		Yes	D	D/E	_	_	R	**E**	**K**	D	**GHs-β-1-2**
C4	290/76/366	295/366	No	D	_	R	H	R	**E**	**K**	D	**GPs-β-1-2**
		70/366						R	**E**	**R**	D	
C5	42/70/112		No	D	_	R	H	R	**E**	**R**	D	**GPs-β-1-3**
C6	85/0/85		No	D	_	R	H	R	**E**	**K**	D	**GPs-β-1-2**
UC1	31/66/97		No	D	_	R	H	R	**Y**	**R**	D	**GPs-β-1-4**
UC2	199/0/199		No	D	_	R	H	R	**E**	**K**	D	GPs-β-1-2
UC3	135/18/153	124/153	No	D	_	R	H	R/D	**E**	**K**	D	GPs-β-1-2
		7/153						R/D	**E**	**R**		GPs-β-1-3
		14/153						R/D	**Y**	**R**		**GPs-β-1-4**
UC4	0/53/53		Yes	_	_	_	_	S	M	N	_	GHs
UC5	61/0/61		No	D	_	R	H	R	**E**	**K**	D	GPs-β-1-2
UC6	0/11/11		No	D	_	R	H	R	**I**	**R**	D	GPs
UC7	9/10/19		No	D	_	R	H	R	**E**	**R**	D	**GPs-β-1-3**
UC8	26/7/33		No	D	_	R	H	R	**E**	**K**	D	GPs-β-1-2
UC9	29/0/29		No	D	_	R	H	R	**E**	**K**	D	GPs-β-1-2


+1 Subsite
GHs_β-1,2-Man$_n$ C3
GPs_β-1,2-Man$_n$ C4
C6
GPs_β-1,3-Man$_n$ C5
UC7
GPs_β-1,4-Man-Glc C1
GPs_β-1,4-Man$_n$ / β-1,4-Man-GlcNAc) UC3
C2
GPs_β-1,4-Man-GlcA UC1
ACT94389.1-Dfer_3176-C3
ALJ48509.1-BACOVA_03624-C3
AAO78885.1-BT3780-C3
CAC96089.1-Lin0857-C4
ABY93074.1-Teth514_1789-C6
ABY93073.1-Teth514_1788-C6
CAZ94304.1-Zobellia_231-C5
340101.Vvad_PD3074-U7-UC7
VCV21228.1-RIL182_01099-C1
ADU21379.1-RaMP1-C1
CAH06518.1-BF0772-C1
AAS19693.1-Unk1-C1
MH0373_GL0093988-U3-UC3
VCV21229.1-RIL182_01100-C2
AAD36300.1-TM1225-C2
AAO76140.1-BT1033-C2
ADD61463.1-Uhgb_MP-C2
ADU20661.1-RaMP2-C2
WP_026485574.1-CalpoDRAFT_0075-C2
WP_026486530.1-CalpoDRAFT_1209-C2
MH0431_GL0150624-U1-UC1


Proton donor
+1 Subsite
GHs_β-1,2-Man$_n$ C3
GPs_β-1,2-Man$_n$ C4
C6
GPs_β-1,3-Man$_n$ C5
UC7
GPs_β-1,4-Man-Glc C1
GPs_β-1,4-Man$_n$ / β-1,4-Man-GlcNAc) UC3
C2
GPs_β-1,4-Man-GlcA UC1
ACT94389.1-Dfer_3176-C3
ALJ48509.1-BACOVA_03624-C3
AAO78885.1-BT3780-C3
CAC96089.1-Lin0857-C4
ABY93074.1-Teth514_1789-C6
ABY93073.1-Teth514_1788-C6
CAZ94304.1-Zobellia_231-C5
340101.Vvad_PD3074-U7-UC7
VCV21228.1-RIL182_01099-C1
ADU21379.1-RaMP1-C1
CAH06518.1-BF0772-C1
AAS19693.1-Unk1-C1
MH0373_GL0093988-U3-UC3
VCV21229.1-RIL182_01100-C2
AAD36300.1-TM1225-C2
AAO76140.1-BT1033-C2
ADD61463.1-Uhgb_MP-C2
ADU20661.1-RaMP2-C2
WP_026485574.1-CalpoDRAFT_0075-C2
WP_026486530.1-CalpoDRAFT_1209-C2
MH0431_GL0150624-U1-UC1


+1 Subsite
Putative catalytic base
Pi binding
GHs_β-1,2-Man$_n$ C3
GPs_β-1,2-Man$_n$ C4
C6
GPs_β-1,3-Man$_n$ C5
UC7
GPs_β-1,4-Man-Glc C1
GPs_β-1,4-Man$_n$ / β-1,4-Man-GlcNAc) UC3
C2
GPs_β-1,4-Man-GlcA UC1
ACT94389.1-Dfer_3176-C3
ALJ48509.1-BACOVA_03624-C3
AAO78885.1-BT3780-C3
CAC96089.1-Lin0857-C4
ABY93074.1-Teth514_1789-C6
ABY93073.1-Teth514_1788-C6
CAZ94304.1-Zobellia_231-C5
340101.Vvad_PD3074-U7-UC7
VCV21228.1-RIL182_01099-C1
ADU21379.1-RaMP1-C1
CAH06518.1-BF0772-C1
AAS19693.1-Unk1-C1
MH0373_GL0093988-U3-UC3
VCV21229.1-RIL182_01100-C2
AAD36300.1-TM1225-C2
AAO76140.1-BT1033-C2
ADD61463.1-Uhgb_MP-C2
ADU20661.1-RaMP2-C2
WP_026485574.1-CalpoDRAFT_0075-C2
WP_026486530.1-CalpoDRAFT_1209-C2
MH0431_GL0150624-U1-UC1


Pi binding
GHs_β-1,2-Man$_n$ C3
GPs_β-1,2-Man$_n$ C4
C6
GPs_β-1,3-Man$_n$ C5
UC7
GPs_β-1,4-Man-Glc C1
GPs_β-1,4-Man$_n$ / β-1,4-Man-GlcNAc) UC3
C2
GPs_β-1,4-Man-GlcA UC1
ACT94389.1-Dfer_3176-C3
ALJ48509.1-BACOVA_03624-C3
AAO78885.1-BT3780-C3
CAC96089.1-Lin0857-C4
ABY93074.1-Teth514_1789-C6
ABY93073.1-Teth514_1788-C6
CAZ94304.1-Zobellia_231-C5
340101.Vvad_PD3074-U7-UC7
VCV21228.1-RIL182_01099-C1
ADU21379.1-RaMP1-C1
CAH06518.1-BF0772-C1
AAS19693.1-Unk1-C1
MH0373_GL0093988-U3-UC3
VCV21229.1-RIL182_01100-C2
AAD36300.1-TM1225-C2
AAO76140.1-BT1033-C2
ADD61463.1-Uhgb_MP-C2
ADU20661.1-RaMP2-C2
WP_026485574.1-CalpoDRAFT_0075-C2
WP_026486530.1-CalpoDRAFT_1209-C2
MH0431_GL0150624-U1-UC1

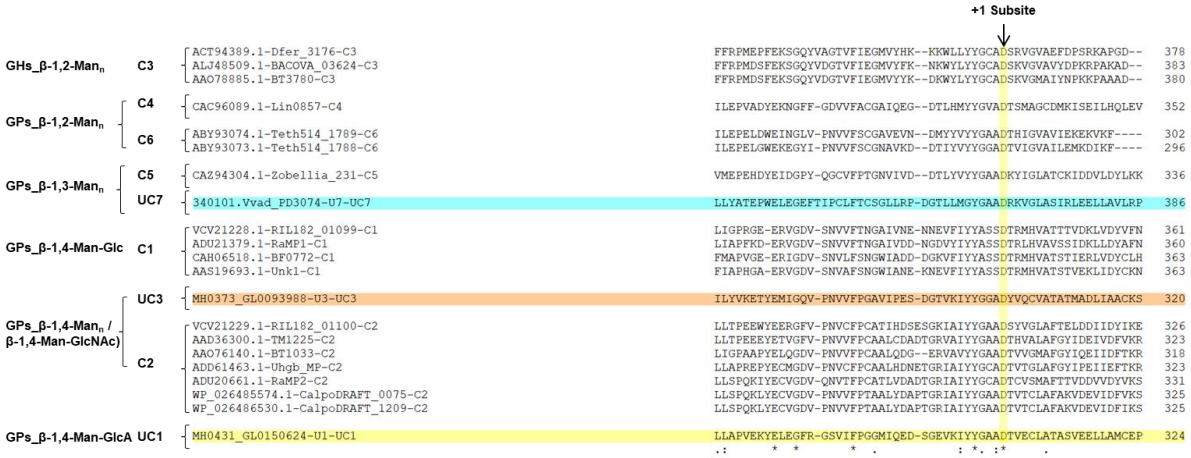


Figure 4: Multiple sequence alignment of all the characterized GH130 enzymes, including U1, U3 and U7.

Prediction of linkage specificity

In order to predict the linkage specificity of the enzymes assigned to uncharacterized meta-nodes, we studied the conservation of the residues involved in the recognition and accommodation of the acceptor at the +1 subsite (R59, Y103, R150, D304 residues, ADD61463.1 numbering, Table 4), based on the functional and structural knowledge of the previously characterized GH130 enzymes (Cuskin et al. 2015, 201; Ye et al. 2016; Tsuda et al. 2015; Ladevèze et al. 2015; Nakae et al. 2013). Firstly, the UC4 sequences do not contain any of these conserved residues. As they also lack the GH130 catalytic residues, we consider these sequences as very distant from the other GH130 enzymes, which correlates with the fact that this meta-node was the first to separate from the other large meta-nodes in the SSN constructed with an E-value threshold of 10^{-30}, corresponding to 32% sequence identity (Figure 3). This was confirmed by constructing a phylogenetic tree representing the distances between 10 sequences of each meta-node (Figure 5).

In all the characterized meta-nodes and in UC1, UC3, UC7 and UC8, residue R59 is more than 98% conserved, regardless of the linkage specificity of the enzymes. The Y103 residue is also conserved in more than 98% of the sequences of the two meta-nodes containing characterized members targeting β-1,4-mannosides (C1 and C2), but is replaced by an E in more than 99% of sequences of the three meta-nodes with members targeting β-1,2-mannosides (with a gap at this position for two sequences). An E is also observed in more than 80% of the sequences of the C5 meta-node, otherwise it is a Glutamine (Q). These isosteric Q and E residues thus may have the same function, although it is not yet proven. The residue at position Y103 thus seems to be involved in linkage specificity, although it is not its only determinant. Residue R150 is conserved in 99% of the sequences of the meta-nodes targeting β-1,4 and β-1,3 mannosides. This residue is also strictly conserved in uncharacterized meta-nodes UC1, UC6 and UC7, and in 18% of the UC3 sequences. It is replaced by a Lysine (K) in 90% of sequences of the meta-nodes targeting β-1,2-mannosides, the remaining 10%, which belong to C4, having an R. A K is also present in meta-nodes UC2, UC5, UC8 and UC9, and in 82% of UC3. It was previously proposed that this K residue confers β-1,2-linkage specificity to GH130 enzymes (Cuskin et al. 2015). Lastly, D304 is conserved in more than 99% of the sequences of all meta-nodes except UC4, and is thus not involved in linkage specificity.

Based on these results, we identified the pairs of residues corresponding to the residues Y103 and R150 in ADD61463.1 as the determinants of linkage specificity. The Y/R pair is conserved in all characterized GH130s that target β-1,4-mannosides, the E/K pair in all those that target β-1,2-mannosides, and the E/R pair in all those that target β-1,3-mannosides (+1 subsite region in Figure 4). The conservation of this pair of residues was then used to predict the linkage specificity of the uncharacterized meta-nodes. Meta-node UC1 would thus target β-1,4-linkages, meta-node UC7 β-1,3 linkages and meta-nodes UC2, UC5, UC8, and UC9 the β-1,2 linkages. In meta-node UC3, again, 81% of the sequences contain the E/K pair, and would thus target β-1,2 linkages, while those containing the E/R pair would target β-1,3 linkages and those containing the Y/R pair would target β-1,4 linkages. The prediction of linkage specificity for the UC6' sequences is not possible, since they have the conserved R150, but lack the Y103 or E, which is replaced by an Isoleucine (I) in 10 of the 11 sequences and by a Methionine (M) in the remaining one. We thus propose that they may feature a new linkage specificity in the GH130 family.

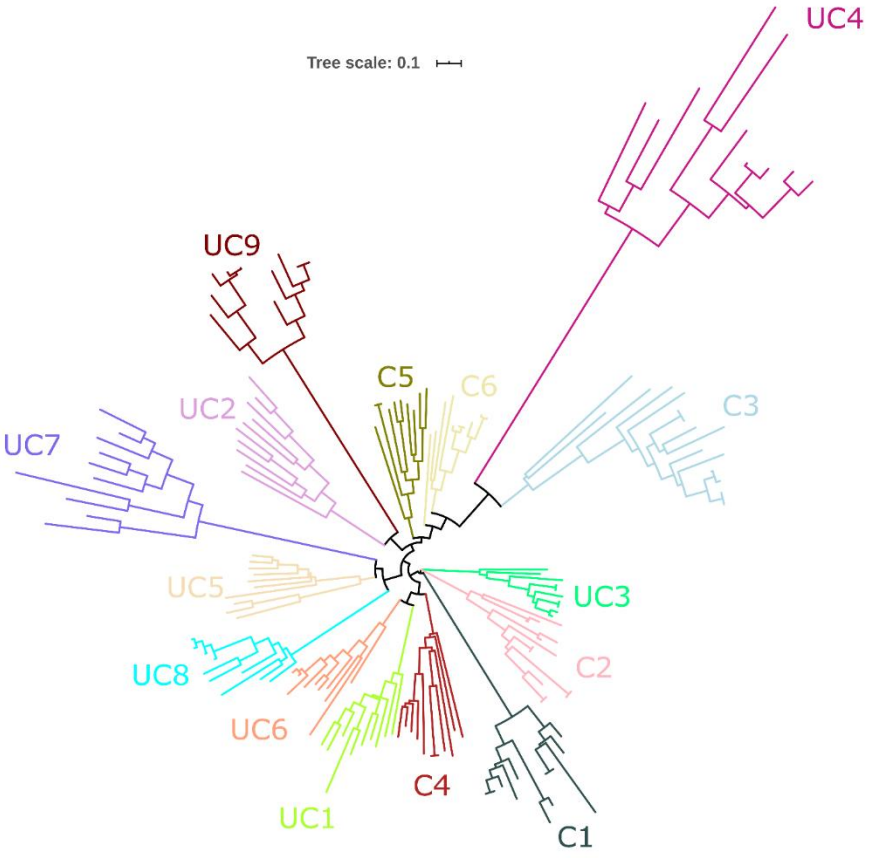


Figure 5: Maximum-likelihood phylogenetic tree, generated using 10 representative sequences for each SSN meta-node, defined with an E-value threshold of 10^{-70}.

Biochemical characterization of sequences from uncharacterized SSN meta-nodes

In order to experimentally validate our predictions regarding the mechanism and linkage specificity of the enzymes of the uncharacterized meta-nodes, and to discover potential new functions in the GH130 family, we selected one sequence to characterize in each meta-node containing at least 20 metagenomic sequences (Table 5).

Table 5: Characteristics of the sequences chosen for biochemical characterization.

Meta-node	Sequence ID	Prevalence in the human gut metagenome (% of individuals)	Best Blast hit with a GH130 sequence from the CAZy DB		Best Blast hit with a characterized GH130		Activity
			Query coverage	% Identity	Query coverage	% Identity	
UC1	MH0431_GL0150624	17 %	AVM43672.1		AAO76140.1		β-1,4-mannosyl-glucuronic acid phosphorylase
			95%	59.8%	94%	33.1%	
UC3	MH0373_GL0093988	17 %	APF19181.1		WP_026485574.1		β-1,4-mannooligosaccharide phosphorylase
			95%	43.4%	97%	39.3%	
UC4	MH0011_GL0029022	53 %	AMM53793.1		ACT94389.1		/
			69%	31.1%	57%	30.0%	
UC6	MH0409_GL0096961	4 %	ADO84614.1		VCV21229.1		/
			98%	36.6%	98%	36.6%	
UC7	340101.Vvad_PD3074	12 %	AIQ48603.1		CAZ94304.1		β-1,3-mannosyl-glucose phosphorylase β-1,3-mannooligosaccharide phosphorylase
			97%	46.7%	97%	28.6%	

We thus selected 5 human gut metagenomic sequences from meta-nodes UC1, UC3, UC4, UC6 and UC7, and named them U1, U3, U4, U6, and U7, respectively. The remaining meta-nodes UC2, UC5, UC8 and UC9 contain no or only a few metagenomic sequences (Table 6). The chosen sequences are very different from those of the 18 GH130 enzymes already characterized, with less than 40% identity (Table 5). The target genes were synthesized and expressed in *E. coli*. Unfortunately, U6 was insoluble, and could not be characterized. The U1, U3, U4 and U7 enzymes were produced in soluble form and were further characterized.

Table 6: Distribution of the predicted phosphorylases and hydrolases in the SSN meta-nodes. The number of sequences include the full-length and truncated ones.

GH130	C1	C2	C4	C5	C6	UC1	UC2	UC3	UC5	UC6	UC7	UC8	UC9	Total	C3	UC4	Total
Predicted mechanism	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GP**	**GH**	**GH**	**GH**
CAZy	222	269	290	42	85	31	199	135	61	0	9	26	29	1,398	113	0	113
Human mg	347	332	53	51	2	31	0	24	0	20	6	8	0	874	113	31	144
Mouse mg	149	227	18	5	0	1	0	0	0	0	1	2	0	403	26	2	28
Pig mg	235	244	13	13	0	24	0	3	0	1	4	0	0	537	39	5	44
Bovine mg	970	643	19	61	0	118	0	7	0	11	13	0	0	1,842	78	53	131
Total	1,923	1,715	393	172	87	205	199	169	61	32	33	36	29	5,054	369	91	460

Based on our mechanistic prediction, the members of meta-nodes UC1, UC3, and UC7 are probably phosphorylases, while the members of meta-node UC4 are expected to be hydrolases (although no signal peptide was predicted in the U4 sequence). In order to test these hypotheses, we screened for GP activity by means of an assay based on the measurement of released inorganic phosphate in the direction of glycoside synthesis, using the *E. coli* cell extracts (De Groeve et al. 2010) (Figure 1E). We first tested αMan1P as a donor, and 11 different monosaccharides as acceptors, to detect phosphate release and

identify the best acceptor for the validated GPs. Then, using the best acceptor, we assessed promiscuity towards glycosyl donors by comparing the phosphate release rates from αMan1P, αGlc1P, αGal1P, αGlcNAc1P, and βGlc1P. The release of phosphate from αMan1P was proven for the U1, U3 and U7 enzymes (Figure 6), in the presence of different acceptors. No other donor was identified for the synthesis reaction, indicating that these enzymes are strict mannoside phosphorylases. In the assay conditions, U1's best acceptor was GlcA (specifc activity 1300±40 Ug^{-1}). A slight reverse phosphorolysis activity was also detected with GalA (specifc activity 270±80 U g^{-1}). U3's best acceptor was Man (specifc activity 3700±130 U g^{-1}). It also reacted, very weakly, with GlcA (specifc activity 130±20 U g^{-1}). In contrast, U7 is a promiscuous enzyme, which is able to efciently accommodate Glc (specifc activity 46,480±4030 U g^{-1}), but also, to a lesser extent, Man (specifc activity 26700±860 U g^{-1}) and GlcNAc (specifc activity 18620±100 U g^{-1}). Without acceptor, U1, U3 and U7 also displayed a slight activity ofαMan1P hydrolysis (specifc activities 80±60, 100±50 and 440±80 U g^{-1}1, respectively), a side reaction already described for mannosidephosphorylases. No phosphate was released with the U4 target, regardless of the acceptor tested. In case this enzyme might accommodate another sugar-phosphate, we also tested it in the presence of various donors and mannose, this monosaccharide being the best acceptor for most characterized GH130 GPs. No phosphate release was detected, indicating that U4 is probably not a GP, but rather, as hypothesized, a glycosidase. We further tested the ability of the purified U4 to hydrolyze β-1,2/1,3 and 1,4-mannobiose and various *p*NP-glycosides, but no substrate consumption was observed by HPAEC-PAD analysis of the reaction media, or by the chromogenic assays performed with *p*NP-glycosides (data not shown).

To confirm the acceptor and donor specificity of U1, U3 and U7, and to determine the degree of polymerization of their products and their linkage specificity, the enzymes were purified and the glycoside synthesis reaction media were analyzed by HPAEC-PAD, and ^{1}H / ^{13}C NMR or MS, as appropriate.

For U1, a 30% decrease in αMan1P and GlcA concentrations was observed on the HPAEC-PAD chromatograms after reaction, confirming that reverse-phosphorolysis occurred (Figure 7). MS analysis was then performed in order to elucidate the structure of the glycoside produced from αMan1P and GlcA. The ESI MS measurement in positive ionization mode highlighted a species at m/z 379.09 (Figure 8_A, red spectrum). This ion corresponds to the $[M+Na]^+$ of a disaccharide composed of an hexose and an uronic acid. The structural characterization of this disaccharide was performed by MS/MS. The intracyclic fragments $^{0,2}A_1$ and $^{1,4}X_0$ confirmed that the uronic acid is at the reducing end of the disaccharide and the $^{3,5}A_2$ fragment proves that, as predicted, the linkage between the two subunits is of type 1,4 (Figure 8_B). Since all the characterized GH130 enzymes share an inverting mechanism, we deduced from these results that the compound produced by U1 from αMan1P and GlcA is β-1,4-Man-GlcA.

The HPAEC-PAD chromatogram of the U3 reaction indicated that a disaccharide and a trisaccharide were produced from αMan1P and Man (Figure 7). Their retention times correspond to those of the β-1,4-mannobiose and β-1,4-mannotriose standards. The β-1,4 linkage specificity of U3 was confirmed by ^{1}H NMR (Figure 9). The signals of the anomeric protons (H1α, H1β, H1') were assigned by comparison with signals obtained from β-1,4-mannobiose standard. This validates our prediction, U3 harboring the Y/R pair responsible for β-1,4 linkage specificity.

The U7 synthesis reactions were performed with αMan1P and Man or Glc acceptors. In both cases, one single product was observed on the HPAEC-PAD chromatograms (Figures 7B and 7C), and its retention times did not correspond to that of β-1,4-mannobiose. By comparing the anomeric region in ^{1}H NMR

with standards, we proved that β-1,3-mannobiose was produced from mannose as an acceptor. For the reaction performed in the presence of glucose as an acceptor, for which there is no available commercial standard, the ^{1}H and^{13}C NMR (1D and 2D) analyses of the reaction medium proved that U7 produced the β-1,3-Man-Glc disaccharide (Figure 9). This linkage specificity confirms our prediction for meta-node UC7' sequences. Te traces of free Man detected by ^{1}H NMR in this reaction medium result from αMan1P hydrolysis.

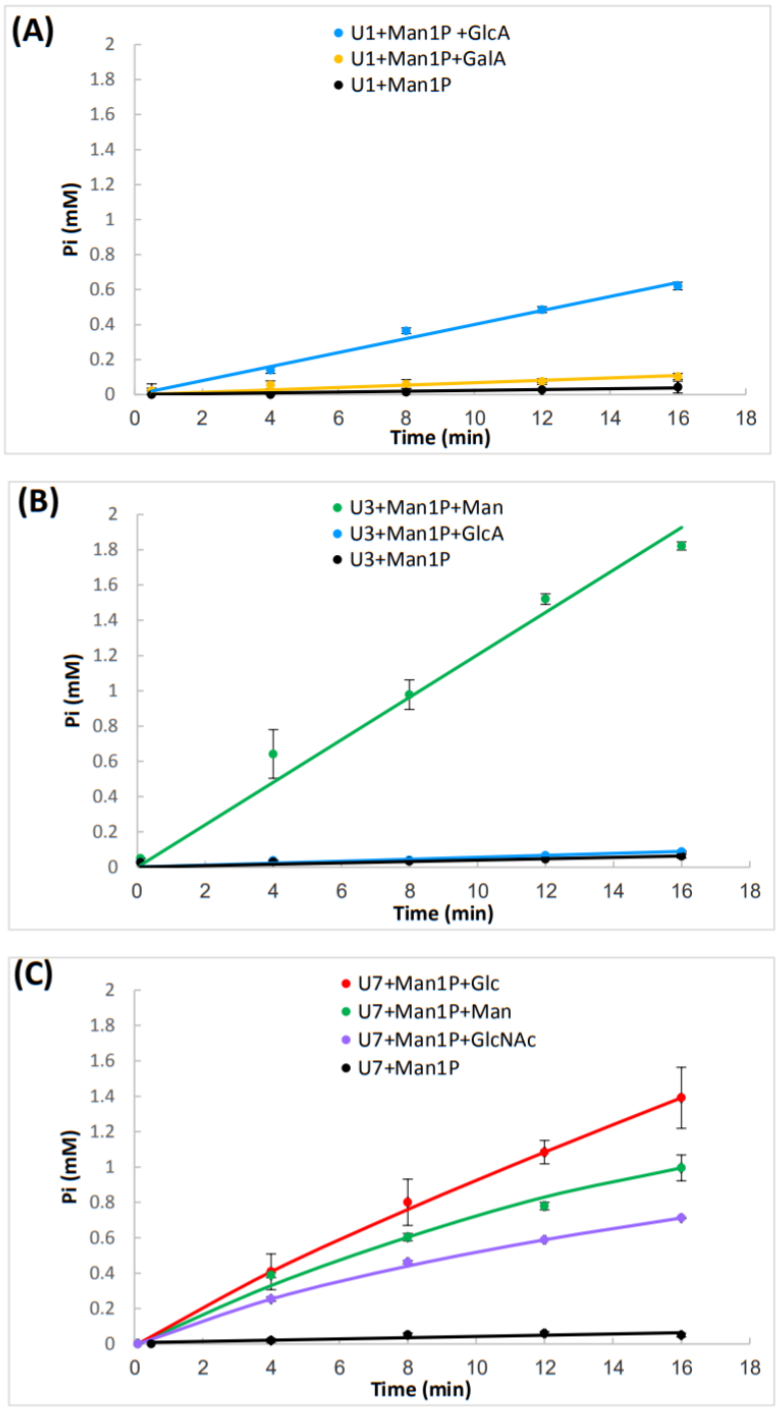


Figure 6: Determination of the acceptor specifcity of the purifed enzymes U1 (a), U3 (b) and U7 (c). The reactions were performed using 10 mM of αMan1P (glycosyl donor) and 10 mM of acceptor. The release of inorganic phosphate was followed using the molybdenum blue activity assay. Reactions were performed in triplicate. When no reverse phosphorolysis activity was detected (similar inorganic phosphate release rate to that in the presence of αMan1P as sole substrate), the data do not appear in this fgure.

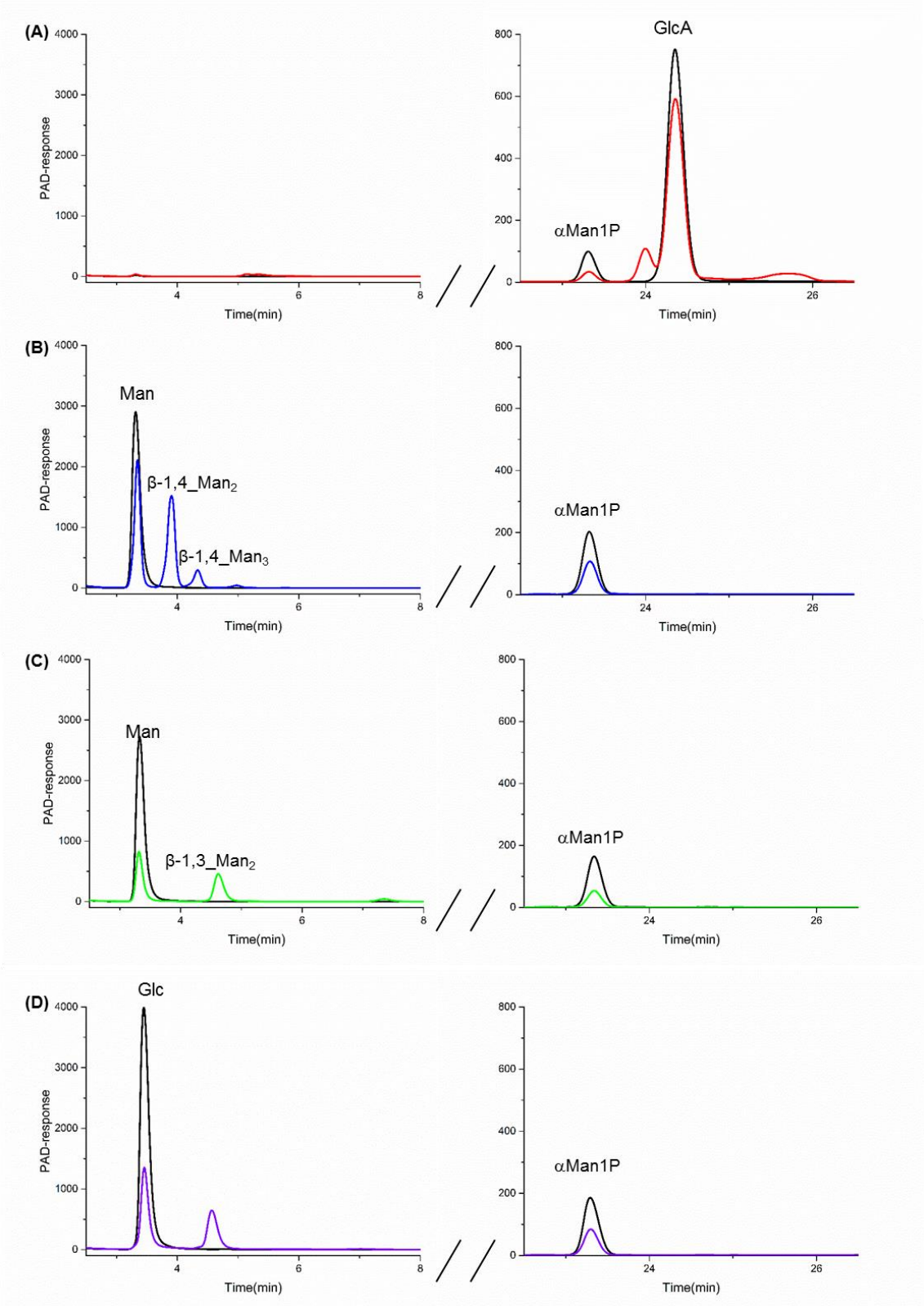


Figure 7: HPAEC-PAD analysis of the reverse-phosphorolysis reactional mixtures after incubation for 0 min (in black) and 24hrs (in color) with U1 (A), U3 (B) and U7 (C, D). The reactions were performed in the presence of 10 mM of αMan1P (glycosyl donor) and 10 mM of GlcA, Glc or Man (acceptors). Only the peaks at the retention times corresponding to commercial standards are labelled.

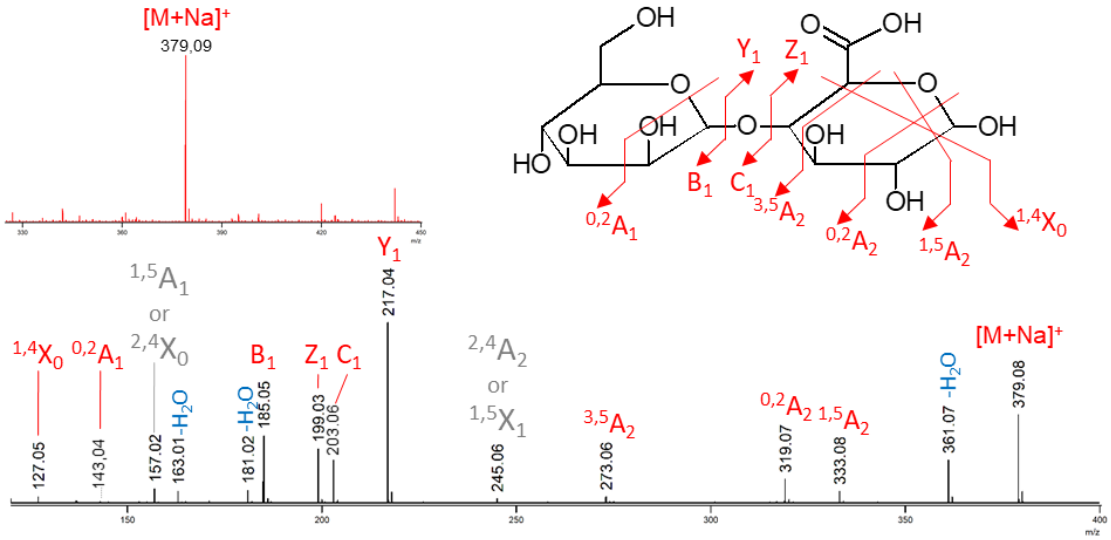


Figure 8: Mass spectrometry measurements of the U1 reverse-phosphorolysis reactional mixture (substrates: 10 mM αMan1P and 10 mM glucuronic acid), in positive ionization mode. The upper spectrum in red corresponds to the ESI MS measurement. The lower spectrum in black corresponds to the ESI MS/MS of the precursor ion isolated as a $[M+Na]^+$ at m/z 379.1. The detailed structure is shown with the specific fragments. Details of the annotation: red, unambiguous fragments; gra, ambiguous fragments (not reported on the corresponding structure); blue, water losses.

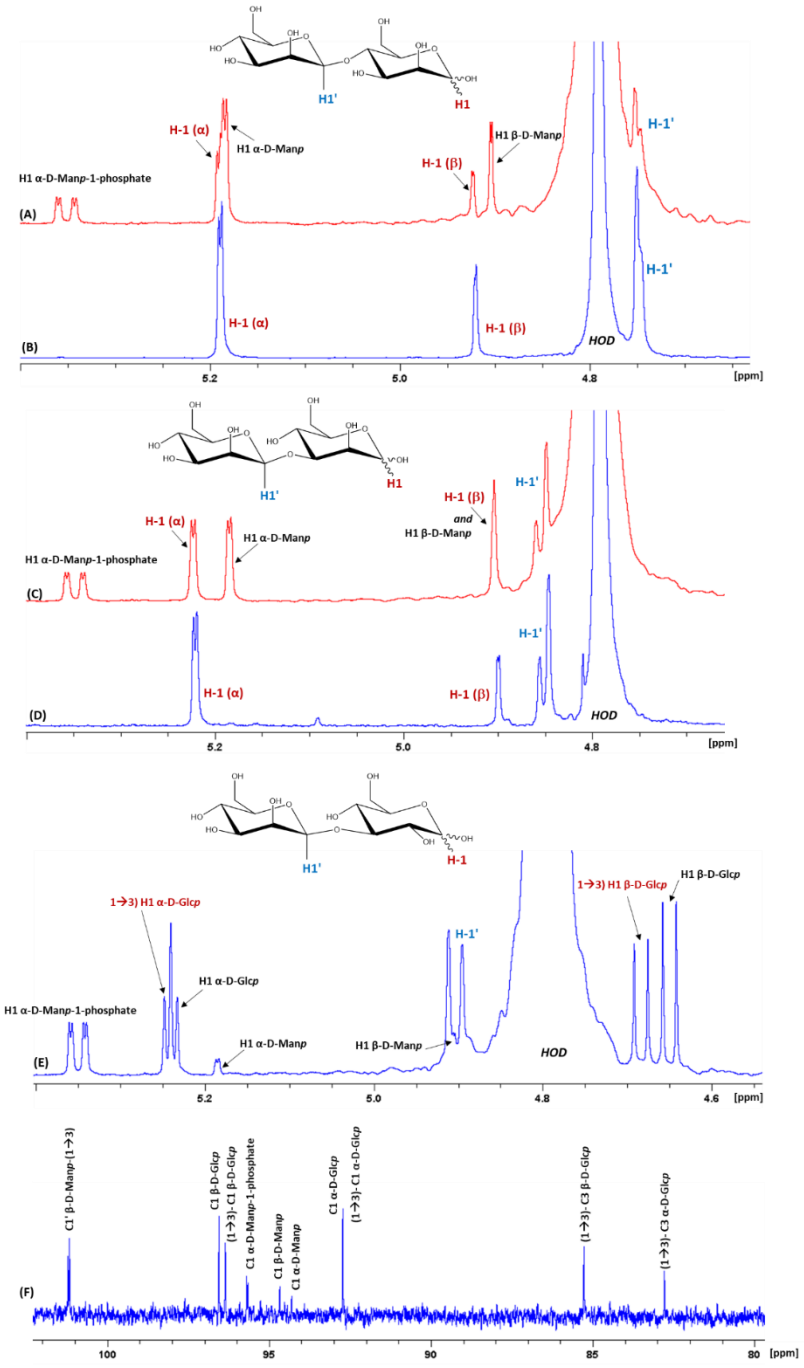


Figure 9: NMR analysis of the reverse-phosphorolysis reactional mixtures incubation for 24hrs with U7 and U3. The reactions were performed in presence of 10 mM of αMan1P (glycosyl donor) and 10 mM Glucose or Mannose (acceptors). (A, B, C, D, E) 500 MHz ^{1}H NMR spectra of the anomeric region (5.5-4.5 ppm) and (F) ^{13}C NMR spectra region (102-80 ppm). (A), U3 reaction medium: at 5.19 ppm (H1α), 4.92 ppm (H1β), 4.74 ppm (H1'). (B), β-D-mannopyranosyl-1,4-D-mannose standard (Megazyme). (C), U7 reaction medium performed with mannose as acceptor: at 5.22 ppm (H1α), 4.90 ppm (H1β), 4.85 ppm (H1'). (D), β-D-mannopyranosyl-1,3-D-mannose standard (Carbosynth). (E, F), U7 reaction medium performed with glucose as acceptor: at 5.24 ppm (H1α), 4.68 ppm (H1β), 4.90 ppm (H1') and 85.2 ppm (1→3)- C3 β-D-Glcp, 82,7 ppm (1→3)- C3 β-D-Glcp.

Discussion

Use of SSNs to analyze the functional diversity of mono-modular CAZymes

The efficiency of SSNs for rapid clustering of a large number of sequences has already been shown (Levin et al. 2017), and was recently applied to the creation of subfamilies within the large multi-functional CAZy family GH16 (Viborg et al. 2019). In the present paper, we describe the use of SSNs to:

i) analyze the functional diversity of CAZyme encoding metagenomic sequences, whose role in microbial ecosystems had never been investigated;
ii) identify new CAZyme functions.

Here, we have targeted a family which mainly contains putative glycoside phosphorylases, in order to extend our knowledge of their structure-function relationships, to investigate their physiological functions, and because they are biotechnological tools of interest for glycoside synthesis. Mannoside-phosphorylases of the family GH130 are intracellular enzymes described as targeting, *in cellulo*, only di- or trisaccharides (Ladevèze et al. 2015). The breakdown of such simple carbohydrate structures does not require complementary catalytic modules or carbohydrate binding modules, which are found to bind and depolymerize complex glycans in numerous cell surface-associated GHs. Thus, more than 99% of the sequences proven or predicted in this study to encode mannoside-phosphorylases are monomodular. This specific feature means that GH130 sequences can be directly analyzed with the EFI-EST web tool, without isolating the target catalytic domain, contrary to what has been done in respect of the often multi-modular GH16 sequences (Viborg et al. 2019).

Despite these advantages, the SSN analysis presents two limitations. Firstly, as explained in Viborg et al. 2019, it cannot be used to establish evolutionary relationships between functional meta-nodes. The formation of stable sub-families, whose boundaries will not evolve at the rate of the rapid increase of the number of genomic and metagenomic sequences, requires the use of complementary approaches, such as phylogenetic and Hidden Markov Model analyses. Secondly, as previously shown for the GH16 family and confirmed here with GH130s, SSNs do not sufficiently differentiate sequences with local structural differences, which can however have major implications in terms of enzyme specificity and mechanism. Regardless of the E-value threshold (10^{-30} to 10^{-90}) used, it was impossible to cluster GH130 sequences according to linkage specificity (β-1,2, 1,3 or 1,4), or acceptor specificity (Figures 2 and 3). We thus had to devise another strategy to predict these determinants of enzymatic function.

This generic approach could be used regardless of CAZy family. The first stage is to eliminate hydrolases by excluding the meta-nodes that contain sequences with i) signal peptides; ii) and conserved carboxylic acid residues which could act as catalytic base residues in inverting hydrolases. This second criterion can only be used for inverting families (representing 7 of the 11 GP-containing families), because, like GHs, retaining GPs involve a catalytic machinery that contains both an acid/base and a nucleophile. Another criterion that could be used for easy discrimination of hydrolases from phosphorylases could be to eliminate sequences containing CBMs, as it is indeed unlikely that intracellular enzymes acting on short oligosaccharides possess CBMs. This hypothesis was not tested here, as none of the GH130 sequences contains a CBM. The second stage aims to predict linkage specificity. For the first time, thanks to the increasing number of crystallographic structures available in the GH130 family, we have found that the linkage specificity of GH130 members, be they phosphorylases or hydrolases, is linked to the presence of a specific pair of amino acids located in the +1 subsite.

These predictions were biochemically validated with sequences belonging to three meta-nodes, and allowed us to isolate, from thousands of GH130 sequences, GPs that catalyse a reaction never described before, namely the synthesis and degradation of Man-β-1,4-GlcA by the U1 enzyme.

The reliability of this prediction remains to be validated for the members of the five meta-nodes which do not yet include any characterized GP member. Sequences of meta-node UC6 are of particular interest, since multiple sequence alignments show that the amino acids conferring specificity towards β-1,2 / 1,3 / 1,4 linkages are not conserved. These enzymes may target β-1,6-linked mannosyl residues, since heteromannoside structures containing this motif do exist in fungal polysaccharides, such as in *Lentinus enodes* (Lee et al. 2009) and in bacterial exopolysaccharides, in particular those of the enterobacterium *Edwardsiella tarda* (Guo et al. 2010) (Figure 10).

GH130 enzymes ensure diverse ecological functions, targeting mammal, plant, mold, yeast and bacterial mannosides

The SSN analysis presented here allowed us to represent the diversity of GH130 enzymes in mammalian gut microbiomes. A large number of metagenomic sequences within a SSN meta-node indicates their dissemination in numerous bacteria, and thus, that the function might play a role in habitat colonization by these species. The largest meta-nodes by far are C1 and C2, with 36% and 31% of the metagenomic sequences, respectively. These meta-nodes contain the enzymes that target plant cell wall mannans and N-glycans. This is not surprising, given the abundance of plant-derived fibers in the human, bovine, pig and mouse diet, and the fact that the gut microbes are in close contact with mammal cells harbouring N-glycans.

The meta-nodes containing the sequences shown or predicted to be involved in the degradation of β-1,2-mannosides, by either phosphorolysis or hydrolysis, are much more restricted, amounting to 8% of the metagenomic sequences. In these meta-nodes, the only characterized member from a mammal gut bacterium has been shown to target *Candida albicans* mannan (Cuskin et al. 2015). However, fungal mannosides are probably not the only targets of the enzymes from these eight meta-nodes, since several other yeast (Mille et al. 2008; Shibata et al. 2003), mould (Lee et al. 2009) and bacterial heteromannosides (Katzenellenbogen et al. 2009; Liu et al. 2014) harbour β-1,2-linked mannosyl residues (Figure 10). Furthermore, a fungal GH130 sequences is contained in meta-node UC9 (Figure 11). These fungal enzymes may be involved in mannoside recycling, via the release of αMan1P, which in turn could be transformed by pyrophosphorylases into GDP-mannose, which would further be used as GT substrate for heteromannan biosynthesis.

The UC1, UC7 and C5 meta-nodes are the final SSN meta-nodes for which we inferred a function from the analyses described here. Together, they only represent a few hundred of the sequences in mammal gut microbiomes. These meta-nodes also contain a few dozen genomic sequences from various bacterial phyla, including *Firmicutes*, *Bacteroidetes* and *Actinobacteria*, which are highly abundant in the gut microbiota (Figure 11, Table 1). The UC1 meta-node, the U1 member of which was shown in this study to target the Man-β-1,4-GlcA disaccharide, only represents 4% of our metagenomic dataset. The Man-β-1,4-GlcA motif is described in *Nicotiana plumbaginifolia* (Sims and Bacic 1995), but this tobacco plant species is not a dietary constituent for mammals. This motif also exists in bacterial lipopolysaccharides (Figure 10), in particular the O-antigen polysaccharide of *Shigella boydii* (Alberta et al. 1994), one of the four *Shigella* species considered as major enteropathogens causing childhood diarrhea worldwide (Lima, Havt, and Lima 2015). The UC7 and the C5 meta-nodes, which target β-1,3-

linked mannosides, together represent just 3% of the metagenomic dataset. These two meta-nodes each contain promiscuous GPs, of which the substrates are the Man-β-1,3-Glc and Man-β-1,3-Man disaccharides (U7 and zobellia_231 (Awad et al. 2017)). These motifs are found in some moulds (Figure 10, (Lee et al. 2009; Chen et al. 2017), and also in the lipopolysaccharides of certain pathogenic bacterial species, like *Pseudomonas aeruginosa* (Kocharova et al. 1988) (Figure 11). Thanks to their ability to produce these particular oligosaccharides through reverse phosphorolysis, the U1 and U7 GPs characterized in this study are thus tools of interest for the synthesis of antigenic oligosaccharides.

In this study, we thus revisited the functional diversity of the GH130 CAZy family by integrating that from mammalian gut microbiomes. By combining a series of *in silico* and *in vitro* approaches, we predicted that the vast majority of the GH130 enzymes are glycoside-phosphorylases and not hydrolases. We identified that some of them are specific for oligosaccharidic motifs found in bacterial lipopolysaccharides, in particular those of certain pathogenic species. Co-culture and transcriptomic studies targeting the commensals that produce these enzymes will be needed to confirm the role of mannoside-phosphorylases in bacterial interactions.

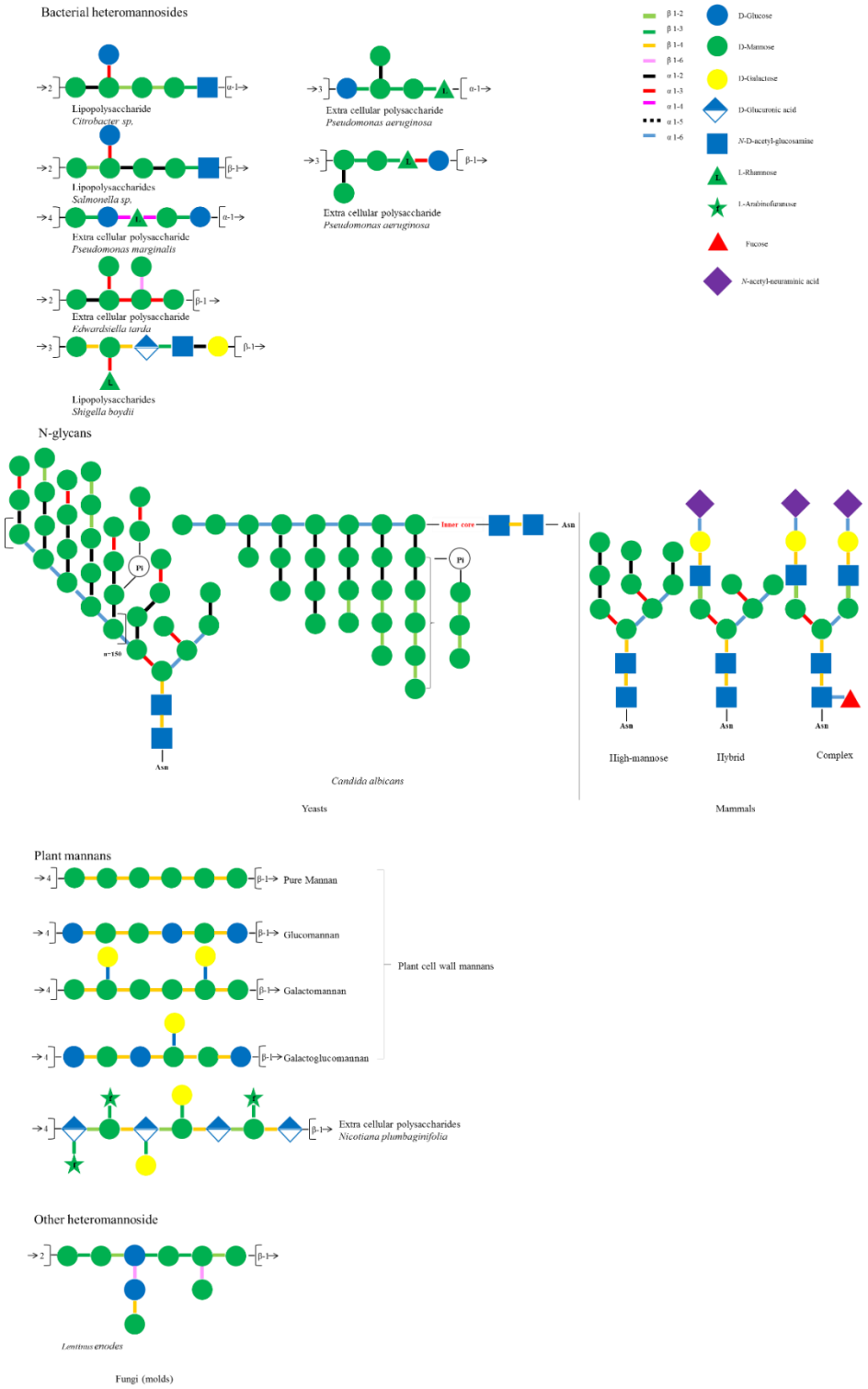


Figure 10: Examples of mannoside and heteromannoside structures containing the motifs targeted by GH130 enzymes.

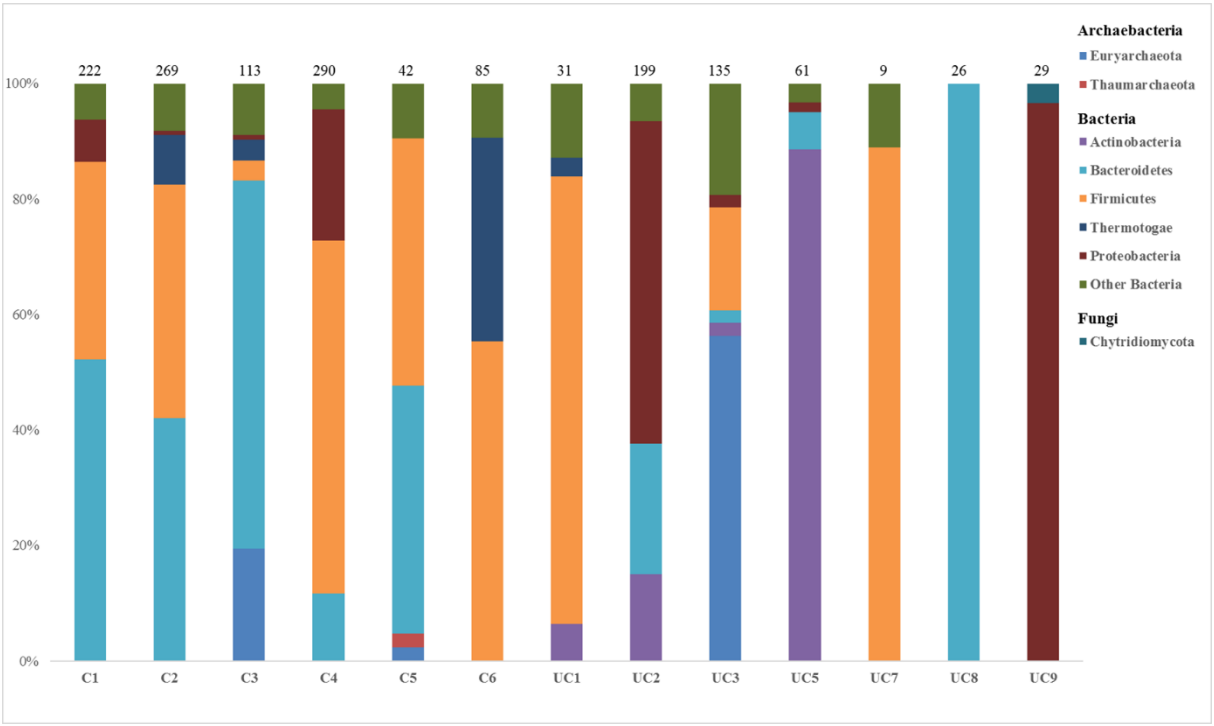


Figure 11: Taxonomical origin of the GH130 sequences listed in the CAZy database, for each SSN meta-node.

References

Alberta, M. John, Tord Holme, Bengt Lindberg, Johan Lindberg, M. Mosihuzzaman, Firdausi Qadri, and M. Mahbubur Rahman. 1994. "Structural Studies of the Shigella Boydii Type 5 O-Antigen Polysaccharide." *Carbohydrate Research* 265 (1): 121–27. https://doi.org/10.1016/0008-6215(94)00223-1.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Atkinson, Holly J., John H. Morris, Thomas E. Ferrin, and Patricia C. Babbitt. 2009. "Using Sequence Similarity Networks for Visualization of Relationships across Diverse Protein Superfamilies." *PLOS ONE* 4 (2): e4345. https://doi.org/10.1371/journal.pone.0004345.

Awad, Faisal Nureldin, Pedro Laborda, Meng Wang, Ai Min Lu, Qian Li, Zhi Peng Cai, Li Liu, and Josef Voglmeir. 2017. "Discovery and Biochemical Characterization of a Mannose Phosphorylase Catalyzing the Synthesis of Novel β-1,3-Mannosides." *Biochimica et Biophysica Acta (BBA) - General Subjects* 1861 (12): 3231–37. https://doi.org/10.1016/j.bbagen.2017.09.013.

Brett, Christopher T., and Keith W. Waldron. 1996. *Physiology and Biochemistry of Plant Cell Walls*. Springer Science & Business Media.

Chekan, Jonathan R., In Hyuk Kwon, Vinayak Agarwal, Dylan Dodd, Vanessa Revindran, Roderick I. Mackie, Isaac Cann, and Satish K. Nair. 2014. "Structural and Biochemical Basis for Mannan Utilization by Caldanaerobius Polysaccharolyticus Strain ATCC BAA-17." *Journal of Biological Chemistry* 289 (50): 34965–77. https://doi.org/10.1074/jbc.M114.579904.

Chen, Yan, Xue-Hui Li, Li-Yuan Zhou, Wan Li, Liu Liu, Dan-Dan Wang, Wen-Na Zhang, Sajid Hussain, Xiao-He Tian, and Yong-Ming Lu. 2017. "Structural Elucidation of Three Antioxidative Polysaccharides from Tricholoma Lobayense." *Carbohydrate Polymers* 157 (February): 484–92. https://doi.org/10.1016/j.carbpol.2016.10.011.

Chiku, Kazuhiro, Takanori Nihira, Erika Suzuki, Mamoru Nishimoto, Motomitsu Kitaoka, Ken'ichi Ohtsubo, and Hiroyuki Nakai. 2014. "Discovery of Two β-1,2-Mannoside Phosphorylases Showing Different Chain-Length Specificities from Thermoanaerobacter Sp. X-514." *PLOS ONE* 9 (12): e114882. https://doi.org/10.1371/journal.pone.0114882.

Cuskin, Fiona, Arnaud Baslé, Simon Ladevèze, Alison M. Day, Harry J. Gilbert, Gideon J. Davies, Gabrielle Potocki-Véronèse, and Elisabeth C. Lowe. 2015. "The GH130 Family of Mannoside Phosphorylases Contains Glycoside Hydrolases That Target β-1,2-Mannosidic Linkages in Candida Mannan." *Journal of Biological Chemistry* 290 (41): 25023–33. https://doi.org/10.1074/jbc.M115.681460.

De Groeve, M. R. M., G. H. Tran, A. Van Hoorebeke, J. Stout, T. Desmet, S. N. Savvides, and W. Soetaert. 2010. "Development and Application of a Screening Assay for Glycoside Phosphorylases." *Analytical Biochemistry* 401 (1): 162–67. https://doi.org/10.1016/j.ab.2010.02.028.

Domon, Bruno, and Catherine E. Costello. 1988. "A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates." *Glycoconjugate Journal* 5 (4): 397–409. https://doi.org/10.1007/BF01049915.

Gawronski, Jeffrey D., and David R. Benson. 2004. "Microtiter Assay for Glutamine Synthetase Biosynthetic Activity Using Inorganic Phosphate Detection." *Analytical Biochemistry* 327 (1): 114–18. https://doi.org/10.1016/j.ab.2003.12.024.

Gerlt, John A., Jason T. Bouvier, Daniel B. Davidson, Heidi J. Imker, Boris Sadkhin, David R. Slater, and Katie L. Whalen. 2015. "Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1854 (8): 1019–37. https://doi.org/10.1016/j.bbapap.2015.04.015.

Grimaud, Florent, Sandra Pizzut-Serin, Laurence Tarquis, Simon Ladevèze, Sandrine Morel, Jean-Luc Putaux, and Gabrielle Potocki-Veronese. 2019. "In Vitro Synthesis and Crystallization of β-1,4-Mannan." *Biomacromolecules* 20 (2): 846–53. https://doi.org/10.1021/acs.biomac.8b01457.

Guo, Shoudong, Wenjun Mao, Yin Han, Xiaohua Zhang, Chunli Yang, Yin Chen, Yanli Chen, et al. 2010. "Structural Characteristics and Antioxidant Activities of the Extracellular Polysaccharides Produced by Marine Bacterium Edwardsiella Tarda." *Bioresource Technology* 101 (12): 4729–32. https://doi.org/10.1016/j.biortech.2010.01.125.

Huang, Ying, Beifang Niu, Ying Gao, Limin Fu, and Weizhong Li. 2010. "CD-HIT Suite: A Web Server for Clustering and Comparing Biological Sequences." *Bioinformatics* 26 (5): 680–82. https://doi.org/10.1093/bioinformatics/btq003.

Katoh, Kazutaka, and Daron M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30 (4): 772–80. https://doi.org/10.1093/molbev/mst010.

Katzenellenbogen, Ewa, Nina A. Kocharova, Philip V. Toukach, Sabina Górska, Agnieszka Korzeniowska-Kowal, Maria Bogulska, Andrzej Gamian, and Yuriy A. Knirel. 2009. "Structure of an Abequose-Containing O-Polysaccharide from Citrobacter Freundii O22 Strain PCM 1555." *Carbohydrate Research* 344 (13): 1724–28. https://doi.org/10.1016/j.carres.2009.06.005.

Kawahara, Ryosuke, Wataru Saburi, Rei Odaka, Hidenori Taguchi, Shigeaki Ito, Haruhide Mori, and Hirokazu Matsui. 2012. "Metabolic Mechanism of Mannan in a Ruminal Bacterium, Ruminococcus Albus, Involving Two Mannoside Phosphorylases and Cellobiose 2-Epimerase: Discovery of a New Carbohydrate Phosphorylase, β-1,4-Mannooligosaccharide Phosphorylase." *Journal of Biological Chemistry* 287 (50): 42389–99. https://doi.org/10.1074/jbc.M112.390336.

Kocharova, Nina A., Yuri A. Knirel, Alexander S. Shashkov, Nikola K. Kochetkov, and Gerald B. Pier. 1988. "Structure of an Extracellular Cross-Reactive Polysaccharide from Pseudomonas Aeruginosa Immunotype 4." *Journal of Biological Chemistry* 263 (23): 11291–95.

Konagurthu, Arun S., James C. Whisstock, Peter J. Stuckey, and Arthur M. Lesk. 2006. "MUSTANG: A Multiple Structural Alignment Algorithm." *Proteins: Structure, Function, and Bioinformatics* 64 (3): 559–74. https://doi.org/10.1002/prot.20921.

La Rosa, Sabina Leanti, Maria Louise Leth, Leszek Michalak, Morten Ejby Hansen, Nicholas A. Pudlo, Robert Glowacki, Gabriel Pereira, et al. 2019. "The Human Gut Firmicute Roseburia Intestinalis

Is a Primary Degrader of Dietary β-Mannans." *Nature Communications* 10 (1): 905. https://doi.org/10.1038/s41467-019-08812-y.

Ladevèze, Simon, Gianluca Cioci, Pierre Roblin, Lionel Mourey, Samuel Tranier, and Gabrielle Potocki-Véronèse. 2015. "Structural Bases for N-Glycan Processing by Mannoside Phosphorylase." *Acta Crystallographica Section D: Biological Crystallography* 71 (6): 1335–46. https://doi.org/10.1107/S1399004715006604.

Ladevèze, Simon, Elisabeth Laville, Jordane Despres, Pascale Mosoni, and Gabrielle Potocki-Véronèse. 2017. "Mannoside Recognition and Degradation by Bacteria." *Biological Reviews* 92 (4): 1969–90. https://doi.org/10.1111/brv.12316.

Ladevèze, Simon, Laurence Tarquis, Davide A. Cecchini, Juliette Bercovici, Isabelle André, Christopher M. Topham, Sandrine Morel, et al. 2013. "Role of Glycoside Phosphorylases in Mannose Foraging by Human Gut Bacteria." *Journal of Biological Chemistry* 288 (45): 32370–83. https://doi.org/10.1074/jbc.M113.483628.

Lee, Hee Hwan, Jong Seok Lee, Jae Youl Cho, Jae Yeol Cho, Young Eon Kim, and Eock Kee Hong. 2009. "Structural Characteristics of Immunostimulating Polysaccharides from Lentinus Edodes." *Journal of Microbiology and Biotechnology* 19 (5): 455–61. https://doi.org/10.4014/jmb.0809.542.

Letunic, Ivica, and Peer Bork. 2016. "Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees." *Nucleic Acids Research* 44 (W1): W242–45. https://doi.org/10.1093/nar/gkw290.

Levin, B. J., Y. Y. Huang, S. C. Peck, Y. Wei, A. Martínez-del Campo, J. A. Marks, E. A. Franzosa, C. Huttenhower, and E. P. Balskus. 2017. "A Prominent Glycyl Radical Enzyme in Human Gut Microbiomes Metabolizes Trans-4-Hydroxy-l-Proline." *Science* 355 (6325): eaai8386. https://doi.org/10.1126/science.aai8386.

Li, Junhua, MetaHIT Consortium, Huijue Jia, Xianghang Cai, Huanzi Zhong, Qiang Feng, Shinichi Sunagawa, et al. 2014. "An Integrated Catalog of Reference Genes in the Human Gut Microbiome." *Nature Biotechnology* 32 (8): 834–41. https://doi.org/10.1038/nbt.2942.

Li, Junhua, Huanzi Zhong, Yuliaxis Ramayo-Caldas, Nicolas Terrapon, Vincent Lombard, Gabrielle Potocki-Veronese, Jordi Estellé, et al. 2020. "A Catalog of Microbial Genes from the Bovine Rumen Unveils a Specialized and Diverse Biomass-Degrading Environment." *GigaScience Database*, May. https://doi.org/10.1101/272690.

Lima, Ila F. N., Alexandre Havt, and Aldo A. M. Lima. 2015. "Update on Molecular Epidemiology of Shigella Infection." *Current Opinion in Gastroenterology* 31 (1): 30. https://doi.org/10.1097/MOG.0000000000000136.

Liu, Bin, Yuriy A. Knirel, Lu Feng, Andrei V. Perepelov, Sof'ya N. Senchenkova, Peter R. Reeves, and Lei Wang. 2014. "Structural Diversity in SalmonellaO Antigens and Its Genetic Basis." *FEMS Microbiology Reviews* 38 (1): 56–89. https://doi.org/10.1111/1574-6976.12034.

Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. 2014. "The Carbohydrate-Active Enzymes Database (CAZy) in 2013." *Nucleic Acids Research* 42 (D1): D490–95. https://doi.org/10.1093/nar/gkt1178.

Macdonald, Spencer S., Zachary Armstrong, Connor Morgan-Lang, Magdalena Osowiecka, Kyle Robinson, Steven J. Hallam, and Stephen G. Withers. 2019. "Development and Application of a High-Throughput Functional Metagenomic Screen for Glycoside Phosphorylases." *Cell Chemical Biology* 26 (7): 1001-1012.e5. https://doi.org/10.1016/j.chembiol.2019.03.017.

Mille, Céline, Piotr Bobrowicz, Pierre-André Trinel, Huijuan Li, Emmanuel Maes, Yann Guerardel, Chantal Fradin, et al. 2008. "Identification of a New Family of Genes Involved in β-1,2-Mannosylation of Glycans in Pichia Pastoris and Candida Albicans." *Journal of Biological Chemistry* 283 (15): 9724–36. https://doi.org/10.1074/jbc.M708825200.

Nakae, Setsu, Shigeaki Ito, Mariko Higa, Takeshi Senoura, Jun Wasaki, Atsushi Hijikata, Masafumi Shionyu, Susumu Ito, and Tsuyoshi Shirai. 2013. "Structure of Novel Enzyme in Mannan Biodegradation Process 4-o-β-d-Mannosyl-d-Glucose Phosphorylase MGP." *Journal of Molecular Biology* 425 (22): 4468–78. https://doi.org/10.1016/j.jmb.2013.08.002.

Niedermeyer, Timo H. J., and Martin Strohalm. 2012. "MMass as a Software Tool for the Annotation of Cyclic Peptide Tandem Mass Spectra." Edited by John Matthew Koomen. *PLoS ONE* 7 (9): e44913. https://doi.org/10.1371/journal.pone.0044913.

Nielsen, Henrik. 2017. “Predicting Secretory Proteins with SignalP.” In *Protein Function Prediction: Methods and Protocols*, edited by Daisuke Kihara, 59–73. Methods in Molecular Biology. https://doi.org/10.1007/978-1-4939-7015-5_6.

Nihira, Takanori, Kazuhiro Chiku, Erika Suzuki, Mamoru Nishimoto, Shinya Fushinobu, Motomitsu Kitaoka, Ken’ichi Ohtsubo, and Hiroyuki Nakai. 2015. “An Inverting β-1,2-Mannosidase Belonging to Glycoside Hydrolase Family 130 from Dyadobacter Fermentans.” *FEBS Letters* 589 (23): 3604–10. https://doi.org/10.1016/j.febslet.2015.10.008.

Nihira, Takanori, Erika Suzuki, Motomitsu Kitaoka, Mamoru Nishimoto, Ken’ichi Ohtsubo, and Hiroyuki Nakai. 2013. “Discovery of β-1,4-D-Mannosyl-N-Acetyl-D-Glucosamine Phosphorylase Involving in the Metabolism of N-Glycans.” *Journal of Biological Chemistry* 288 (38): 27366–74. https://doi.org/10.1074/jbc.M113.469080.

Saitou, Naruya, and Masatoshi Nei. 1987. “The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees.” *Molecular Biology and Evolution* 4 (4): 406–25. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Senoura, Takeshi, Shigeaki Ito, Hidenori Taguchi, Mariko Higa, Shigeki Hamada, Hirokazu Matsui, Tadahiro Ozawa, et al. 2011. “New Microbial Mannan Catabolic Pathway That Involves a Novel Mannosylglucose Phosphorylase.” *Biochemical and Biophysical Research Communications* 408 (4): 701–6. https://doi.org/10.1016/j.bbrc.2011.04.095.

Sernee, M. Fleur, Julie E. Ralton, Tracy L. Nero, Lukasz F. Sobala, Joachim Kloehn, Marcel A. Vieira-Lara, Simon A. Cobbold, et al. 2019. “A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in Leishmania Parasites.” *Cell Host & Microbe* 26 (3): 385-399.e9. https://doi.org/10.1016/j.chom.2019.08.009.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” *Genome Research* 13 (11): 2498–2504. https://doi.org/10.1101/gr.1239303.

Sharma, Vandana, Mie Ichikawa, and Hudson H. Freeze. 2014. “Mannose Metabolism: More than Meets the Eye.” *Biochemical and Biophysical Research Communications*, Integrative Glycobiology and Future Perspectives, 453 (2): 220–28. https://doi.org/10.1016/j.bbrc.2014.06.021.

Shibata, Nobuyuki, Hidemitsu Kobayashi, Yoshio Okawa, and Shigeo Suzuki. 2003. “Existence of Novel β-1,2 Linkage-Containing Side Chain in the Mannan of Candida Lusitaniae, Antigenically Related to Candida Albicans Serotype A.” *European Journal of Biochemistry* 270 (12): 2565–75. https://doi.org/10.1046/j.1432-1033.2003.03622.x.

Sims, Ian M., and Antony Bacic. 1995. “Extracellular Polysaccharides from Suspension Cultures of Nicotiana Plumbaginifolia.” *Phytochemistry* 38 (6): 1397–1405. https://doi.org/10.1016/0031-9422(94)00832-E.

Svartström, Olov, Johannes Alneberg, Nicolas Terrapon, Vincent Lombard, Ino de Bruijn, Jonas Malmsten, Ann-Marie Dalin, et al. 2017. “Ninety-Nine de Novo Assembled Genomes from the Moose (Alces Alces) Rumen Microbiome Provide New Insights into Microbial Plant Biomass Degradation.” *The ISME Journal* 11 (11): 2538–51. https://doi.org/10.1038/ismej.2017.108.

Tang, Shu-Lun, and Nicola L. B. Pohl. 2015. “Automated Solution-Phase Synthesis of β-1,4-Mannuronate and β-1,4-Mannan.” *Organic Letters* 17 (11): 2642–45. https://doi.org/10.1021/acs.orglett.5b01013.

Tsuda, Tomohiro, Takanori Nihira, Kazuhiro Chiku, Erika Suzuki, Takatoshi Arakawa, Mamoru Nishimoto, Motomitsu Kitaoka, Hiroyuki Nakai, and Shinya Fushinobu. 2015. “Characterization and Crystal Structure Determination of β-1,2-Mannobiose Phosphorylase from Listeria Innocua.” *FEBS Letters* 589 (24PartB): 3816–21. https://doi.org/10.1016/j.febslet.2015.11.034.

Viborg, Alexander Holm, Nicolas Terrapon, Vincent Lombard, Gurvan Michel, Mirjam Czjzek, Bernard Henrissat, and Harry Brumer. 2019. “A Subfamily Roadmap for Functional Glycogenomics of the Evolutionarily Diverse Glycoside Hydrolase Family 16 (GH16).” *Journal of Biological Chemistry*, September, jbc.RA119.010619. https://doi.org/10.1074/jbc.RA119.010619.

Wildberger, Patricia, Anamaria Todea, and Bernd Nidetzky. 2012. “Probing Enzyme–Substrate Interactions at the Catalytic Subsite of Leuconostoc Mesenteroides Sucrose Phosphorylase with Site-Directed Mutagenesis: The Roles of Asp49 and Arg395.” *Biocatalysis and Biotransformation* 30 (3): 326–37. https://doi.org/10.3109/10242422.2012.674720.

Xiao, Liang, Jordi Estellé, Pia Kiilerich, Yuliaxis Ramayo-Caldas, Zhongkui Xia, Qiang Feng, Suisha Liang, et al. 2016. “A Reference Gene Catalogue of the Pig Gut Microbiome.” *Nature Microbiology* 1 (12): 1–6. https://doi.org/10.1038/nmicrobiol.2016.161.

Xiao, Liang, Qiang Feng, Suisha Liang, Si Brask Sonne, Zhongkui Xia, Xinmin Qiu, Xiaoping Li, et al. 2015. “A Catalog of the Mouse Gut Metagenome.” *Nature Biotechnology* 33 (10): 1103–8. https://doi.org/10.1038/nbt.3353.

Ye, Yuxin, Wataru Saburi, Rei Odaka, Koji Kato, Naofumi Sakurai, Keisuke Komoda, Mamoru Nishimoto, Motomitsu Kitaoka, Haruhide Mori, and Min Yao. 2016. “Structural Insights into the Difference in Substrate Recognition of Two Mannoside Phosphorylases from Two GH130 Subfamilies.” *FEBS Letters* 590 (6): 828–37. https://doi.org/10.1002/1873-3468.12105.

Zeleznick, L. D., S. M. Rosen, M. Saltmarsh-Andrew, M. J. Osborn, and B. L. Horecker. 1965. “Biosynthesis of Bacterial Lipopolysaccharide, Iv. Enzymatic Incorporation of Mannose, Rhamnose, and Galactose in a Mutant Strain of Salmonella Typhimurium.” *Proceedings of the National Academy of Sciences of the United States of America* 53 (1): 207–14.

Zhang, Kai, and Stephen M. Beverley. 2019. “Mannogen-Ing Central Carbon Metabolism by Leishmania.” *Trends in Parasitology* 35 (12): 947–49. https://doi.org/10.1016/j.pt.2019.10.001.

Chapter 2

In the previous chapter, I described the analysis of GH130 sequence diversity in mammalian gut microbiomes, with a specific focus on uncharacterized sequence clusters. In order to choose the target to characterize, we analyzed the prevalence and abundance of all the GH130 sequences identified in the human gut metagenome. I identified several highly prevalent sequences in the microbiome, whatever are the geographical origins and health status of the individuals, and characterized the function of three of them. By performing this analysis, I also highlighted GH130 sequences which are strikingly overprevalent and overabundant in the microbiome of patients suffering from inflammatory bowel diseases. In this chapter, I describe this analysis and the biochemical characterization of one of the corresponding enzymes, that we named IBD_P1. In this study, Elisabeth Laville provided me with the sequence abundance and prevalence data. I performed the analysis of these data to highlight the sequences of interest, and predicted the enzyme mechanisms. Elisabeth and I analyzed the sequence similarities, taxonomical origin, and genomic context. I produced and purified the IBD_P1 enzyme and performed its biochemical characterization, based on chromogenic assays and HPAEC-PAD analysis of the reaction products. David Guieysse (TBI) performed the NMR analysis. Simon Ollivier and David Ropartz (BIBS platform, INRAE Nantes) determined the structure of one of the trisaccharides synthesized by this enzyme. The experimental procedure and each step of this highly challenging ion mobility-mass spectrometry analysis will not be detailed in this chapter. The results presented here will be coupled with the detailed description of the mannooligosaccharide structural determination, as one single publication which will be submitted once we will have performed the few validation experiments that are still missing.

A dual glycoside-phosphorylase/transmannosylase enzyme from an uncultured gut bacteria, associated with inflammatory bowel diseases

Introduction

Mannose-containing glycosides are found in many living organisms. They are involved in multiple biological functions, such as structuration of plant cell walls (as hemicellulose components), cell signaling (as component of proteoglycans, lipopolysaccharides and capsular polysaccharides of the external surface of cellular membranes) (Brett and Waldron 1996; Sharma, Ichikawa, and Freeze 2014) and, in some parasites, as energy storage polymer in the form of mannogen (Sernee et al. 2019). Mannosides are thus ubiquitous sources of energy for the organisms that store them, and for those which feed on them. To harvest them, bacteria produce a large diversity of CAZymes, mainly glycoside hydrolases (GHs). β-mannanases and β-mannosidases are found in the families GH1, GH2, GH5, GH26, GH113, GH130 and GH134 of the CAZy classification (Lombard et al. 2014), while α-mannanases and α-mannosidases are found in the GH31, GH38, GH47, GH63, GH76, GH92, GH99 and GH125 families. Bacteria can also breakdown β-mannosides by phosphorolysis, involving glycoside-phosphorylases (GPs) of the GH130 family. The family 108 of glycosyltransferases also contains enzymes acting on mannogen, a linear β-1,2-mannoside, by combining phosphorolysis and GDP-mannose dependent mannosyltransferase activities (Sernee et al. 2019). Nevertheless, no bacterial enzyme has yet been characterized in this family.

In presence of inorganic phosphate (Pi), β-mannoside-phosphorylases breakdown their substrates with an inverting mechanism, to generate α-D-mannose-1-phosphate (αMan1P) (Sernee et al. 2019; Senoura et al. 2011) and a β-mannoside of reduced length. To date, 21 bacterial enzymes of the GH130 family have been characterized, of which 3 are β-1,2-mannosidases (Cuskin et al. 2015; Nihira et al. 2015) and 18 are GPs acting on a large variety of substrates, such as β-1,4-mannooligosaccharides (Chekan et al. 2014; Grimaud et al. 2019; Kawahara et al. 2012; La Rosa et al. 2019; A. Li et al. 2020; Ye et al. 2016), β-1,4-mannosyl-N-acetyl-glucosamine and β-1,4-mannosyl-N,N'-diacetylchitobiose (Ladevèze et al. 2013; Nihira et al. 2013), β-1,2-mannobiose and β-1,2-oligomannans (Chiku et al. 2014), β-1,3-mannobiose or β-1,3-Mannosyl-Glucose (A. Li et al. 2020; Tsuda et al. 2015), and β-1,4-mannosyl-glucuronicacid (A. Li et al. 2020). *In vitro*, mannoside-phosphorylases also have the ability to synthesize β-mannosides, by reverse-phosphorolysis. This reaction corresponds to the transfer of a mannosyl residue from the αMan1P donor to a glycosyl acceptor, by releasing Pi (Ye et al. 2016; Nihira et al. 2013; Ladevèze et al. 2013; Chiku et al. 2014; Tsuda et al. 2015, 201; Jaito et al. 2014; Awad et al. 2017). GP-based β-mannoside synthesis processes are attractive compared to those based on mannosyl-transferases, which use expensive mannose diphosphonucleotides as donnors (Revers et al. 1999; Zhao and Thorson 1999), or on transmannosylation catalyzed by the rare native or engineered mannosidases or mannanases (Dilokpimol et al. 2011; Eneyskaya et al. 2009; Ishimizu et al. 2004; Murata and Usui 1997; Sasaki, Ishimizu, and Hase 2005; Usui et al. 1994).

As we recently showed by analyzing their sequence diversity, GH130 enzymes are widespread in bacteria residing in the gut of mammals, where there can feed on diverse mannosides of plant, mammal and microbial origins (A. Li et al. 2020). In the human gut microbiome, the prevalence and abundance of GH130 sequences is variable (Ladevèze et al. 2013; A. Li et al. 2020), but some of them are present in more than 50 % of individuals, indicating that they probably play a critical role in mannose foraging in the gut. Interestingly too, the abundance and prevalence of some sequences in the microbiome is significantly higher in patients suffering from inflammatory bowel diseases (IBD, including Crohn desease and ulcerative colitis) than in healthy individuals (Ladevèze et al. 2013). However, to date, only one of these IBD-associated enzymes has been biochemically characterized (the Uhgb_MP enzyme, (Ladevèze et al. 2013), revealing its specificity towards the host N-glycans that line the intestinal epithelium.

In this chapter, we present the biochemical characterization of a mannoside-acting enzyme of which the sequence is a striking biomarker for IBD. Its ability to degrade, but also to synthesize mannosides, in particular β-1,2-linked ones, will be highlighted, and its ecological role in the human gut microbiota will be discussed with regard to IBD-related dysbiosis.

Material and Methods

Sequence analysis

The IBD_P1 sequence is available in the database of the Integrated reference catalog of the human gut microbiome (J. Li et al. 2014), under gene identifier MH0105_GL0062872, locus scaffold61381_1:8141:9193 (http://meta.genomics.cn/meta/dataTools). This sequence was retrieved from the human gut metagenomic gene catalog, according to the procedure described in Li et al (A. Li et al. 2020) for the prediction of GH130 sequences. Signal peptide was searched using the SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen 2017). The sequence abundance and prevalence in the human gut metagenome were retrieved from the 9.9 million gene abundance table in 760 European individuals (http://meta.genomics.cn/meta/dataTools).

Chemicals

The IBD_P1 specificity towards acceptors was tested in reverse-phosphorolysis using i) different acceptors: D-glucose (Glc) and D-fructose (Fru) (VWR Chemicals BDH), D-mannose (Man), D-galactose (Gal), N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-galactosamine (GalNAc), L-arabinose (Ara), L-rhamnose (Rha), D-xylose (Xyl), D-glucuronic acid (GlcA) (Sigma Aldrich), D-glucosamine (GlcN), N-acetyl-D-mannosamine (ManNAc) and D-galacturonic acid (GalA) (Carbosynth), and ii) donors: α-D-glucose-1-phosphate (αGlc1P), α-D-mannose-1-phosphate (αMan1P), α-D-galactose-1-phosphate (αGal1P), α-D-N-acetyl-glucosamine-1-phosphate (αGlcNAc1P) (Sigma Aldrich); β-D-glucose-1-phosphate (βGlc1P) (Carbosynth). Sodium molybdate and L-ascorbic acid were purchased from Sigma Aldrich. β-1,2-mannobiose and β-1,2-mannotriose used as standards and substrates to assess the phosphorolytic and hydrolysis activities were obtained by purification of the products of the reverse-synthesis reaction performed from 10 mM αM1P and 10 mM Man by the mannoside-phosphorylase Teth514-1788, according to the protocol described in Chiku et al.(Chiku et al. 2014). α-1,2-mannobiose was purchased from Carbosynth.

Recombinant enzyme production and purification

The gene encoding IBD-P1 was synthetized by Biomatik Limited (Cambridge, Ontario, Canada), with optimization of codon usage for expression in *E. coli*. The IBD-P1 encoding gene was cloned in the pET-23a (+) vector, in fusion with N- and C-terminal (His) 6 tags, and expressed in the *E. coli* BL21-DE3 star strain. The recombinant *E. coli* cells were cultured overnight at 37°C in ZYM5052 auto-induction medium, supplemented with100 µg/ml ampicillin. To prepare the cytoplasmic extracts used for activity screening, the cells were harvested by centrifugation for five minutes at 5,000 rpm, before being re-suspended and concentrated in 20 mM Tris-HCl buffer (pH 7.0) to obtain a final OD_{600nm} of 80. Cell lysis was carried out using sonication. Cell debris were centrifuged at 13,000 rpm for 10 minutes, and cytoplasmic extracts were filtered using a 0.20 µm Minisart RC4 syringe filter.

To purify the enzyme, the recombinant cells were harvested, re-suspended in TALON buffer (20 mM Tris-HCl, pH 7.0, 300 mM NaCl) and disrupted by sonication. The supernatant was collected by centrifugation and subsequently loaded on a TALON resin loaded with cobalt (GE Healthcare) and previously washed with 20 mM Tris-HCl, pH 7.0, 300 mM NaCl. After washing with 20 mM Tris-HCl, pH 7.0, 300 mM NaCl, 20 mM imidazole, the purified protein was eluted by 20 mM Tris-HCl, pH 7.0, 300 mM NaCl, and 200 mM imidazole. At last, the purified protein was desalted using a PD-10 column (GE Healthcare) to eliminate imidazole, and change the buffer to 20 mM Tris-HCl, pH 7.0. The protein concentration was determined by spectrometry using NanoDrop (Thermo Fisher Scientific, Waltham, MA).

Enzymatic assays

Screening for acceptor and donor specificity

All enzymatic reactions were carried out at 37 °C (corresponding to the human gut temperature) in 20 mM Tris-HCl buffer, pH 7.0. The Molybdenum Blue Activity Assay was used to assess the release of Pi during the reverse-phosphorolysis of the GP activity (De Groeve et al. 2010; Macdonald et al. 2019). The method is based on quantification of the Pi resulting from the reverse-phosphorolysis reaction. In acidic conditions, Pi forms molybdenum blue, which is quantified by absorbance measurement at 655nm. The reverse-phosphorolysis reactions were performed at 37°C. The reaction mixtures were composed

of 45 µL of cytoplasmic extract and 255 µL of substrate solution, resulting in a final reaction medium containing 20 mM Tris-HCl buffer (pH 7.0), 200 mM sodium molybdate, 10 mM of donor (list above) and 10 mM of acceptor (list above). After 0, 15, 30 and 60 min of incubation, 50 µL of the reaction mixture were transferred to 96-wells microplates containing, per well, 150 µL of a 0.24% (w/v) L-ascorbic acid solution in 0.1 N HCl. After incubation at room temperature for 5 min, 150 µL of stop solution (2% (v/v) acetic acid and 2% (w/v) sodium citrate tribasic dihydrate) were added to stop the reaction. At last, the absorbance was measured at 655 nm using a microplate reader (InfiniteM200pro, TECAN). The best acceptor was first identified using αM1P as donor. Then, the donor specificity was determined using the best acceptor identified with αM1P as donor. Negative control assays contained no donor, or no acceptor.

Mannooligosaccharide synthesis, phosphorolysis and hydrolysis assays

The synthesis of manno-oligosaccharides was performed in a reaction volume of 100 µL with 0.1 mg/ml purified enzyme, 10mM of αM1P and 10 mM of mannose in 20 mM Tris-HCl buffer, pH 7.0, during 24h at 37°C. The mannooligosaccharide degradation reactions were performed in 20 mM Tris-HCl buffer, pH 7.0 during 24h at 37°C, in reaction volumes of 100 µL, containing either 0.02 or 0.1 mg/mL purified enzyme, 0.1 mg/ml of a mixture containing β-1,2-mannobiose and mannose (massic ratio 87/13, which corresponds to final concentrations of 72 µM mannose and 254 µM mM β-1,2-mannobiose in the reaction mixture), with (for phosphorolysis) or without (for hydrolysis) 10 mM Pi.

HPAEC-PAD analysis

The variation of αM1P, carbohydrate acceptor and product amounts was assessed by high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Carbohydrates and αM1P were separated on a 4*250 mm Dionex CarboPak PA100 column. A gradient of sodium acetate (from 0 to 150mM in 15 min) and an isocratic step of 300 mM sodium acetate in 150 mM NaOH was applied at a 1mL min-1 flow rate. Detection was performed by using a DionexED40 module with a gold working electrode and an Ag/AgCl pH reference.

NMR Spectroscopy

Spectra were acquired using a Bruker Advance 500-MHz spectrometer with a 5-mm z-gradient TBI probe at 298 K, and an acquisition frequency of 500.13 MHz for ^{1}H and 125.75 MHz for ^{13}C NMR. The data were processed using TopSpin 3.0 software. The freeze-dried reaction media and standards were re-suspended in deuterated water. Deuterium oxide was used as the solvent and the chemical shift scales were internally referenced to sodium 2,2,3,3-tetradeuterio-3-trimethylsilylpropanoate (d4-TSP) for ^{1}H and to the non-deuterated acetone (singlet at 30.89 ppm) for ^{13}C NMR. The various signals were assigned by comparison with signals obtained from D-mannose, αMan1P, β-D-mannopyranosyl-1,4-D-mannose (Megazyme, Irleland), and β-D-mannopyranosyl-1,3-D-mannose (Carbosynth, UK), and the β-D-mannopyranosyl-1,2-D-mannose used as substrate for the degradation reactions. Usual 2D techniques such as TOCSY, HSQC, and HMBC were used.

Mass spectrometry

The synthesized products were analyzed by high resolution ion mobility-mass spectrometry (IMS-MS) by David Ropartz and Simon Ollivier at the BIBS platform (https://www.bibs.inrae.fr). Experiments were performed on a Select Series Cyclic IMS (Waters Corp., Manchester, UK), using electrospray ionization. The analysis procedures will be detailed in a collaborative paper with the results presented

below and the structural characterization of the products synthesized from β-1,2-mannobiose, in presence and absence of inorganic phosphate.

Results

GH130 sequence prevalence in the human gut microbiome

Among the 1,205 bacterial GH130 sequences identified in the human intestinal metagenome (A. Li et al. 2020), we searched for the one with the highest prevalence in IBD individuals (n= 359, against n=401 for healthy ones). Two sequences were found quite exclusively in the metagenome of the IBD individuals (Figure 1). The MH0055_GL0072634 sequence clusters in metanode C3 of the sequence similarity network (SSN) of the GH130 family, together with the β-1,2-mannosidases (A. Li et al. 2020). It is found in the gut metagenome of 147 individuals, of which 90% are from the IBD cohort. Our attempts to express gene MH0055_GL0072634 in *E. coli* to characterize the encoded enzyme failed. The second one (MH0105_GL0062872) clusters in metanode C4 of the GH130 SSN, together with the β-1,2-mannoside-phosphorylases Lin0857 from *Listeria innocua* Clip11262 (CAC96089.1, (A. Li et al. 2020)). The MH0105_GL0062872 sequence was found in 120 individuals, of which 85 % are from the IBD cohort. We named the encoded enzyme IBD_P1. Its sequence was assigned to Bacteroidetes by Li et al. (J. Li et al. 2014). In the non-redundant NCBI database, the nearest sequence of IBD_P1 in a sequenced genome is the Bacsa1715 protein from *Bacteroides salanitronis* DSM 18170 (ADY36113.1), with 99% coverage and 87% identity. The second one is the XB1A BXY_23240 protein from *Bacteroides xylanisolvens* (CBK67392.1), with 98% coverage and 87% identity. Both of these sequences also belong to metanode C4 of the SSN, as IBD_P1, but the corresponding proteins have not yet been biochemically characterized. The distribution of the most similar sequences of Bacsa1715 and BXY_23240 in the human gut metagenome (MH0290_GL0113903 and O2.UC21-2_GL0054119, respectively) is very different from that of IBD_P1. MH0290_GL0113903, which was initially discovered in the chicken gut microbiota. MH0290_GL0113903 is very rare, almost absent from the human gut metagenome (Figure 2). In contrast, O2.UC21-2_GL0054119 is very prevalent and abundant, whatever is the medical status of the individuals. The IBD_P1 sequence distribution in the human metagenome is also very different from that of all other characterized members of the GH130 family (Figure 3), of which only one (Uhgb_MP, ADD61463.1) is differentially prevalent and abundant in IBD and healthy individuals, as previously described (Ladevèze et al. 2013). Nevertheless, this trait is dramatically less marked than for IBD_P1, for which the correlation between its presence in the gut metagenome and the IBD status is striking. IBD_P1 likely belongs to a bacterium of the Bacteroides genus, even though the species and strain from which it is issued is yet unknown, given that no identical sequence was found in a sequenced genome.

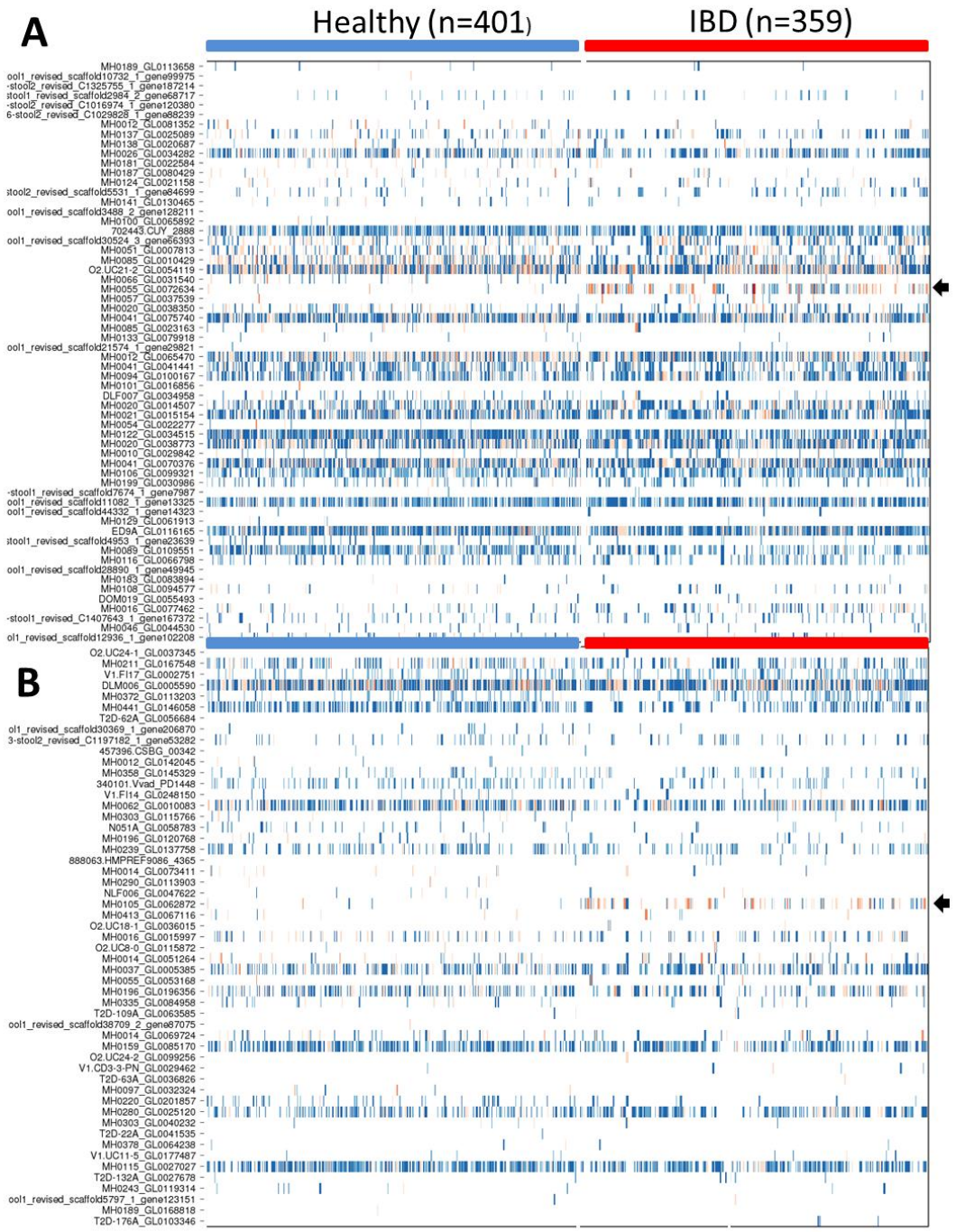


Figure 1: Abundance and prevalence of GH130 sequences in the gut metagenome of European healthy and IBD individuals. (A) Genes of the C3 and metanode; (B) genes of the C4 metanode. Genes are in rows. Individuals are in columns. The normalized abundance is represented by the following color scale: white, not detected; dark blue, light blue, orange and red, increasing abundance with a 100-fold change between colors. Black arrow, IBD markers.

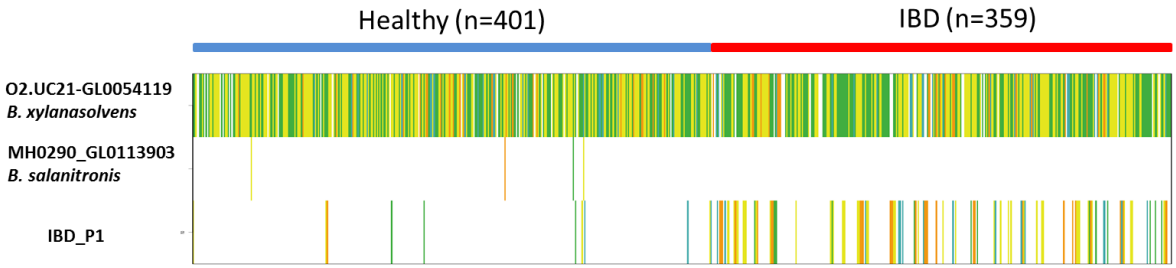


Figure 2: Abundance and prevalence in the gut metagenome of European healthy and IBD individuals of the IBD_P1 encoding gene, and of its two best blast hit sequences. Gene abundance in the gut metagenomes is represented by the following color scale: white, not detected; blue, turquoise, green, yellow and orange, increasing abundance with a 10-fold change between colors.

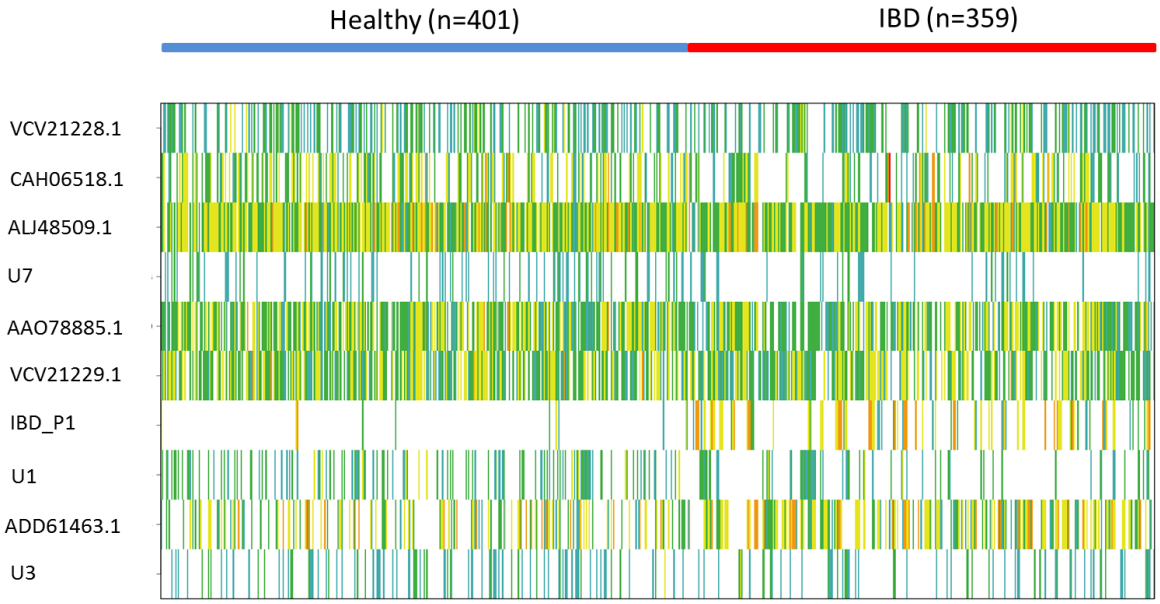
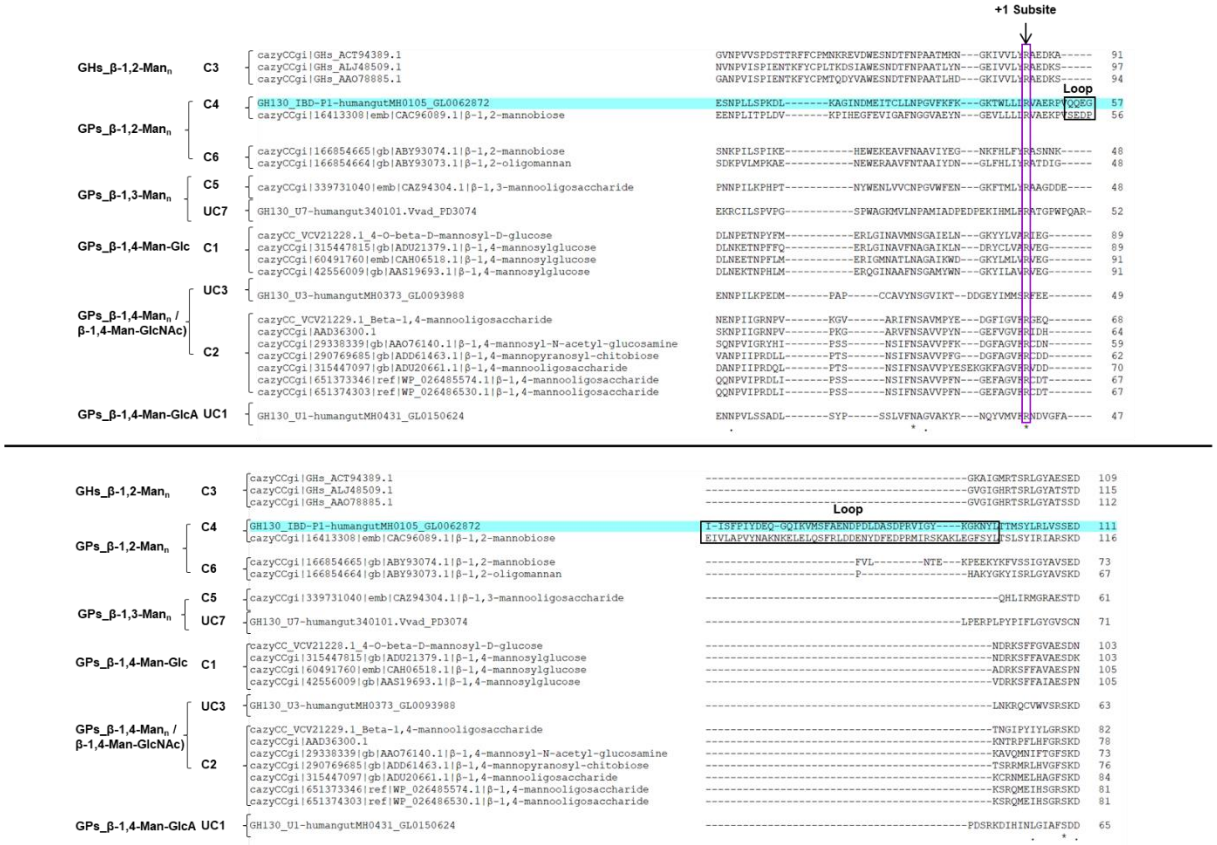


Figure 3: Abundance and prevalence of genes encoding characterized GH130 enzymes in the gut microbiome of European healthy and IBD individuals. Gene abundance in the gut metagenomes is represented by the following color scale: white, not detected; blue, turquoise, green, yellow and orange, increasing abundance with a 10-fold change between colors.

Prediction of IBD_P1 mechanism and linkage specificity

IBD_P1 is quite distantly related to the 21 GH130 enzymes characterized to date, since it shares only 44% sequence identity (98% coverage) with Lin0857, the most similar enzyme. The Lin0857 enzyme preferentially phosphorolyzes β-1,2-mannobiose over β-1,2-mannotriose (Tsuda et al. 2015). Like all known GP encoding sequences and the other sequences of the C4 metanode, that of IBD_P1 has no signal peptide. In addition, it contains the well conserved Asp amino acid acting as proton donor in all GH130 enzymes, and no Asp or Glu residues could be found at the location of the putative catalytic base residue in GH130 mannoside-hydrolases (Figure 4). The His and Arg amino acids involved in phosphate binding in GH130 GPs are conserved in the IBD_P1 sequence. The conserved Arg and Asp residues of the +1 subsite of GH130 enzymes are also present, together with the "EK" motif (Glu/Lys) conferring specificity for β-1,2-mannosides (A. Li et al. 2020). The canonical catalytic machinery of GH130 β-1,2-mannoside-phosphorylases is thus perfectly conserved in IBD_P1. Furthermore, the IBD_P1 sequence contains the 61 amino acids loop described in the Lin0857 enzyme (Figure 4). The presence of this loop is a specific trait of the C4 metanode's sequences, since the other metanodes containing β-1,2-mannosidases (C3) and β-1,2-mannoside phosphorylases (C6) lack it. In Lin0857, this loop covers the active site and confines a small pocket for di- or trisaccharide binding (Tsuda et al. 2015). These structural informations allowed predicting the IBD_P1 enzyme as a phosphorylase specific for β-1,2-mannooligosaccharides with a polymerization degree (DP) of 2 or 3.



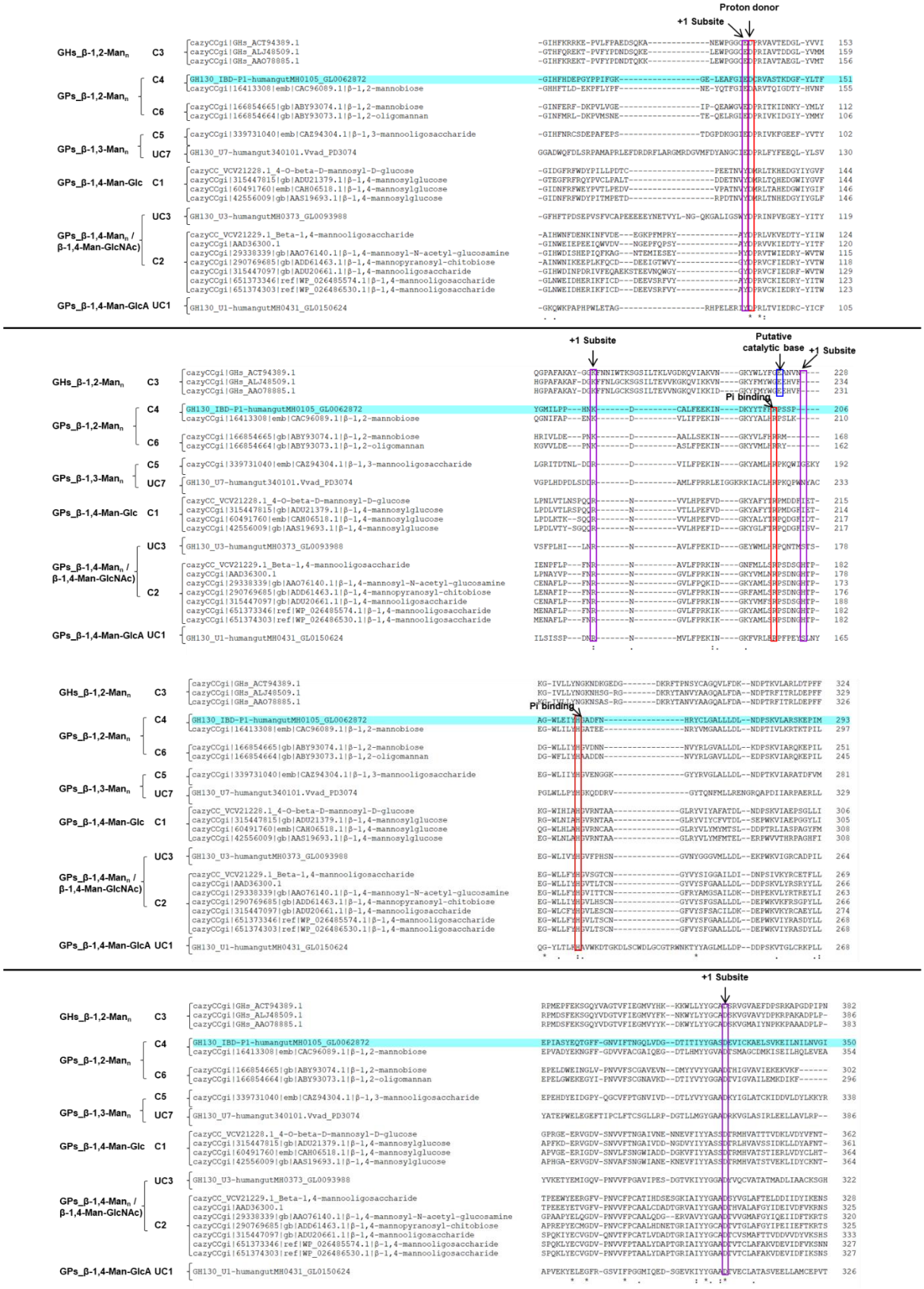


Figure 4: Multiple sequence alignment of all the characterized GH130 enzymes, including IBD_P1.

Profiling of IBD_P1 specificity towards glycosyl acceptors and donors in reverse-phosphorolysis

The IBD_P1 encoding gene was synthetized and expressed in *E. coli*. A soluble recombinant protein of 40 KDa was produced. To investigate the phosphorylase activity of IBD_P1, we first examined its specificity towards acceptors in the reverse-phosphorolysis reaction, using the usual donor of GH130 GPs, αMan1P. We screened the IBD_P1 activity in crude enzymatic extracts, using the Molybdenum Blue colorimetric method based on the detection of released Pi during reverse-phosphorolysis (A. Li et al. 2020; De Groeve et al. 2010; Macdonald et al. 2019). Among the twelve carbohydrate acceptor candidates, only Man was efficiently used as substrate. Using Man as acceptor, we confirmed that αMan1P is the best donor (Figure 5A and B). These results confirmed that IBD_P1 is a glycoside-phosphorylase quite specific for mannosyl residues, both as donor and acceptor. In addition, a slight αMan1P hydrolysis activity was highlighted, since Pi was released from αMan1P as sole substrate (Figure 5B).

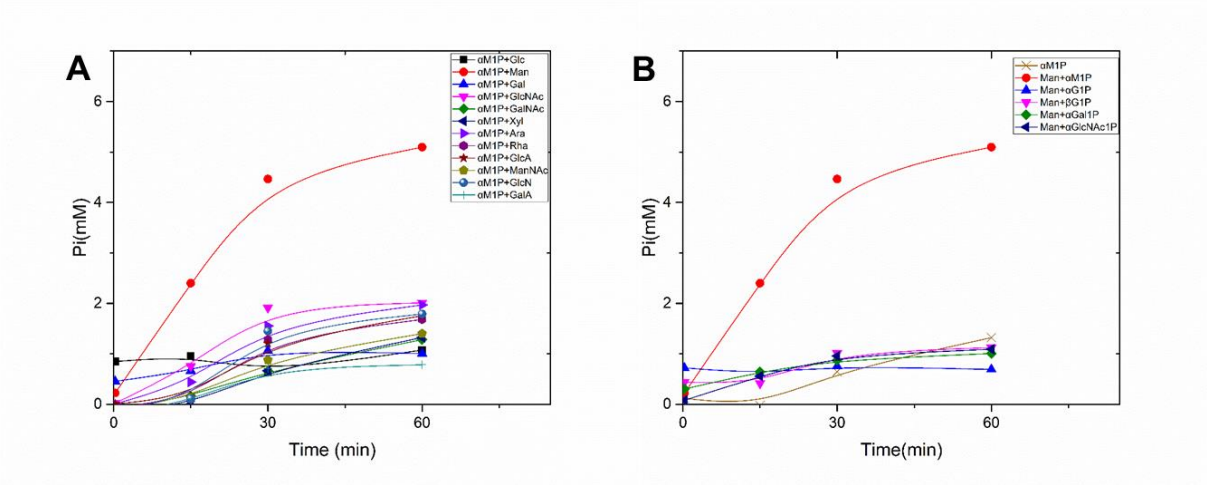


Figure 5: screening of the acceptor (A) and donor (B) specificity of IBD_P1, in reverse-phosphorolysis. The release of inorganic phosphate was followed using the molybdenum blue activity assay.

Determination of IBD_P1 linkage specificity, and characterization of its mannoside synthesis and degradation abilities

To determine the polymerization degree of the reverse-phosphorolysis products and the IBD_P1 linkage specificity, the enzyme was purified, tested for reverse-phosphorolysis, phosphorolysis and hydrolysis, and the reaction media were analyzed by HPAEC-PAD. At the enzyme concentration usually used to characterize GPs (0.1mg/mL, (A. Li et al. 2020; Ladevèze et al. 2013)), the reverse-phosphorolysis reaction was very fast, since oligosaccharides were synthesized from αMan1P and Man after only 30 seconds of reaction (Figure 6A). The retention time of these products were compared to those of the β-1,2-mannobiose and β-1,2-mannotriose synthesized by the Teth514_1788 β-1,2-mannoside-phosphorylase from αMan1P and Man (Chiku et al. 2014). The retention time of one of the products synthesized by IBD_P1 strictly corresponds to that of β-1,2-mannobiose, which is different from those of β-1,3-mannobiose and β-1,4-mannobiose (A. Li et al. 2020). The presence of β-1,2-mannobiose in the reaction mixture was also confirmed by ^{1}H NMR (Figure 7A, (spectra 2 & 3) with the signals of the α- and β-anomeric protons at 5.30 and 5.0 ppm, respectively, from the reducing terminal mannose),

confirming the β-1,2 linkage specificity of IBD_P1 that was predicted by sequence analysis. An oligosaccharide with a retention time corresponding to that of β-1,2-mannotriose was also produced, but surprisingly, the corresponding peaks on the HPAEC-PAD chromatograms at 30s and 24h of reaction are not fully resolved, highlighting the presence of at least one other oligosaccharide, likely a DP3 mannoside, in the reaction mixture. According to published date (Faille et al. 1992; Shibata et al. 1992) the presence of β-1,2-mannotriose was confirmed by 1H NMR (Figure 7A, spectrum 3) with the signals of the α- and β-anomeric protons at 5.288 and 4.996 ppm, respectively, from the reducing terminal mannose and also H2'β (4.427 ppm) and H2'α (4.292 ppm) for the beta anomer and alpha anomer of mannotriose respectively (Figure 7C).As β-1,2-mannotriose, this product could have been produced by reverse-phosphorolysis from β-1,2-mannobiose and αMan1P. In this case, it would be either β-1,3-mannotriose, β-1,4-mannotriose or β-1,6-mannotriose. The other hypothesis is that this putative DP3 mannoside resulted from transmannosylation, that is the transfer of a mannosyl residue from an oligosaccharide to an hydroxylated acceptor, without involvement of phosphate. GH130 enzymes being inverting enzymes, if the donor oligosaccharide is α-linked, the synthesized linkage would be in the β-configuration, and reversely. In this case, the DP3 would thus be a mixed β- and α-linked mannotriose.

To further investigate the IBD_P1 mechanism and its ability to degrade and to synthesize mannooligosaccharides, we tested its ability to phosphorolyse β-1,2-mannobiose in presence of Pi. We reduced the enzyme concentration from 0.1 to 0.02 mg/mL for this reaction, with the objective to better describe what happened between the initial and final times of reaction. Due to the exorbitant price of purified β-1,2-mannobiose (110 €/mg), we used, as substrate, a mixture of Man and β-1,2-mannobiose that were co-purified from the Teth514_1788-catalyzed reverse-phosphorolysis reaction mixture. As shown in Figure 6B, IBD_P1 efficiently degraded β-1,2-mannobiose. In addition, after only 30s of reaction, one or several DP3 oligosaccharides also appeared, highlighted by a peak at the same retention time as the one appearing after 30s of reaction. Its retention time is not the same as that of β-1,2-mannotriose, indicating that it probably corresponds to a mixture of the same DP3 products as those produced by reverse-phosphorolysis, or only to the DP3 with the highest retention time.

To investigate the hydrolysis and transmannosylation abilities of IBD_P1, we tested the enzyme without Pi, in presence of the same mixture of Man and β-1,2-mannobiose as for phosphorolysis. At an enzyme concentration of 0.02 mg/ml, no reaction occurred (Figure 6C), indicating that IBD_P1 is more efficient to phosphorylate β-1,2-mannobiose than to hydrolyze it. However, with a five times higher enzyme concentration (0.1 mg/ml), β-1,2-mannobiose was significantly hydrolyzed into Man (Figure 6D). A very significant additional peak also appeared at a retention time slightly higher than that of β-1,2-mannotriose. Since there was no αMan1P nor Pi in the reaction medium, the oligosaccharide(s) corresponding to this peak on HPAEC-PAD chromatogram was/were not produced by reverse-phosphorolysis, but by transmannosylation. To check that this/these product(s) did not result from the activity of an *E. coli* CAZyme that could have been co-purified with the recombinant IBD_P1 enzyme, we tested an enzyme which does not display any mannoside-phosphorylase activity (sequence MH0257_GL0045417, unpublished data), produced and purified in the same conditions as IBD_P1, at 0.1 mg/ml. No reaction occurred from β-1,2-mannobiose (data not shown).

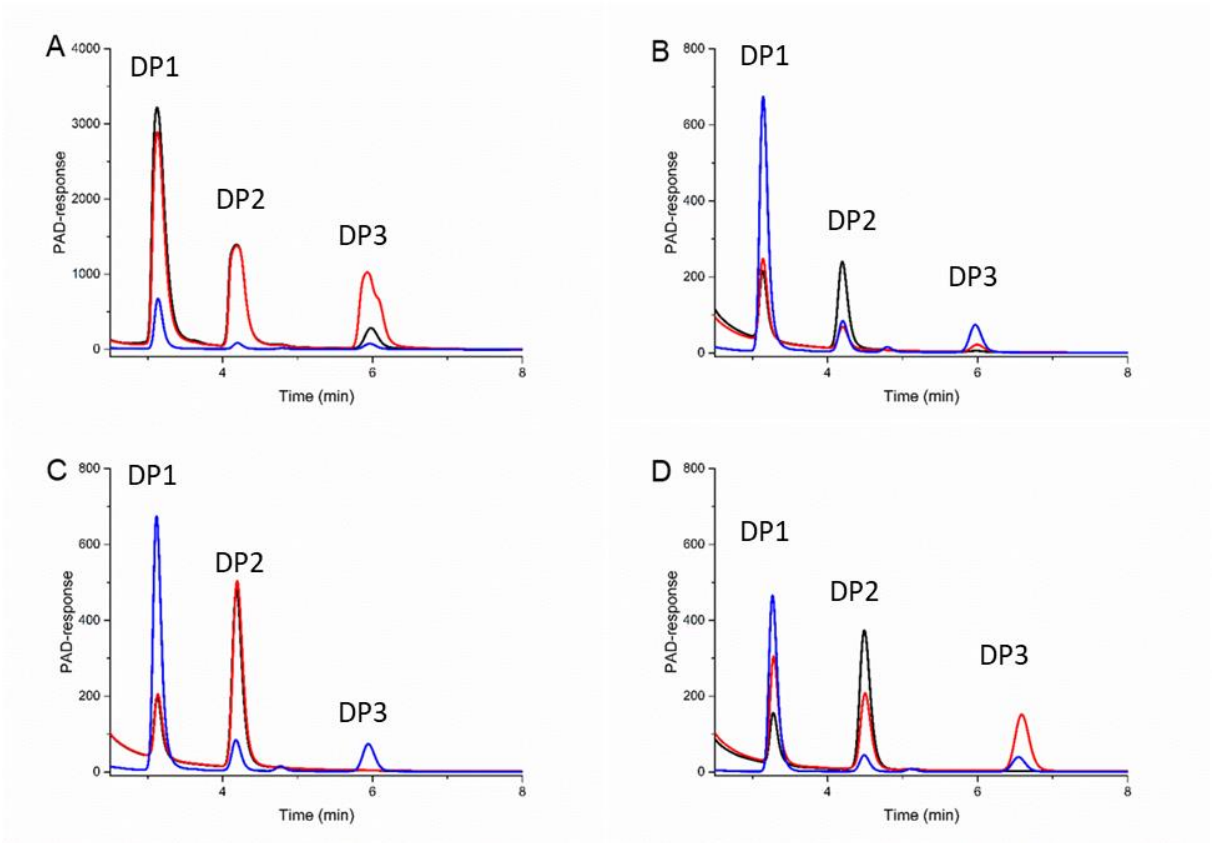
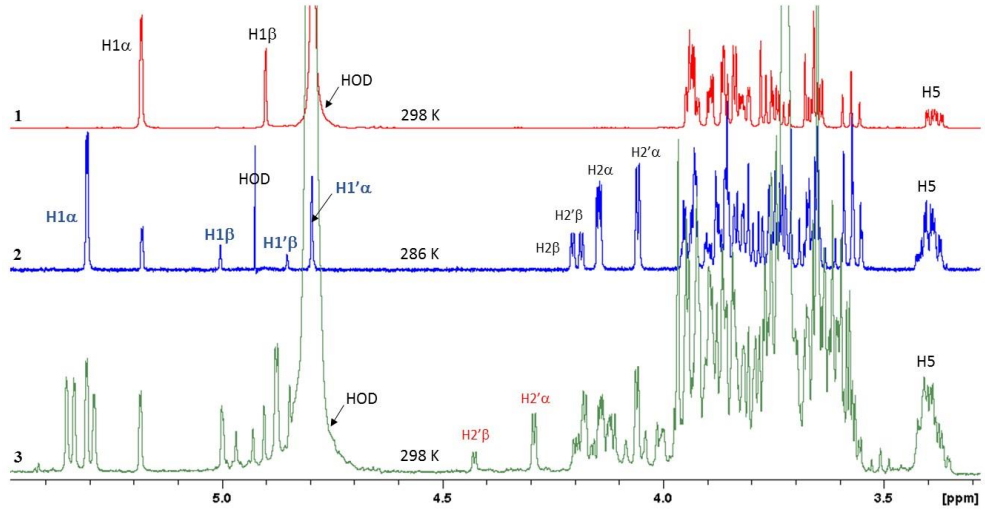
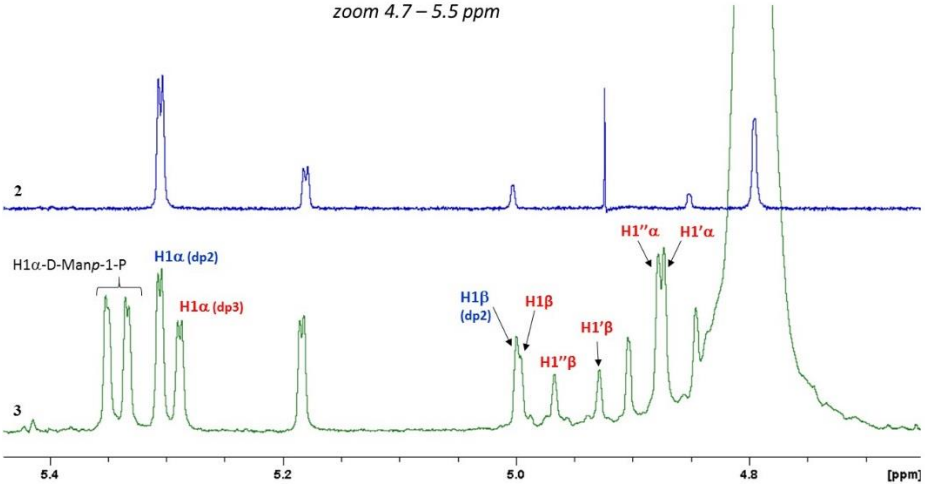


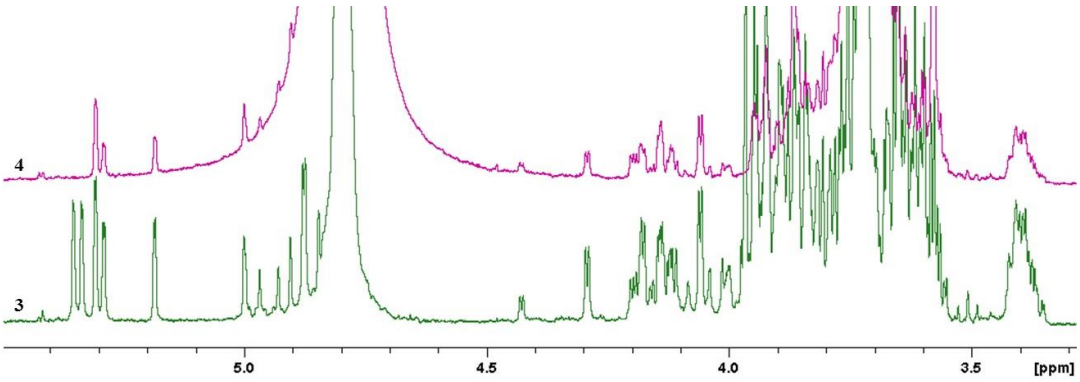
Figure 6: HPAEC-PAD analysis of the monosaccharides and oligosaccharides of the reactional mixtures obtained by IBD_P1-catalyzed reverse-phosphorolysis (A), phosphorolysis (B), transmannosylation/hydrolysis (C, D), after incubation for 30 sec and (in black) and 24hrs (in red). The chromatograms are compared to those containing Man, β-1,2-mannobiose and β-1,2-mannotriose obtained from αMan1P and Man by reverse-phosphorolysis with the Teth514_1788 enzyme (in blue). The reactions were performed in the presence of 0.1 mg/ml (A, D) or 0.02 mg/ml of purified IBD_P1 (B, C). The reverse-phosphorolysis reaction was performed from 10 mM of αMan1P and 10 mM of Man. The phosphorolysis reaction was performed from 254 µM of β-1,2-mannobiose, 72 µM of Man and 10 mM Pi. The transmannosylation/hydrolysis reactions were performed from 254 µM of β-1,2-mannobiose and 72 µM of Man, in absence of Pi. The HPAEC-PAD elution gradient was different in expriment D compared to those of A, B, C. This will be written in the experimental procedures as soon I can access to the laboratory after the CoVID-19 lockout.

A


H1α
H1β
HOD
298 K
H5
H2'α
H2α
H2'β
H2β
H1'α
H1'β
286 K
H2'α
H2'β
1
2
3
5.0
4.5
4.0
3.5
[ppm]


zoom 4.7 – 5.5 ppm
H1α-D-Manp-1-P
H1α (dp2)
H1α (dp3)
H1β (dp2)
H1β
H1'β
H1''β
H1''α
H1'α
2
3
5.4
5.2
5.0
4.8
[ppm]

B


4
3
5.0
4.5
4.0
3.5
[ppm]

C

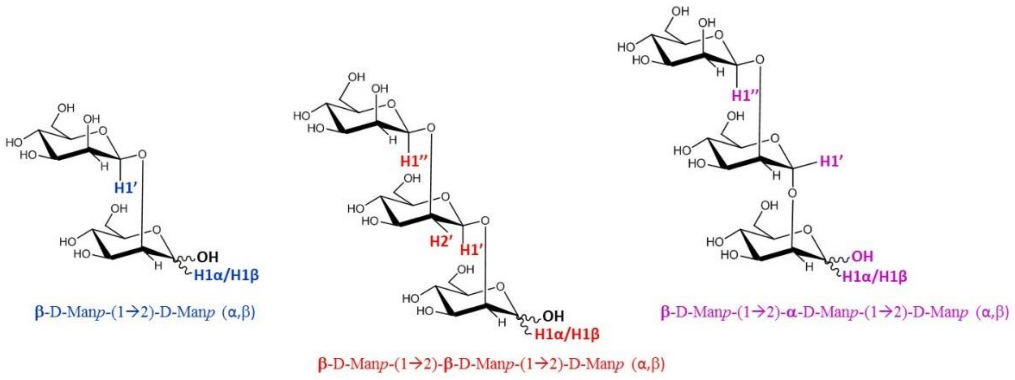


Figure 7: 1H NMR spectra of the reverse-phosphorolysis (A, spectrum 3) and transmannosylation/hydrolysis (B spectrum 4) reactional mixtures incubated for 24h with 0.1 mg/ml of purified IBD_P1. The reverse-phosphorolysis reaction (3) was performed in presence of 10 mM of αMan1P and 10 mM of Man. The transmannosylation/hydrolysis reaction (4) were performed from 254 µM of β-1,2-mannobiose and 72 µM of Man, in absence of Pi . (1) 1H-NMR of mannose at 298K in D2O, (2) 1H-NMR of purified β-1,2-mannobiose at 286 K in D2O. (C) Structures and anomeric annotation of Man-β1,2-Man, Man-β1,2-Man-β1,2-Man and Man-β1,2-Man-α1,2-Man.

To confirm the polymerization degree and determine the structure of the product(s) synthesized by transmannosylation, Simon Ollivier and David Ropartz (BIBS platform, INRAE Nantes) analyzed the reaction mixture obtained after 24 h of reaction with 0.1 mg/ml enzyme from β-1,2-mannobiose and Man. Using a novel approach combining tandem mass spectrometry (MS/MS) with multidimensional ion mobility spectrometry (IMS-IMS), they specifically analyzed the structure of the DP3 oligosaccharides by mobility selection and fragmentation, starting with a very small amount of product, since we provided them with only 1 µg of carbohydrates in total, together with α-1,2-mannobiose and β-1,2-mannotriose (co-purified with β-1,2-mannobiose and β-1,2-mannotetraose from the Teth514_1788-catalyzed reverse-phosphorolysis reaction mixture) as standards.

MS analysis of the reaction mixture indicated that no DP4 was synthesized, according to the HPAEC-PAD results and to the predicted specificity of IBD_P1 towards β-1,2-mannooligosaccharides of DP2 and DP3. The extracted ion mobiligrams (EIM) of the $[DP3+Li]^+$ and $[DP3+Na]^+$ ions and the arrival time distributions (ATD) of each of their fragments highlighted the presence in the reaction mixture of two DP3 mannooligosaccharides, probably Man-β1,2-Man-α1,2-Man and Man-β1,2-Man-β1,2-Man, of which the structures are schematized on Figure 7C. ^{18}O labelling of the reducing ends, which will allow the distinction between C and Y fragments, respectively at the non-reducing and reducing ends, will be performed to definitively confirm the anomeric conformation of Man-β1,2-Man-α1,2-Man. These two DP3 mannooligosaccharides are the same as those produced in the reverse-phosphorolysis reaction, as revealed by NMR analysis of both reaction mixtures (Figure 7B).

In the transmannosylation/hydrolysis sample, the Man-β1,2-Man-α1,2-Man trisaccharide was likely produced by a three-step reaction (Figure 8): i) β-1,2-mannobiose hydrolysis in Man (even though Man was already present in the reaction medium); ii) transmannosylation from β-1,2-mannobiose as donor and Man as acceptor, yielding to α-1,2-mannobiose; iii) transmannosylation from α-1,2-mannobiose as donor and acceptor, yielding to Man-β1,2-Man-α1,2-Man. Since no αMan1P nor Pi was present in the

reaction medium to allow IBD_P1 performing reverse-phosphorolysis, we speculate that Man-β1,2-Man-β1,2-Man is the product of a transmannosylation reaction from α1,2-mannobiose as donor and β-1,2-mannobiose as acceptor (Figure 8). With the analyses performed to date, no trace of α-1,2-mannotriose or Man-α1,2-Man-β1,2-Man was detected, indicating that no transmannosylation reaction occurred from β1,2-mannobiose as donor and α-1,2-mannobiose as acceptor, or from β1,2-mannobiose as donor and as acceptor (Figure 8). ^{18}O labelling of the reducing end of the DP3 products will allow us to confirm this result.

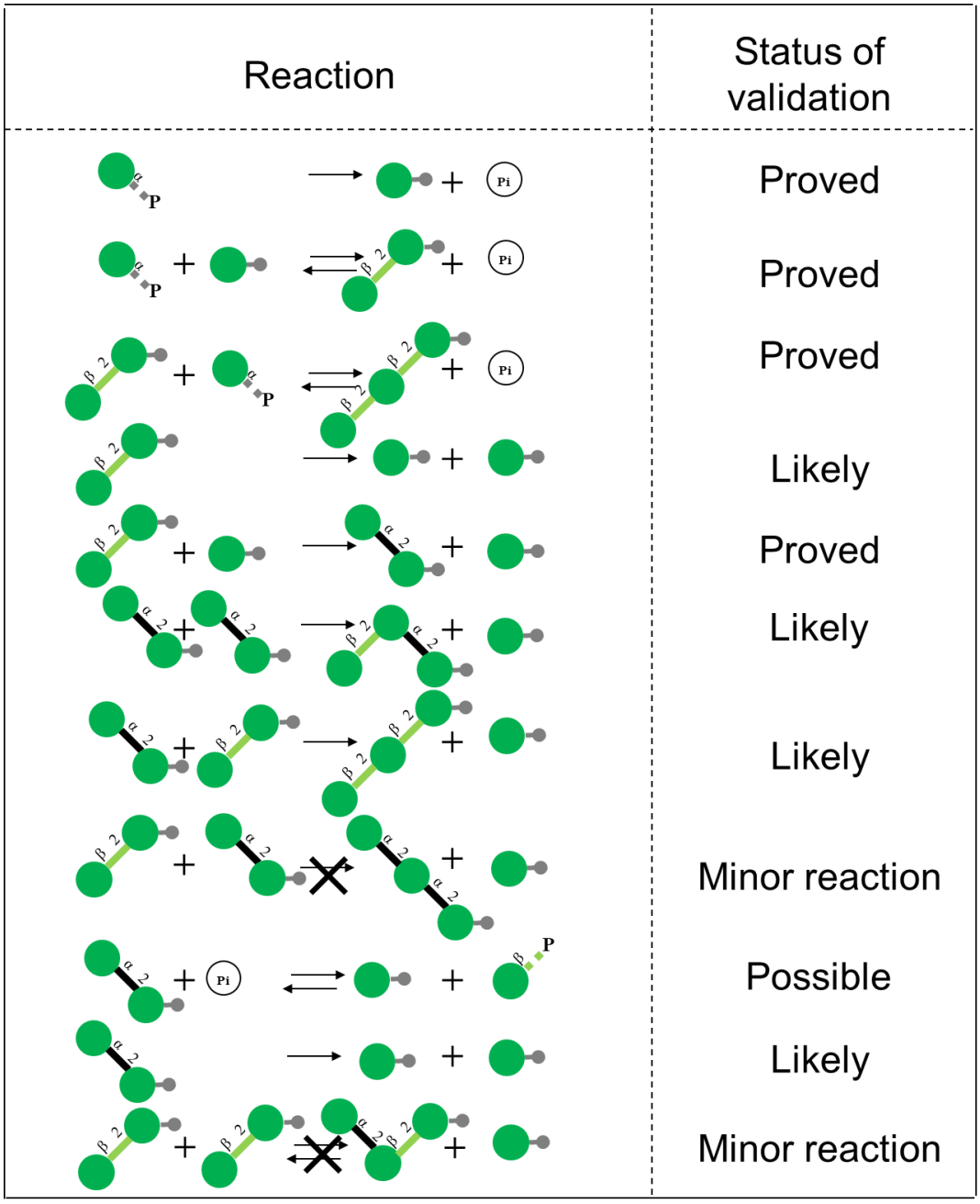


Figure 8: List of the reverse-phosphorolysis, phosphorolysis, hydrolysis and transmannosylation reactions proved or hypothesized to be catalyzed by the IBD_P1 enzyme. The reactions proved to be impossible are also mentioned.

Discussion

IBD_P1: the first known dual-activity mannoside-phosphorylase/transmannosylase

We discovered and characterized IBD_P1, a new enzyme of the GH130 family, issued from an uncultured human gut bacterium. By analyzing the SSN previously constructed for the GH130 family, conservation of residues involved in catalysis, phosphate and acceptor binding, and of a specific loop restricting access to long substrates, we predicted, and further confirmed by biochemical characterization, that this enzyme is an inverting glycoside-phosphorylase specific for β-1,2-mannosides of DP 2 and 3. Results of screening for acceptors and donors in reverse-phosphorolysis highlighted its strict specificity for αMan1P and Man as sole donor and acceptor, respectively. However, contrary to all other characterized mannoside-phosphorylases, this enzyme does not only catalyze phosphorolysis and reverse-phosphorolysis reactions, but also transmannosylation, and likely, oligosaccharide hydrolysis. This is the first mention of an enzyme presenting dual transmannosylase and mannoside-phosphorylase activities. Glycoside-phosphorylases with transglycosidase activity were already described, for example sucrose-phosphorylases, which can be employed in the enzymatic synthesis of α-D-glucosides (Aerts et al. 2011; Kitao et al. 1993). However, the term transglycosylation is often employed in litterature to describe glycosylation by reverse-phosphorolysis from sugar-phosphates, and should not be confused with a real transglycosylation reaction, which occurs from oligosaccharides or polysaccharides as glycosyl donors.

In the present study, transmannosylation products were evidenced both by HPAEC-PAD analysis and using a novel, highly resolutive pipeline for oligosaccharide structure determination, combining MS/MS and IMS/IMS. Transmannosylation occured in a significant yield, since both the reverse-phosphorolysis and transmannosylation DP3 products were synthesized in presence of αMan1P and Man, even at the very beginning of the reaction. In absence of αMan1P or Pi, transmannosylation was also catalyzed with a quite low enzyme concentration, usually used to characterize GPs. However, it appeared that the enzyme/substrate ratio is an important parameter to control transmannosylation, since we observed this reaction with a molar ratio of 10.10^{-3} but not with 2.10^{-3}. The only way for IBD_P1 to synthesize Man-β1,2-Man-α1,2-Man is by transmannosylation from α-1,2-mannobiose as donor and acceptor. In addition, in absence of αMan1P, the only way for it to synthesize Man-β1,2-Man-β1,2-Man is by transmannosylation from α1,2-mannobiose as donor and β-1,2-mannobiose as acceptor. Preliminary IMS/IMS experiments targeting the DP2 oligosaccharides showed the presence of α-1,2-mannobiose in the reaction mixture from β1,2-mannobiose and Man. We hypothesize that the α1,2-mannobiose retention time on HPAEC-PAD chromatograms would be very near from that of β1,2-mannobiose, explaining why we did not observe two peaks corresponding to putative DP2 oligosaccharides. This will be checked with the commercial α1,2-mannobiose standard. In addition, we will have to confirm the ability of IBD_P1 to perform hydrolysis, phosphorolysis and transmannosylation from α1,2-mannobiose as sole carbohydrate substrate, to definitively confirm the transmannosylation mechanism used by IBD_P1 to synthesize Man-β1,2-Man-α1,2-Man. Furthermore, we showed that IBD_P1 is unable to produce α-1,2-mannotriose or Man-α1,2-Man-β1,2-Man by transmannosylation, while it is able to produce α-1,2-mannobiose from β-1,2-mannobiose and Man. These results proved that the -1 subsite of IBD_P1 is specific for α-linked mannosides, when its +1 and +2 subsites are occupied by α- or β-linked mannobiose. Nevertheless, it could also accommodate β-linked mannosides when only Man is bound in the +1 subsite (Figure 9). The resolution of the IBD_P1 crystallographic structure will be necessary to identify the molecular determinants of such a flexibility of the donor and acceptor binding sites.

At this stage, the catalytic duality of IBD_P1 is only an *in vitro* curiosity and any biological implication remains to be proved. Nevertheless, the original catalytic properties of this new enzyme make it a very interesting enzyme for the synthesis, in a one-pot reaction from αMan1P and Man, of Man-β1,2-Man-α1,2-Man. To our knowledge, IBD_P1 is the first known enzyme which is able to synthesize mixed β- and α-linked oligosaccharides, and the only known member of the GH130 family which targets α-linked oligosaccharides. The Man-β1,2-Man-α1,2-Man trisaccharide is a particularly interesting target for synthesis of antigenic epitopes. Indeed, it is a key motif linking the α-1,2- and β-1,2-linked mannosyl chains in yeast N-glycans (Shibata et al. 2003), in particular in the antigenic factor 6 of the pathogenic yeast *Candida albicans* (Kobayashi, Shibata, and Suzuki 1992), which uses the β-mannosyltransferase *CaBmt1* from the GT91 family to synthesize this motif from α-1,2-linked mannosides (Fabre et al. 2014). The Man-β1,2-Man-α1,2-Man motif is also found in other microbial glycans, such as in the O antigen part of the lipopolysaccharide of the pathogenic *Salmonella* species (Liu et al. 2014) (Figure 10).

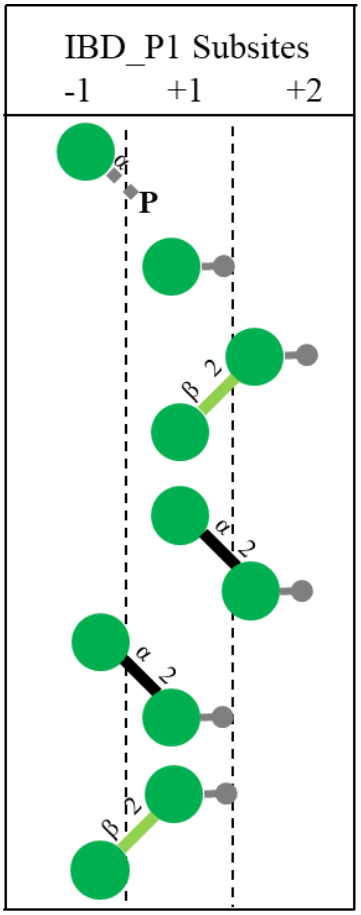


Figure 9: Specificity of the IBD_P1 donor and acceptor binding sites towards α- and β-linked mannosides.

Physiological function of IBD_P1

The IBD_P1 enzyme is classified in the C4 metanode of the GH130 SSN, which contains another characterized member, the Lin0857 enzyme from *Listeria innocua*. This enzyme is a canonical β-1,2-mannoside-phosphorylase, of which the biological function was not discussed by Tsuda et al., who discovered and characterized it (Tsuda et al. 2015). Prediction of the physiological role of a bacterial CAZyme can be done by analyzing the genomic context of the target gene. The structure of the targeted

glycan can indeed be deduced from the known (or predicted) activities of the CAZymes expressed from a same operon (Lapébie et al. 2019). The IBD_P1 sequence is a metagenomic sequence issued from the gene catalog of the human gut microbiome, in which the listed genes are disconnected from their genomic environment. Therefore, we analyzed the genomic environment of the two best blast hits of IBD_P1, Bacsa1715 from *Bacteroides salanitronis* DSM 18170 (ADY36113.1) and BXY_23240 from *Bacteroides xylanisolvens* XB1A (CBK67392.1). Both genes belong to predicted polysaccharide utilization loci (PUL) listed in the PULDB (Terrapon et al. 2015), the predicted PUL9 (Bacsa_1536 to 1541) and PUL31 (BXY_23190 to 23240), respectively. These PUL both contain SusC/SusD and MFS transporter sequences, but no other CAZyme encoding ones, which could have provided us with information on the structure of the PUL's cognate substrates. This is also the case for the other Bacteroidetes GH130 sequences of the C4 metanode, found in 13 PUL of the PULDB, which contain no other CAZy sequence to functionally complement the putative operon. This consistency in the genomic environment of the Bacteroidetes GH130 of the C4 cluster suggests that these GH130 may, *in vivo*, ensure the final phosphorolysis of short (< DP4) β-1,2- mannooligosaccharides internalized in bacterial cells, providing αM1P which can enter directly the central carbon metabolism without consumption of ATP. The IBD_P1 enzyme, in particular, would also be involved in the metabolization of α-1,2- mannobiose and of the Man-β1,2-Man-α1,2-Man trisaccharide. This original ability makes of it an enzyme perfectly adapted to the metabolisation of the N-glycans of *Candida albicans* (Figure 10). This yeast is one of the dominant fungal species in the human gut microbiota, since *C. albicans* operational taxonomic units are present in more than 80% of the stool samples of healthy individuals (Nash et al. 2017). However, in patients with IBD, the fungal microbiota is skewed, with an increased proportion of *C. albicans* compared with healthy individuals (Sokol et al. 2017). This yeast is also a possible initiator of the inflammatory process (Gerard et al. 2015). The increased proportion of *C. albicans* in the microbiota of IBD individuals is thus in accordance with the strikely overprevalence and overabundance of the IBD_P1 sequence in the microbiota of IBD individuals, given the substrate targeted by this enzyme. It would also explain the IBD marker status of the other sequence highlighted in the present study, which, belonging to metanode C3 of the GH130 SSN (A. Li et al. 2020), is predicted to target β-1,2-linked mannosides, which are highly abundant in *C. albicans* N-glycans. Besides, as explained above, the mannosides targeted by IBD_P1 are also present in the lipopolysaccharides of *Salmonella* species, which do not contain long β-1,2-linked mannan chains, contrary to *C. albicans* N-glycans (Figure 10). *Salmonella* is the cause of typhoid fever, paratyphoid fever, and the foodborne illness salmonellosis (Liu et al. 2014). Interestingly, *Salmonella* infection could promote the onset of IBD (Schultz et al. 2017; Gradel et al. 2009) and increase their severity (Szilagyi et al. 1985). The discovery and functional characterization of the IBD-associated enzyme IBD_P1 allowed us to decipher a novel mechanism of mannoside metabolization, which may thus be the key of disease-specific inter-kingdom and inter-genera interactions.

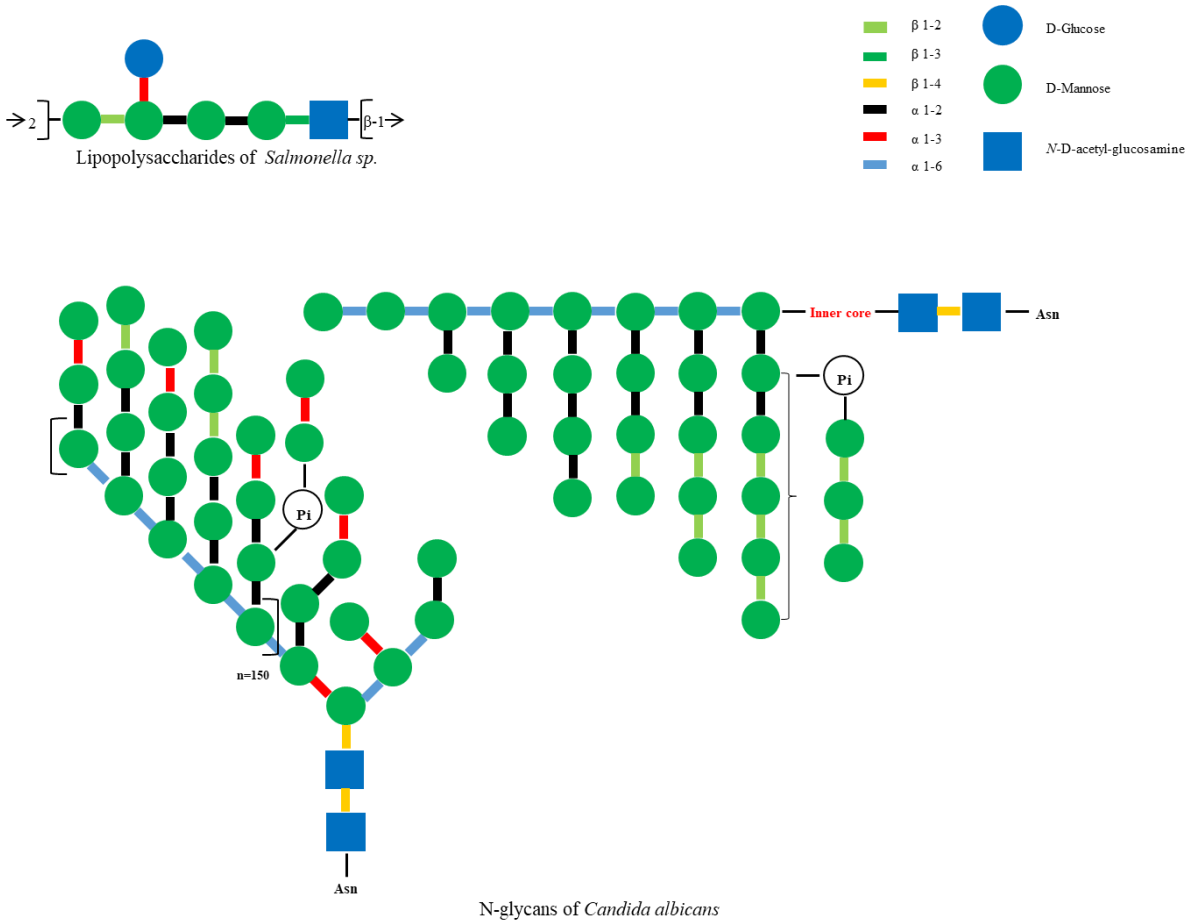


Figure 10: Schematic representation of the structures of the N-glycan of Candida albicans *and of the O-antigen part of the* Salmonella *lipooligosaccharides.*

References

Aerts, Dirk, Tom F. Verhaeghe, Bart I. Roman, Christian V. Stevens, Tom Desmet, and Wim Soetaert. 2011. "Transglucosylation Potential of Six Sucrose Phosphorylases toward Different Classes of Acceptors." *Carbohydrate Research* 346 (13): 1860–67. https://doi.org/10.1016/j.carres.2011.06.024.

Awad, Faisal Nureldin, Pedro Laborda, Meng Wang, Ai Min Lu, Qian Li, Zhi Peng Cai, Li Liu, and Josef Voglmeir. 2017. "Discovery and Biochemical Characterization of a Mannose Phosphorylase Catalyzing the Synthesis of Novel β-1,3-Mannosides." *Biochimica et Biophysica Acta (BBA) - General Subjects* 1861 (12): 3231–37. https://doi.org/10.1016/j.bbagen.2017.09.013.

Brett, Christopher T., and Keith W. Waldron. 1996. *Physiology and Biochemistry of Plant Cell Walls*. Springer Science & Business Media.

Chekan, Jonathan R., In Hyuk Kwon, Vinayak Agarwal, Dylan Dodd, Vanessa Revindran, Roderick I. Mackie, Isaac Cann, and Satish K. Nair. 2014. "Structural and Biochemical Basis for Mannan Utilization by Caldanaerobius Polysaccharolyticus Strain ATCC BAA-17." *Journal of Biological Chemistry* 289 (50): 34965–77. https://doi.org/10.1074/jbc.M114.579904.

Chiku, Kazuhiro, Takanori Nihira, Erika Suzuki, Mamoru Nishimoto, Motomitsu Kitaoka, Ken'ichi Ohtsubo, and Hiroyuki Nakai. 2014. "Discovery of Two β-1,2-Mannoside Phosphorylases

Showing Different Chain-Length Specificities from Thermoanaerobacter Sp. X-514." *PLOS ONE* 9 (12): e114882. https://doi.org/10.1371/journal.pone.0114882.

Cuskin, Fiona, Arnaud Baslé, Simon Ladevèze, Alison M. Day, Harry J. Gilbert, Gideon J. Davies, Gabrielle Potocki-Véronèse, and Elisabeth C. Lowe. 2015. "The GH130 Family of Mannoside Phosphorylases Contains Glycoside Hydrolases That Target β-1,2-Mannosidic Linkages in Candida Mannan." *Journal of Biological Chemistry* 290 (41): 25023–33. https://doi.org/10.1074/jbc.M115.681460.

De Groeve, M. R. M., G. H. Tran, A. Van Hoorebeke, J. Stout, T. Desmet, S. N. Savvides, and W. Soetaert. 2010. "Development and Application of a Screening Assay for Glycoside Phosphorylases." *Analytical Biochemistry* 401 (1): 162–67. https://doi.org/10.1016/j.ab.2010.02.028.

Dilokpimol, Adiphol, Hiroyuki Nakai, Charlotte H. Gotfredsen, Martin J. Baumann, Natsuko Nakai, Maher Abou Hachem, and Birte Svensson. 2011. "Recombinant Production and Characterisation of Two Related GH5 Endo-β-1,4-Mannanases from Aspergillus Nidulans FGSC A4 Showing Distinctly Different Transglycosylation Capacity." *Biochimica Et Biophysica Acta* 1814 (12): 1720–29. https://doi.org/10.1016/j.bbapap.2011.08.003.

Eneyskaya, Elena V., Gustav Sundqvist, Alexander M. Golubev, Farid M. Ibatullin, Dina R. Ivanen, Konstantin A. Shabalin, Harry Brumer, and Anna A. Kulminskaya. 2009. "Transglycosylating and Hydrolytic Activities of the Beta-Mannosidase from Trichoderma Reesei." *Biochimie* 91 (5): 632–38. https://doi.org/10.1016/j.biochi.2009.03.009.

Fabre, Emeline, Ghenima Sfihi-Loualia, Marilyne Pourcelot, Bernadette Coddeville, Frédéric Krzewinski, Julie Bouckaert, Emmanuel Maes, et al. 2014. "Characterization of the Recombinant Candida Albicans β-1,2-Mannosyltransferase That Initiates the β-Mannosylation of Cell Wall Phosphopeptidomannan." *The Biochemical Journal* 457 (2): 347–60. https://doi.org/10.1042/BJ20131012.

Gerard, Romain, Boualem Sendid, Jean-Frederic Colombel, Daniel Poulain, and Thierry Jouault. 2015. "An Immunological Link between Candida Albicans Colonization and Crohn's Disease." *Critical Reviews in Microbiology* 41 (2): 135–39. https://doi.org/10.3109/1040841X.2013.810587.

Gradel, Kim O., Hans L. Nielsen, Henrik C. Schønheyder, Tove Ejlertsen, Brian Kristensen, and Henrik Nielsen. 2009. "Increased Short- and Long-Term Risk of Inflammatory Bowel Disease After Salmonella or Campylobacter Gastroenteritis." *Gastroenterology* 137 (2): 495–501. https://doi.org/10.1053/j.gastro.2009.04.001.

Grimaud, Florent, Sandra Pizzut-Serin, Laurence Tarquis, Simon Ladevèze, Sandrine Morel, Jean-Luc Putaux, and Gabrielle Potocki-Veronese. 2019. "In Vitro Synthesis and Crystallization of β-1,4-Mannan." *Biomacromolecules* 20 (2): 846–53. https://doi.org/10.1021/acs.biomac.8b01457.

Ishimizu, Takeshi, Akiko Sasaki, Satoshi Okutani, Mami Maeda, Mai Yamagishi, and Sumihiro Hase. 2004. "Endo-β-Mannosidase, a Plant Enzyme Acting on N-Glycan Purification, Molecular Cloning, And Characterization." *Journal of Biological Chemistry* 279 (37): 38555–62. https://doi.org/10.1074/jbc.M406886200.

Jaito, Nongluck, Wataru Saburi, Rei Odaka, Yusuke Kido, Ken Hamura, Mamoru Nishimoto, Motomitsu Kitaoka, Hirokazu Matsui, and Haruhide Mori. 2014. "Characterization of a Thermophilic 4-O-β-d-Mannosyl-d-Glucose Phosphorylase from Rhodothermus Marinus." *Bioscience, Biotechnology, and Biochemistry* 78 (2): 263–70. https://doi.org/10.1080/09168451.2014.882760.

Kawahara, Ryosuke, Wataru Saburi, Rei Odaka, Hidenori Taguchi, Shigeaki Ito, Haruhide Mori, and Hirokazu Matsui. 2012. "Metabolic Mechanism of Mannan in a Ruminal Bacterium, Ruminococcus Albus, Involving Two Mannoside Phosphorylases and Cellobiose 2-Epimerase: Discovery of a New Carbohydrate Phosphorylase, β-1,4-Mannooligosaccharide Phosphorylase." *Journal of Biological Chemistry* 287 (50): 42389–99. https://doi.org/10.1074/jbc.M112.390336.

Kitao, Satoshi, Toshiaki Ariga, Takanao Matsudo, and Hiroshi Sekine. 1993. "The Syntheses of Catechin-Glucosides by Transglycosylation with Leuconostoc Mesenteroides Sucrose Phosphorylase." *Bioscience, Biotechnology, and Biochemistry* 57 (12): 2010–15. https://doi.org/10.1271/bbb.57.2010.

Kobayashi, Hidemitsu, Nobuyuki Shibata, and Shigeo Suzuki. 1992. “Evidence for Oligomannosyl Residues Containing Both Beta-1,2 and Alpha-1,2 Linkages as a Serotype A-Specific Epitope(s) in Mannans of Candida Albicans.” *Infection and Immunity* 60 (5): 2106–9.

La Rosa, Sabina Leanti, Maria Louise Leth, Leszek Michalak, Morten Ejby Hansen, Nicholas A. Pudlo, Robert Glowacki, Gabriel Pereira, et al. 2019. “The Human Gut Firmicute Roseburia Intestinalis Is a Primary Degrader of Dietary β-Mannans.” *Nature Communications* 10 (1): 905. https://doi.org/10.1038/s41467-019-08812-y.

Ladevèze, Simon, Laurence Tarquis, Davide A. Cecchini, Juliette Bercovici, Isabelle André, Christopher M. Topham, Sandrine Morel, et al. 2013. “Role of Glycoside Phosphorylases in Mannose Foraging by Human Gut Bacteria.” *Journal of Biological Chemistry* 288 (45): 32370–83. https://doi.org/10.1074/jbc.M113.483628.

Lapébie, Pascal, Vincent Lombard, Elodie Drula, Nicolas Terrapon, and Bernard Henrissat. 2019. “Bacteroidetes Use Thousands of Enzyme Combinations to Break down Glycans.” *Nature Communications* 10 (1): 1–7. https://doi.org/10.1038/s41467-019-10068-5.

Li, Ao, Elisabeth Laville, Laurence Tarquis, Vincent Lombard, David Ropartz, Nicolas Terrapon, Bernard Henrissa, et al. 2020. “Analysis of the Diversity of the Glycoside Hydrolase Family 130 in Mammal Gut Microbiomes Reveals a Novel Mannoside-Phosphorylase Function.” *Under Review*.

Li, Junhua, MetaHIT Consortium, Huijue Jia, Xianghang Cai, Huanzi Zhong, Qiang Feng, Shinichi Sunagawa, et al. 2014. “An Integrated Catalog of Reference Genes in the Human Gut Microbiome.” *Nature Biotechnology* 32 (8): 834–41. https://doi.org/10.1038/nbt.2942.

Liu, Bin, Yuriy A. Knirel, Lu Feng, Andrei V. Perepelov, Sof’ya N. Senchenkova, Peter R. Reeves, and Lei Wang. 2014. “Structural Diversity in SalmonellaO Antigens and Its Genetic Basis.” *FEMS Microbiology Reviews* 38 (1): 56–89. https://doi.org/10.1111/1574-6976.12034.

Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. 2014. “The Carbohydrate-Active Enzymes Database (CAZy) in 2013.” *Nucleic Acids Research* 42 (D1): D490–95. https://doi.org/10.1093/nar/gkt1178.

Macdonald, Spencer S., Zachary Armstrong, Connor Morgan-Lang, Magdalena Osowiecka, Kyle Robinson, Steven J. Hallam, and Stephen G. Withers. 2019. “Development and Application of a High-Throughput Functional Metagenomic Screen for Glycoside Phosphorylases.” *Cell Chemical Biology* 26 (7): 1001-1012.e5. https://doi.org/10.1016/j.chembiol.2019.03.017.

Murata, Takeomi, and Taichi Usui. 1997. “Preparation of Oligosaccharide Units Library and Its Utilization.” *Bioscience, Biotechnology, and Biochemistry* 61 (7): 1059–66. https://doi.org/10.1271/bbb.61.1059.

Nash, Andrea K., Thomas A. Auchtung, Matthew C. Wong, Daniel P. Smith, Jonathan R. Gesell, Matthew C. Ross, Christopher J. Stewart, et al. 2017. “The Gut Mycobiome of the Human Microbiome Project Healthy Cohort.” *Microbiome* 5 (1): 153. https://doi.org/10.1186/s40168-017-0373-4.

Nielsen, Henrik. 2017. “Predicting Secretory Proteins with SignalP.” In *Protein Function Prediction: Methods and Protocols*, edited by Daisuke Kihara, 59–73. Methods in Molecular Biology. https://doi.org/10.1007/978-1-4939-7015-5_6.

Nihira, Takanori, Kazuhiro Chiku, Erika Suzuki, Mamoru Nishimoto, Shinya Fushinobu, Motomitsu Kitaoka, Ken’ichi Ohtsubo, and Hiroyuki Nakai. 2015. “An Inverting β-1,2-Mannosidase Belonging to Glycoside Hydrolase Family 130 from Dyadobacter Fermentans.” *FEBS Letters* 589 (23): 3604–10. https://doi.org/10.1016/j.febslet.2015.10.008.

Nihira, Takanori, Erika Suzuki, Motomitsu Kitaoka, Mamoru Nishimoto, Ken’ichi Ohtsubo, and Hiroyuki Nakai. 2013. “Discovery of β-1,4-D-Mannosyl-N-Acetyl-D-Glucosamine Phosphorylase Involving in the Metabolism of N-Glycans.” *Journal of Biological Chemistry* 288 (38): 27366–74. https://doi.org/10.1074/jbc.M113.469080.

Revers, Leigh, Roslyn M. Bill, Iain B. H. Wilson, Gregory M. Watt, and Sabine L. Flitsch. 1999. “Development of Recombinant, Immobilised β-1,4-Mannosyltransferase for Use as an Efficient Tool in the Chemoenzymatic Synthesis of N-Linked Oligosaccharides.” *Biochimica et Biophysica Acta (BBA) - General Subjects* 1428 (1): 88–98. https://doi.org/10.1016/S0304-4165(99)00048-3.

Sasaki, Akiko, Takeshi Ishimizu, and Sumihiro Hase. 2005. “Substrate Specificity and Molecular Cloning of the Lily Endo-Beta-Mannosidase Acting on N-Glycan.” *Journal of Biochemistry* 137 (1): 87–93. https://doi.org/10.1093/jb/mvi008.

Schultz, Bárbara M., Carolina A. Paduro, Geraldyne A. Salazar, Francisco J. Salazar-Echegarai, Valentina P. Sebastián, Claudia A. Riedel, Alexis M. Kalergis, Manuel Alvarez-Lobos, and Susan M. Bueno. 2017. “A Potential Role of Salmonella Infection in the Onset of Inflammatory Bowel Diseases.” *Frontiers in Immunology* 8. https://doi.org/10.3389/fimmu.2017.00191.

Senoura, Takeshi, Shigeaki Ito, Hidenori Taguchi, Mariko Higa, Shigeki Hamada, Hirokazu Matsui, Tadahiro Ozawa, et al. 2011. “New Microbial Mannan Catabolic Pathway That Involves a Novel Mannosylglucose Phosphorylase.” *Biochemical and Biophysical Research Communications* 408 (4): 701–6. https://doi.org/10.1016/j.bbrc.2011.04.095.

Sernee, M. Fleur, Julie E. Ralton, Tracy L. Nero, Lukasz F. Sobala, Joachim Kloehn, Marcel A. Vieira-Lara, Simon A. Cobbold, et al. 2019. “A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in Leishmania Parasites.” *Cell Host & Microbe* 26 (3): 385-399.e9. https://doi.org/10.1016/j.chom.2019.08.009.

Sharma, Vandana, Mie Ichikawa, and Hudson H. Freeze. 2014. “Mannose Metabolism: More than Meets the Eye.” *Biochemical and Biophysical Research Communications*, Integrative Glycobiology and Future Perspectives, 453 (2): 220–28. https://doi.org/10.1016/j.bbrc.2014.06.021.

Shibata, Nobuyuki, Hidemitsu Kobayashi, Yoshio Okawa, and Shigeo Suzuki. 2003. “Existence of Novel β-1,2 Linkage-Containing Side Chain in the Mannan of Candida Lusitaniae, Antigenically Related to Candida Albicans Serotype A.” *European Journal of Biochemistry* 270 (12): 2565–75. https://doi.org/10.1046/j.1432-1033.2003.03622.x.

Sokol, Harry, Valentin Leducq, Hugues Aschard, Hang-Phuong Pham, Sarah Jegou, Cecilia Landman, David Cohen, et al. 2017. “Fungal Microbiota Dysbiosis in IBD.” *Gut* 66 (6): 1039–48. https://doi.org/10.1136/gutjnl-2015-310746.

Szilagyi, Andrew, Marvin Gerson, Jack Mendelson, and Nasim A. Yusuf. 1985. “Salmonella Infections Complicating Inflammatory Bowel Disease.” *Journal of Clinical Gastroenterology* 7 (3): 251–255.

Terrapon, Nicolas, Vincent Lombard, Harry J. Gilbert, and Bernard Henrissat. 2015. “Automatic Prediction of Polysaccharide Utilization Loci in Bacteroidetes Species.” *Bioinformatics* 31 (5): 647–55. https://doi.org/10.1093/bioinformatics/btu716.

Tsuda, Tomohiro, Takanori Nihira, Kazuhiro Chiku, Erika Suzuki, Takatoshi Arakawa, Mamoru Nishimoto, Motomitsu Kitaoka, Hiroyuki Nakai, and Shinya Fushinobu. 2015. “Characterization and Crystal Structure Determination of β-1,2-Mannobiose Phosphorylase from Listeria Innocua.” *FEBS Letters* 589 (24PartB): 3816–21. https://doi.org/10.1016/j.febslet.2015.11.034.

Usui, Taichi, Masahiro Suzuki, Toshinari Sato, Hirokazu Kawagishi, Kyoko Adachi, and Hiroshi Sano. 1994. “Enzymic Synthesis of the Trisaccharide Core Region of the Carbohydrate Chain OfN-Glycoprotein.” *Glycoconjugate Journal* 11 (2): 105–10. https://doi.org/10.1007/BF00731150.

Ye, Yuxin, Wataru Saburi, Rei Odaka, Koji Kato, Naofumi Sakurai, Keisuke Komoda, Mamoru Nishimoto, Motomitsu Kitaoka, Haruhide Mori, and Min Yao. 2016. “Structural Insights into the Difference in Substrate Recognition of Two Mannoside Phosphorylases from Two GH130 Subfamilies.” *FEBS Letters* 590 (6): 828–37. https://doi.org/10.1002/1873-3468.12105.

Zhao, Yongxin, and Jon S Thorson. 1999. “Chemoenzymatic Synthesis of the Salmonella Group E1 Core Trisaccharide Using a Recombinant β-(1→4)-Mannosyltransferase.” *Carbohydrate Research* 319 (1): 184–91. https://doi.org/10.1016/S0008-6215(99)00116-0.

Chapter 3

In the Chapter 1 of the Results section, I presented a novel approach based on sequence similarity networks analysis to analyze GP sequence diversity, and to highlight unexplored sequence spaces to mine for function discovery. The proof of concept of this approach was established with the GH130 family, which contains mannoside-acting enzymes.

In this chapter, I describe the application of this novel strategy to the analysis of five other CAZy families containing inverting GPs. In this study, I performed all the sequence analyses, except annotation of the GH65, GH94, GH112, GH149 and GH161 sequences in the targeted metagenomes, which was performed by Vincent Lombard, from the CAZy team in Marseille (France). Diego Morgavi, from INRAE Saint-Genès-Champanelle (France), provided us with the bovine rumen metagenomic dataset. Jeremy Esque helped me to manage large sequence datasets with the EFI-EST and Cytoscape tools. The figures presented in this chapter are not optimal, since because of the lockout due to the COVID pandemy, I could not finalize them with the computing facilities available at the Toulouse Biotechnology Institute.

Analysis of the diversity of the Glycoside Hydrolase Families 65, 94, 112, 149 and 161 in genomes and mammalian gut metagenomes

Introduction

Glycoside-phosphorylases (GPs) are very particular enzymes, which break down glycosides by transferring a glycosyl moiety from the non-reducing end of the substrate to inorganic phosphate, producing a glycosyl-1-phosphate and a shortened glycoside. GPs are also able to synthetize glycosides by reverse-phosphorolysis, from a glycosyl-1-phosphate and an appropriate hydroxylated acceptor, releasing inorganic phosphate. GPs are intracellular enzymes, requiring the cytoplasmic inorganic phosphate to perform phosphorolysis. They are widespread in bacteria, in the genomes of which the GP encoding genes often cluster in operon-like loci with genes encoding glycoside hydrolases and transporters, with which they participate to glycoside catabolism (Simon Ladevèze et al. 2017). It has also been suggested that some bacterial GPs may salvage polysaccharides to synthesize their own lipopolysaccharides or capsule polysaccharides (Nihira, Nakai, and Kitaoka 2012). In addition, in eukaryotes but not only, some GPs are involved in the intracellular recycling of energy storage or osmoregulator glycosides, such as trehalose (Schiraldi, Di Lernia, and De Rosa 2002), sucrose (Goedl et al. 2010), glycogen (Daghlas and Mohiuddin 2020), starch (E. O'Neill 2013) (ref), cyclic β-1,2-glucans (Cho et al. 2016), laminarin (Papaspyridi, Zerva, and Topakas 2018) and mannogen (Sernee et al. 2019).

GPs perform the phosphorolysis and reverse phosphorolysis reactions using the same catalytic features as glycoside-hydrolases (GHs) and glycosyltransferases (GTs), with which they share high sequence and mechanistic similarities (Puchart 2015). In the CAZy database, GPs are thus classified in GHs and GTs families. To date, there are 8 GH and 3 GT families containing GPs. GPs can be divided into retaining GPs (GH3, GH13_18, GT4, GT35) and inverting GPs (GH65, GH94, GH112, GH130, GH149, GH161, GT108) depending on whether the glycosidic bond is formed with retention or inversion of the anomeric configuration of the glycosyl-1-phosphate substrate, in the reverse-phosphorolysis reaction,

and reversely in the phosphorolysis reaction (Koshland 1953). There is much less functional diversity in retaining GPs than in inverting ones. Indeed, the specificity towards glycosyl-phosphates of retaining GPs acting on α-glycosides (GH13_18, GT4 and GT35) is limited to α-D-glucose-1-phosphate, and their linkage specificity to α-1,2, α-1,1 and α-1,4. The only retaining GPs acting on β-glycosides are the rare ones from the GH3 family, which are specific to β-D-glucose-1-phosphate or β-D-N-acetylglucosamine-1-phosphate, and to β-1,4 or β-1,3 glycosidic linkages. The remaining families contain inverting GPs, which change the anomeric configuration in the product, yielding α-glycosyl-phosphates from β-glycosides or β-glycosyl-phosphates from α-glycosides. The linkage and acceptor specificities of inverting GPs are very diverse, even intra family, even though only one type of glycosyl-phosphate is recognized by all the biochemically characterized GPs of each family (β-D-glucose-1-phosphate for GH65, α-D-glucose-1-phosphate for GH149 and GH161, α-D-mannose-1-phosphate for GH130 and GT108), except for GH94 GPs which act either on α-D-glucose-1-phosphate or on α-D-N-acetylglucosamine-1-phosphate. Recently, we used sequence similarity networks (SSNs) to analyze the diversity of GH130 genomic and metagenomic sequences, and to highlight unexplored sequence spaces, yielding to the discovery of a new enzymatic function in this family (A. Li et al. 2020).

In the present study, we used SSNs to analyze the sequence diversity of five CAZy families containing inverting GPs (GH65, GH94, GH112, GH149, GH161). The newly created GT108 family has not yet been included in this study. For each family, we analyzed the sequences of the CAZy database and those retrieved from four metagenomic datasets, selected from the gut of mammals, as ecosystems that are rich in glycosides of various origins and structures. By segregating the sequences into iso-functional groups, we isolated unexplored ones likely to offer functional novelty.

Material and Methods

GH sequence mining in mammal gut metagenomes

The GH sequences were identified from the metagenomic sequence datasets following a procedure previously described for other metagenomics analyses (Svartström et al. 2017). The procedure consisted of a BLAST (Altschul et al. 1990) search of the protein sequences against the full-length GH65, GH94, GH112, GH149 and GH161 sequences included in the CAZy database (http://www.cazy.org) in November 2019, using a cut-off E-value of 10^{-6}. Sequences that aligned over their entire length with one of the five family sequence in the database with >50% identity were directly assigned to the concerned family. The remaining sequences were subjected to a sequence similarity search using the Hidden Markov models (HMMER v3) built for each CAZy family, allowing identification of the CAZy family they belong to (Lombard et al. 2014).

Sequence similarity networks

The web-based Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (http://efi.igb.illinois.edu/efi-est) (Gerlt et al. 2015) was used independently in each family to construct sequence similarity networks (SSNs) from the sequences retrieved from the CAZy database in November 2019 and from the four metagenomes, with E-value thresholds between 10^{-120} and 10^{-320}. The node networks were visualized using Cytoscape 3.2 (https://cytoscape.org/ (Shannon et al. 2003)).

Analysis of sequence similarity

The CD-HIT program was used to analyze sequence redundancy within and between the datasets (http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi) (Huang et al. 2010).

Signal peptide detection

Signal peptides were predicted using the SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen 2017).

Results

CAZy sequence mining in mammalian gut microbiomes

Four mammalian gut metagenomes, totaling 33,962,473 genes, were searched for sequences containing at least one catalytic module assigned to one of the GH65, GH94, GH112, GH149 and GH161 families. Respectively, 1394, 7506, 1658, 404 and 374 sequences were retrieved from these metagenomes (Table 1). The relative number of sequences extracted from each dataset is in the same range, varying from 2.10^{-3} to 52.10^{-3} % of the genes in each metagenome for all families except family GH149, which is absent or rare in mammalian gut metagenomes (from 0 % in the mouse database to 4.10^{-5} % in the pig database). The proportion of truncated sequences, which lack N- and/or C-terminal ends, varies between 59.8 to 84.2% of the metagenomic sequences. Sequences of the families GH112, 149 and 161 are strictly monomodular, while maximum 1.9% of the metagenomic GH65 and GH94 sequences are multimodular (Table 2). Whatever the database or the family, very few sequences are fully identical to sequences of the CAZy database; the maximum is 30 sequences (0.1%) in family GH65. Among the metagenomic sequences, no GH112, GH149 and GH161 sequence was found in two or more metagenomes, and only two GH65 and six GH94 sequences are fully identical in several metagenomes. With a threshold of 90% sequence identity, sequence redundancy between metagenomes and the CAZy database was less than 11%, with no common sequence in the GH149 family. Between the metagenomes, it reaches 43.5% (Table 3).

Table 1: Diversity of sequences from the GH65, GH94, GH112, GH149 and GH161 families in mammalian gut metagenomes.

Metagenome	**Number of genes**	**N50 average contig length (kb)**	**Number of GH65 sequences**	**% of GH65**	**Number of full-length sequences**	**Number of multimodularsequences**	**Number of redundant sequences with those of the CAZy database**	
							100 % identity	**≥ 90 % identity**
Human	9,878,647	5.0	573	6.10^{-3}	318 (55.5%)	0	24	79
Mouse	7,685,872	6.6	132	2.10^{-3}	73 (55.3%)	0	3	22
Pig	2,572,074	1.87	195	8.10^{-3}	68 (34.9%)	1	3	8
Bovine	13,825,880	0.91	494	4.10^{-3}	102 (20.6%)	1	0	2
Total	33,962,473	--	1,394	--	561 (40.2%)	2 (0.1%)	30	111 (7.9%)
Redundancy between metagenomes							2 (0.1%)	306 (21.9%)

Metagenome	Number of genes	N50 average contig length (kb)	Number of GH94 sequences	% of GH94	Number of full-length sequences	Number of multimodular sequences	Number of redundant sequences with those of the CAZy database	
							100 % identity	≥ 90 % identity
Human	9,878,647	5.0	1,364	14.10-3	670 (49%)	12	9	87
Mouse	7,685,872	6.6	713	9.10-3	373 (52%)	25	0	6
Pig	2,572,074	1.87	1,349	52.10-3	424 (31%)	57	0	72
Bovine	13,825,880	0.91	4,080	30.10-3	710 (17%)	53	0	345
Total	33,962,473	--	7,506	--	2177 (29%)	147 (1.9%)	9	510 (6.8%)
Redundancy between metagenomes							6 (8.10-4%)	3270 (43.5%)

Metagenome	Number of genes	N50 average contig length (kb)	Number of GH112 sequences	% of GH112	Number of full-length sequences	Number of multimodular sequences	Number of redundant sequences with those of the CAZy database	
							100 % identity	≥ 90 % identity
Human	9,878,647	5.0	672	7.10^{-3}	142 (21%)	0	5	151
Mouse	7,685,872	6.6	441	6.10^{-3}	99 (22%)	0	0	9
Pig	2,572,074	1.87	221	9.10^{-3}	75 (34%)	0	0	16
Bovine	13,825,880	0.91	324	2.10^{-3}	113 (35%)	0	0	7
Total	33,962,473	--	1,658	--	429 (25.9%)	0 (%)	5	183 (11%)
Redundancy between metagenomes							0 (%)	702 (42.3%)

Metagenome	Number of genes	N50 average contig length (kb)	Number of GH149 sequences	% of GH149	Number of full-length sequences	Number of multimodular sequences	Number of redundant sequences with those of the CAZy database	
							100 % identity	≥ 90 % identity
Human	9,878,647	5.0	4	4.10^{-5}	2 (50%)	0	0	0
Mouse	7,685,872	6.6	0	0	0	0	0	0
Pig	2,572,074	1.87	67	$2.6.10^{-3}$	27 (40.3%)	0	0	0
Bovine	13,825,880	0.91	333	$2.4.10^{-3}$	35 (10.5%)	0	0	0 (%)
Total	33,962,473	--	404	--	64 (15.8%)	0 (%)	0	0 (%)
Redundancy between metagenomes							0 (%)	126 (31.2%)

Metagenome	Number of genes	N50 average contig length (kb)	Number of GH161 sequences	% of GH161	Number of full-length sequences	Number of multimodular sequences	Number of redundant sequences with those of the CAZy database	
							100 % identity	≥ 90 % identity
Human	9,878,647	5.0	102	12.10^{-3}	36 (35.3%)	0	1	2
Mouse	7,685,872	6.6	33	7.10^{-3}	23 (69.7%)	0	0	0
Pig	2,572,074	1.87	46	26.10^{-3}	7 (15.2%)	0	0	0
Bovine	13,825,880	0.91	193	17.10^{-3}	14 (7.2%)	0	0	2
Total	33,962,473	--	374	--	80 (21.4%)	0 (%)	1	4 (1%)
Redundancy between metagenomes							0 (%)	70 (18.7 %)

Table 2: List of the multi-modular sequences containing a GH65 or a GH94 module, and their location in the SSN meta-nodes.

Sequence ID	Modularity	SSN Meta-nodes
GH65		
Pig metagenome		
PIG_031_GL0134317	GH65-GH123	UC1
Rumen metagenome		
552_GL0092692	GH39-GH50-GH65	Singleton
GH94		
Human metagenome		
MH0028_GL0003057	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
764487809_stool1_revised_scaffold5909_1_gene196520	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MH0174_GL0056037	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
158337416_stool1_revised_C1288134_1_gene152125	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MH0355_GL0196938	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MH0371_GL0043287	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MH0229_GL0068780	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MH0373_GL0036976	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
O2_UC44_2_GL0183344	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
V1_FI17_GL0166068	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MH0149_GL0070226	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
O2_UC47_1_GL0061816	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
Mouse metagenome		
G1_3A_GL0084201	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
8_6_GL0008594	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
8_7_GL0089744	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
S_Fe9_GL0201599	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MC_6_2_GL0007599	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
S_Fe10_GL0136174	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MC_6_3_GL0000468	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MC_6_4_GL0003412	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MC_6_5_GL0004986	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MH_6_3_GL0034105	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
MH_6_4_GL0094834	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
Group2_3A_GL0144932	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
Group2_7A_GL0098350	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
1_3_GL0036569	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
28_GL0027913	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
2B_dyr23_07_GL0045266	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
37_GL0058896	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
4_7_GL0009465	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
6_1_GL0015423	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
7_2_GL0090807	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
S_Fe12_GL0120881	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
16_GL0012642	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
2A_dyr16_07_GL0030507	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
4_8_GL0013147	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
S_Fe8_GL0189988	GH94-GH31	C7(GPs-Glc-β-1,4-$(Glc)_n$_2)

Sequence ID	Modularity	SSN Meta-nodes
GH94		
Pig metagenome		
PIG_004_GL0051592	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_011_GL0243393	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_015_GL0289875	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_019_GL0060800	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_020_GL0136562	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_022_GL0064627	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_023_GL0070918	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_028_GL0050292	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_030_GL0073751	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_041_GL0117714	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_046_GL0080949	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_053_GL0083850	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_053_GL0167626	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_056_GL0132660	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_058_GL0013445	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_059_GL0010219	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_131_GL0151701	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BHZ_10B_GL0129047	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BHZ_11B_GL0022917	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BHZ_12B_GL0224971	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BHZ_3B_GL0141403	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BHZ_5B_GL0055499	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BMZ_11B_GL0068516	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BMZ_4B_GL0076575	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BMZ_6B_GL0056378	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BMZ_7B_GL0075281	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
DB_11B_GL0000689	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
DB_11B_GL0016812	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
DB_511B_GL0007200	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
DB_512B_GL0090269	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
EYZ_183B_GL0034834	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
EYZ_362B_GL0236707	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
EYZ_378B_GL0161831	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
EYZ_652B_GL0251890	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
SYZ_423B_GL0029902	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
ZXZ_10B_GL0052922	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
ZXZ_10B_GL0061714	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
ZXZ_4B_GL0003134	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_154_GL0132841	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_186_GL0052376	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_116_GL0211365	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_116_GL0249540	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_153_GL0190531	GT84-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_002_GL0090830	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_014_GL0084799	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_023_GL0211078	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_123_GL0211566	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
PIG_079_GL0011828	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BHZ_10B_GL0250266	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
BMZ_4B_GL0126628	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)

Sequence ID	Modularity	SSN Meta-nodes
GH94		
Pig metagenome		
BMZ_5B_GL0010348	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
BMZ_9B_GL0154221	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
DB_510B_GL0207906	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
DB_512B_GL0201683	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
EYZ_120B_GL0209766	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
SYZ_533B_GL0049271	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
ZXZ_4B_GL0101816	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
Bovine metagenome		
554_GL0968146	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
554_GL2815684	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
553_GL0596946	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
555_GL0467606	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
555_GL0567853	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
555_GL2263323	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
553_GL0824374	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
0081_GL0053622	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
0081_GL0177273	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
553_GL0895504	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
553_GL0936267	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
0081_GL2154069	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
553_GL0083189	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
100058_GL1678326	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
100058_GL3066203	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
100058_GL3167091	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
100058_GL3183669	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
100058_GL3298800	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
100058_GL3568686	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
2009040_GL1299512	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
2009040_GL1456528	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
2009040_GL1590182	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
2009040_GL1844098	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
2009040_GL2277690	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
3042_GL1652457	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
7049_GL0175957	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
7049_GL0191990	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
7049_GL0676236	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
7049_GL1871656	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
553_GL0181451	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
553_GL0206120	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
552_GL1372664	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
583_GL0085070	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
583_GL0656402	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
583_GL0733471	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
554_GL0086158	GT84-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
554_GL2456883	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
553_GL0040713	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
0081_GL1012976	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
0081_GL2147646	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
100058_GL1026702	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)
100058_GL2230308	GT84-GH144-GH94	C2(GSs-Glc–β–1,2-$(Glc)_n$)

Sequence ID	Modularity	SSN Meta-nodes
GH94		
Bovine metagenome		
2009040_GL1820487	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
2009040_GL1820769	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
3042_ GL0469036	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
3042_GL2774413	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
583_GL0256027	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
583_GL0638438	GT84-GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
555_GL1726641	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
0081_GL1064759	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
100058_GL2367730	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
552_GL0709473	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)
7049_ GL0638626	GH144-GH94	C2(GSs-Glc−β−1,2-$(Glc)_n$)

Table 3 Redundancy of metagenomic sequences. The redundancy of the GH130 sequences was previously analyzed (A. Li et al. 2020).

Families	Number of CAZy sequences (Nov.2019)	Number of metagenomic sequences	Redundancy (ID ≥ 90%)	
			Metagenomes/CAZy	Within metagenomes
GH65	5,871	1,394	7.5 %	21.9 %
GH94	2,741	7,506	6.8 %	43.5 %
GH112	326	1,658	11 %	42.3 %
GH130	1,594(Nov.2018)	4,714	10.7 %	51.5 %
GH149	122	404	0	31.2 %
GH161	96	374	1 %	18.7 %

Sequence diversity analysis

The analysis of sequence diversity was done by the construction of Sequence Similarity Networks (SSN) for each family, including all sequences from the CAZy database and from the four target metagenomes. As it was shown in the previous study on the GH130 family, the identical sequences do not affect the clusterization, insofar as identical sequences are grouped in the same node. When the number of nodes containing sequences with 100% identity was higher than 5000, we chose to group in the same node the sequences sharing more than 70% identity, in order to be able to use the EFI-ST webserver. This procedure was applied to the family GH94, containing 10,247 non-redundant sequences. For the family GH65, the procedure was not applied due to the great number (in total 2,389) of sequences from the CAZy database sharing 100% identity, leading to an initial number of nodes lower than 5,000. In addition, very few sequences are multimodular (≤ 1,9 %), which should not impact the clustering, as previously shown for the GH130 family. To construct the SSNs, we tested different E-value thresholds (between 10^{-120} and 10^{-320}), and mapped the functions of the characterized enzymes on the obtained meta-nodes. For families GH65, GH94 and GH112, we obtained isofunctional clusters that are defined as meta-nodes containing characterized members with the same osidic linkage and acceptor specificities. In these cases, the optimal E-value threshold was defined as the highest threshold yielding to the separation of the sequences into the maximum number of isofunctional meta-nodes. Two families, GH149 and GH161, have only two characterized members, which have the same substrate and linkage specificities. For these families, the optimal E-value threshold was defined as the lowest threshold that allows maintenance of characterized members in the same meta-nodes. Finally, in the optimal configuration, sequences of each meta-nodes were retrieved to search for the presence of signal peptide, indicating the presence of putative extracellular or membrane-bound CAZymes, which could not be GPs, since GPs need intracellular inorganic phosphate to perform glycoside phosphorolysis (A. Li et al. 2020).

Family GH65

The GH65 family lists 5,871 sequences in the CAZy database, of which 24 are characterized (Table 4). The family contains mainly GPs but eight of the characterized members from eukaryotic organisms are hydrolases (one α-glucosyl-1,2-β-galactosyl-L-hydroxylysine α-glucosidase (Hamazaki and Hotta 1979) and seven α-trehalases (Destruelle, Holzer, and Klionsky 1995; D'Enfert and Fontaine 1997; Eck et al. 1997; Zhao et al. 2007; Sánchez-Fresneda et al. 2015; Zilli et al. 2015). In reverse-phosphorolysis, the unique known donor is β-D-glucose 1-phosphate (βGlc1P), while four different acceptors are found in this family (D-glucose, L-rhamnose, glucose-6-phosphate and glycerol), with various linkage specificities when glucose is the acceptor. GPs of the GH65 family breakdown the disaccharides maltose (Glc-α-1,4-Glc) (Ehrmann and Vogel 1998; Inoue, Yasutake, et al. 2002; Y. Hidaka et al. 2005; Andersen et al. 2012; Nihira, Saito, Kitaoka, Otsubo, et al. 2012; Nihira, Nishimoto, et al. 2014), kojibiose (Glc-α-1,2-Glc) (Chaen et al. 1999; Yamamoto et al. 2004; 2011; Mukherjee, Narindoshvili, and Raushel 2018), trehalose (Glc-α-1,1-Glc) (Nishimoto et al. 1996; Inoue, Ishii, et al. 2002) and nigerose (Glc-α-1,3-Glc) (Nihira, Nakai, et al. 2012). Others characterized enzymes are trehalose 6-phosphate phosphorylase, acting on Glc-α-1,1-Glc6P (Andersson, Levander, and Rådström 2001); 3-O-α-D-glucosyl-L-rhamnose phosphorylase, acting on Glc-α-1,3-Rha (Nihira, Nakai, and Kitaoka 2012) and 2-O-α-glucopyranosylglycerol phosphorylase, acting on Glc-α-1,2-Glycerol (Nihira, Saito, et al. 2014; Touhara et al. 2014) . The SSN analysis of the family GH65 was performed on a total of 7,265 sequences. The nine different activities appear in distinct meta-nodes at the optimal configuration, corresponding to E-value 10^{-180}, with a pairwise sequence identity over 45% (Figure 1). Only the hydrolases acting on Glc-α-1,1-Glc (trehalases) are found in two meta-nodes. Previously, the phylogenetic tree of characterized GH65 enzymes also showed a significant distance between two branches containing trehalases (Nihira, Nakai, and Kitaoka 2012; 2012). In this configuration, we obtained seven meta-nodes containing at least one characterized sequence (C1 to C7), of which two host characterized hydrolases and sequences with a signal peptide (C6, C7). Seven meta-nodes (UC1 to UC7) contain more than 20 uncharacterized sequences (Table 5). Signal peptides were found in some sequences of the UC1 and UC4 meta-nodes, indicating that these meta-nodes probably contain hydrolase sequences. Among the five other uncharacterized meta-nodes, UC3, UC6 and UC7 mainly contain CAZy sequences, while UC2 and UC5 contain mainly metagenomic sequences. The 1,036 remaining sequences are distributed in small meta-nodes (less than 20 sequences), of which three contain one characterized sequence, and in 778 singletons. In singletons, 78,4% of sequences are truncated, while 97.4% of full length sequences are distributed in the meta-nodes (Table 14).

Table 4: List of the biochemically characterized GH65 sequences and their location in the SSN meta-nodes.

GH65_GPs	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
			Donor	Acceptor	Product	
C1(GPs-Glc-α-1,4-Glc)	AAV43670.1	maltose phosphorylase (MalP;LBA1870)	βGlc1P	D-glucose	Maltose (α-D-glucosyl-(1→4)-D-glucose)	(Andersen et al. 2012)

GH65_GPs	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
			Donor	Acceptor	Product	
C1(GPs-Glc-α-1,4-Glc)	ADH99560.1	maltose phosphorylase (Bsel_2056)	βGlc1P	D-glucose	Maltose	(Nihira, Saito, Kitaoka, Otsubo, et al. 2012)
	BAC54904.1	maltose phosphorylase (MPase)	βGlc1P	D-glucose	Maltose	(Inoue, Yasutake, et al. 2002)
	BAD97810.1	maltose phosphorylase (MapA)	βGlc1P	D-glucose	Maltose	(Y. Hidaka et al. 2005)
	CAA11905.1	maltose phosphorylase	βGlc1P	D-glucose	Maltose	(Ehrmann and Vogel 1998)
	AAO80764.1	maltose phosphorylase (MalP;EF0957)	βGlc1P	D-glucose	Maltose	(Nihira, Nishimoto, et al. 2014)
	Q7SIE1	maltose phosphorylase	βGlc1P	D-glucose	Maltose	(Hüwel et al. 1997)
C2(GPs-Glc-α-1,1-Glc)	BAB97299.1	trehalose phosphorylase (TreP)	βGlc1P	D-glucose	trehalose (α-D-glucosyl-(1→1)-D-glucose)	(Nishimoto et al. 1996)
	BAC20640.1	trehalose phosphorylase (TPase)	βGlc1P	D-glucose	trehalose	(Inoue, Ishii, et al. 2002)
C3(GPs-Glc-α-1,1-Glc6P)	AAK04526.1	trehalose-6-phosphate phosphorylase (TrePP;YeeA;L39593;LL0428)	βGlc1P	D-glucose 6-P	trehalose 6-P (α-D-glucosyl-(1→1)-D-glucopyranoside 6-P)	(Andersson, Levander, and Rådström 2001)
C4(GPs-Glc-α-1,2-Glc)	AAE30762.1	kojibiose phosphorylase (KojP;KPase)	βGlc1P	D-glucose	kojibiose (α-D-glucosyl-(1→2)-D-glucose)	(Chaen et al. 1999; Yamamoto et al. 2004)
	AAC74398.1	kojibiose phosphorylase(YcjT)	βGlc1P	D-glucose	kojibiose	(Mukherjee, Narindoshvili, and Raushel 2018)
	ABP66077.1	kojibiose phosphorylase (CsKP;Csac_0439)	βGlc1P	D-glucose	kojibiose	(Yamamoto et al. 2011)
C5(GPs-Glc-α-1,3-Glc)	ABX42243.1	nigerose phosphorylase (Cphy_1874)	βGlc1P	D-glucose	nigerose (α-D-glucosyl-(1→3)-D-glucose)	(Nihira, Nakai, et al. 2012)
Small meta-nodes (GPs-Glc-α-1,3-Rha)	ABX41399.1	3-O-α-glucopyranosyl-L-rhamnose phosphorylase (Cphy_1019)	βGlc1P	L-rhamnoe	α-D-glucopyranosyl-(1→3)-L-rhamnose	(Nihira, Nakai, and Kitaoka 2012)
Small meta-nodes (GPs-Glc-α-1,2-glycerol)	ADI00307.1	1,2-alpha-glucosylglycerol phosphorylase(Bsel_2816)	βGlc1P	D-glycerol	α-D-glucopyranosyl-(1→2)-glycerol	(Nihira, Saito, et al. 2014)
GH65_GHs	**GenBank/ UniProtKB accession**	**Protein Name**	**Hydrolase activity**			**Reference**
			Substrate		**Product**	
C6(GHs-Glc-α-1,1-Glc)	AAV05390.1	acid trehalase (Atc1;CaATC1)	trehalose		D-glucose	(Eck et al. 1997)
	AGG12634.1	acid trehalase (Ath1;CgATH1;CAGLOK05137g)	trehalose		D-glucose	(Zilli et al. 2015)
	CCE43253.1	trehalase (Atc1;CpATC1;CPAR2_208980)	trehalose		D-glucose	(Sánchez-Fresneda et al. 2015)
	CAA89280.1	acid trehalase (vacuolar) (Ath1;YPR026w;YP9367.06)	trehalose		D-glucose	(Destruelle, Holzer, and Klionsky 1995)

GH65_GHs	GenBank/ UniProtKB accession	Protein Name	Hydrolase activity		Reference
			Substrate	Product	
C7(GHs-Glc-α-1,1-Glc)	AAB57642.1	acid trehalase (TreA)	trehalose	D-glucose	(D'Enfert and Fontaine 1997)
	EAA66407.1	AN9340.2 (TreA)	trehalose	D-glucose	(D'Enfert and Fontaine 1997)
	ABO93464.1	acid trehalase (Atm1)	trehalose	D-glucose	(Zhao et al. 2007)
Small meta-nodes(α-glucosyl-1,2-β-galactosyl-L-hydroxylysine α-glucosidase)	BAR88294.1	protein α-glucosyl-1,2-β-galactosyl-L-hydroxylysine α-glucosidase (Athl1;PGGHG)	[collagen]-(5R)-5-O-[α-D-glucosyl-(1→2)- β-D-galactosyl]-5-hydroxy-L-lysine	D-glucose + [collagen]-(5R)-5-O-(β-D-galactosyl)-5-hydroxy-L-lysine	(Hamazaki and Hotta 1979)

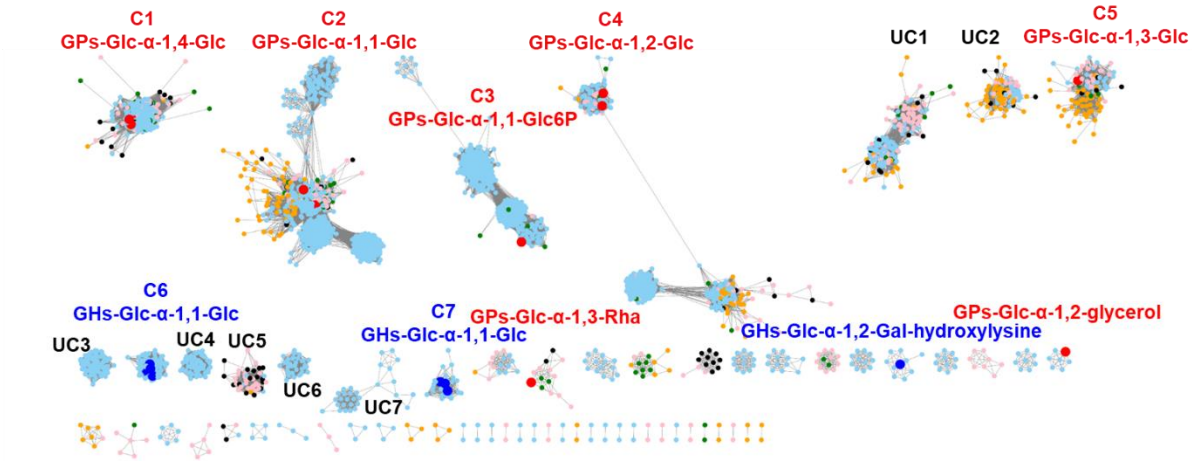


Figure 1: Protein sequence similarity network (SSN) of the 7,265 GH65 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-180}. Each of the 4,866 nodes contains sequences with more than 100% identity. Nodes are connected by an edge when the pairwise sequence identity is over 45%. In this figure, the 778 sequences appearing in 724 singletons have been omitted. The meta-nodes containing at least 20 sequences and with characterized members are labelled as C1 to C7. Red nodes contain biochemically characterized GPs, blue nodes contain biochemically characterized GHs. The meta-nodes containing at least 20 sequences and with no characterized members are labelled as UC1 to UC7. Nodes are colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; green, mouse gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome.

Table 5: Distribution of the predicted GH65 glycoside-phosphorylases and hydrolases in the SSN meta-nodes, obtained with an E-value threshold of 10^{-180}. The number of sequences includes the full-length and truncated ones. Only the sequences belonging to meta-nodes with at least 20 sequences are mentioned.

GH65	C1 (GPs_Glc-α-1,4-Glc)	C2 (GPs_Glc-α-1,1-Glc)	C3 (GPs_Glc-α-1,1 -Glc6P)	C4 (GPs_Glc-α-1,2-Glc)	C5 (GPs_Glc-α-1,3-Glc)	Total
Prediction of mechanism	**GPs**	**GPs**	**GPs**	**GPs**	**GPs**	**GPs**
CAZy	1,627	1,118	879	1,394	27	5,045
Human	81	40	5	48	24	198
Mouse	26	19	4	7	3	59
Pig	23	7	0	6	4	40
Rumen	5	55	0	27	43	130
Total	1,762	1,239	888	1,482	101	5,472

GH65	C6 (GHs_Glc-α-1,1-Glc)	C7 (GHs_Glc-α-1,1-Glc)	Total
Prediction of mechanism	**GHs**	**GHs**	**GHs**
CAZy	142	26	168
Human	0	0	0
Mouse	0	0	0
Pig	0	0	0
Rumen	0	0	0
Total	142	26	168

GH65	UC1	UC2	UC3	UC4	UC5	UC6	UC7	Total	Total
Prediction of mechanism	**GHs**	**GPs**	**GPs**	**GHs**	**GPs**	**GPs**	**GPs**	**GPs**	**GHs**
CAZy	94	28	102	66	4	47	30	211	160
Human	61	21	0	0	24	0	0	45	61
Mouse	7	0	0	0	0	0	0	0	7
Pig	10	6	0	0	15	0	0	21	10
Rumen	15	47	0	0	1	0	0	48	15
Total	187	102	102	66	44	47	30	325	253

Family GH94

The CAZy database lists 2,741 GH94 sequences, of which 27 are characterized (Table 6). All of them are GPs, targeting cellobiose (Glc-β-1,4-Glc) (Dakhova et al. 1993; Reichenbecher, Lottspeich, and Bronnenmeier 1997; Liu et al. 1998; Yernool et al. 2000; Y.-K. Kim et al. 2002; Nidetzky et al. 2004; M. Hidaka et al. 2006; Ha et al. 2013; Sawano et al. 2013; Xu et al. 2013; Devendran et al. 2016; Meng et al. 2018; S.-K. Kim et al. 2019) , cellodextrins (Glc-β-1,4-$(Glc)_n$) (Reichenbecher, Lottspeich, and Bronnenmeier 1997; Kawaguchi et al. 1998; Ha et al. 2013; Sawano et al. 2013; E. C. O'Neill et al. 2017), laminarinbiose (Glc-β-1,3-Glc) (Nihira, Saito, Kitaoka, Nishimoto, et al. 2012; Kuhaudomlarp, Walpole, et al. 2019), cellobionic acid (Glc-β-1,4-GlcOA) (Nihira, Saito, et al. 2013), sophorose (Glc-β-1,2-$(Glc)_n$) (Nakajima et al. 2014; 2017) and N,N'-diacetyl chitobiose (GlcNAc-β-1,4-GlcNAc) (Park, Keyhani, and Roseman 2000; M. Hidaka et al. 2004; Honda, Kitaoka, and Hayashi 2004). In addition, the GH94 family contains bacterial multi-domain sequences (GT84-GH94), including those of the cyclic β-1,2-glucan synthases, which have phosphorolytic activity towards β-1,2-glucooligosacccharides synthesized from UDP-glucose by the GT84 domain (Ciocchini et al. 2007). In the family GH94, the SSN analysis was applied to 10,247 sequences for which the optimal clusterization was obtained by setting the E-value at the threshold of 10^{-210}. The pairwise sequence identity in meta-nodes is over 52%. In the optimal configuration, the different enzyme specificities found in this family are well separated

(Figure 2). The GPs targeting Glc-β-1,4-Glc (cellobiose), Glc-β-1,4-$(Glc)_n$ (cellodextrins) and GPs GlcNAc-β-1,4-GlcNAc (chitobiose) appear in distinct meta-nodes. Nevertheless, the characterized sequences targeting Glc-β-1,4-$(Glc)_n$ (cellodextrins) are distributed in two large meta-nodes, appearing early at the E-value 10^{-120}, and in one small meta-nodes of less than 20 sequences (Figure 3). It was not possible to maintain cellodextrins-targeting GP sequences in one single meta-nodes while separating cellobiose- and chitobiose-targeting GPs. The separation of these three functions is obtained at the E-value threshold 10^{-210} (Figure 2). With this optimal E-value threshold, the GPs targeting Glc-β-1,3-Glc (laminaribiose) and Glc-β-1,4-$(Glc)_n$ (cellodextrins) are nevertheless each splitted in a large meta-nodes and in a small one of less than 20 sequences, each containing one characterized sequence of each specificity. Previously, the phylogenetic tree of characterized GH94 enzymes also showed a significant distance between the two cellodextrin phosphorylases and the two laminaribiose phosphorylases (Nihira, Saito, Kitaoka, Nishimoto, et al. 2012). The optimal SSN configuration consists of eight meta-nodes containing at least one characterized sequence: seven meta-nodes without any signal peptide in their sequences (C1, C3-C8) and the meta-nodes C2 containing the two characterized cyclic β-1,2-glucan synthases and three uncharacterized sequences harboring a signal peptide (Table 7). The meta-nodes C2 contains all the multimodular sequences (GH144-GH94, GT84-GH94, GT84-GH144-GH94), except the GH94-GH31 ones, which belong to the meta-nodes C7. In meta-nodes C2, the 3 sequences harboring a signal peptide are GT84-GH94 multimodular sequences, in which the C-terminal module could confer phosphorylase activity to these proteins. Indeed, the cyclic β-1,2-glucooligosacccharides are osmoregulated periplasmic bacterial glucans, of which the number of glucosyl residues in the cyclic ring is controlled by the C-terminal GH94 domain (conferring GP activity) of the GT84-GH94 multimodular β-1,2-glucan synthases (Ciocchini et al. 2007). In such cases the signal peptide would address the protein to the periplasmic compartment, which is still compatible with a phosphorolytic function. Interestingly, some multimodular sequences present a GH144 module, conferring an additional endo-β-1,2-glucanase activity to these exo-β-1,2-glucan phosphorylases (Abe et al. 2017). Finally, the optimal SSN configuration comprises 4 additional meta-nodes of more than 20 uncharacterized sequences, none with a signal peptide (UC1-UC4). The UC1, UC2 and UC3 mainly contain metagenomic sequences, while UC4 is composed of a balanced assembly of metagenomic and genomic sequences. The 4,130 remaining sequences are in singletons and in 47 small meta-nodes, of which two contain one characterized sequence, as explained above. Among the 3,894 sequences in singletons, 93.84 % are truncated, while 95.1 % of complete sequences are in the meta-nodes (Table 14).

Table 6: List of the biochemically characterized GH94 sequences and their location in the SSN meta-nodes.

GH94_GPs	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
			Donor	Acceptor	Product	
C1(GPs-Glc-β-1,4-Glc)	AAB95491.2	cellobiose phosphorylase (CbpA)	αGlc1P	D-glucose	cellobiose (β-D-glucopyranosyl-(1→4)-D-glucopyranose)	(Yernool et al. 2000)

GH94_GPs	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
			Donor	Acceptor	Product	
C1(GPs-Glc-β-1,4-Glc)	AAC45510.1	cellobiose phosphorylase (CepA)	αGlc1P	D-glucose	cellobiose	(Reichenbecher, Lottspeich, and Bronnenmeier 1997)
	AAD36910.1	cellobiose phosphorylase (CepA;TM1848;Tmari_1863)	αGlc1P	D-glucose	cellobiose	(S.-K. Kim et al. 2018)
	AAL67138.1	cellobiose phosphorylase (Cbp)	αGlc1P	D-glucose	cellobiose	(Y.-K. Kim et al. 2002)
	AAQ20920.1	cellobiose phosphorylase (CbP)	αGlc1P	D-glucose	cellobiose	(Nidetzky et al. 2004)
	ABD80580.1	cellobiose phosphorylase 94A (Cbp;SdCBP;Sde1318) (Cep94A)	αGlc1P	D-glucose	cellobiose	(Ha et al. 2013)
	ABN51514.1	cellobiose phosphorylase (Cbp;CtCBP;Cthe_0275)	αGlc1P	D-glucose	cellobiose	(Wilson et al. 2013)
	ACL76454.1	cellobiose phosphorylase (CbpA;Ccel_2109)	αGlc1P	D-glucose	cellobiose	(Xu et al. 2013)
	ADU20744.1	cellobiose phosphorylase (CBP;RaCBP;Rumal_0187)	αGlc1P	D-glucose	cellobiose	(Sawano et al. 2013)
	BAA25846.1	cellobiose-phosphorylase (Cbp)	αGlc1P	D-glucose	cellobiose	(Cate et al. 2014)
	BAA28631.1	cellobiose phosphorylase (Cbp;CgCBP)	αGlc1P	D-glucose	cellobiose	(Liu et al. 1998)
	CAB16926.1 *	cellobiose phosphorylase (CepA)	αGlc1P	D-glucose	cellobiose	(Dakhova et al. 1993)
C3(GPs-Glc-β-1,4-$(Glc)_n$_1)	ADU22883.1	cellodextrin phosphorylase (CDP;RaCDP;Rumal_2403)	αGlc1P	$[(1\rightarrow4)$-β-D-glucosyl$]_n$	cellodetrin $[(1\rightarrow4)$-β-D-glucosyl$]_{n+1}$	(Sawano et al. 2013)
C4(GPs-Glc-β-1,3-Glc)	BAJ10826.1	laminaribiose phosphorylase (LbpA)	αGlc1P	D-glucose	laminaribiose (β-D-glucopyranosyl-(1→3)-D-glucopyranose)	(Kuhaudomlarp, Walpole, et al. 2019)
C5(GPs-GlcNAc-β-1,4-GlcNAc)	AAG23740.1	diacetylchitobiose phosphorylase (ChbP)	αGlcNAc1P	N-acetyl-D-glucosamine	N,N'-diacetylchitobiose (N-acetyl-β-D-glucosaminyl-(1→4)-N-acetyl-D-glucosamine)	(Keyhani, Li, and Roseman 2000)
	BAC87867.1	chitobiose phosphorylase (ChbP;VpChBP)	αGlcNAc1P	N-acetyl-D-glucosamine	N,N'-diacetylchitobiose	(Honda, Kitaoka, and Hayashi 2004)
C6(GPs-Glc-β-1,2-$(Glc)_n$)	CAC97070.1	β-1,2-oligoglucan phosphorylase (LiSOGP;Lin1839)	αGlc1P	$[(1\rightarrow2)$-β-D-glucosyl$]_n$	$[(1\rightarrow2)$-β-D-glucosyl$]_{n+1}$	(Nakajima et al. 2014)
	ABX41081.1	β-1,2-oligoglucan phosphorylase (LpSOGP;Cphy_0694)	αGlc1P	$[(1\rightarrow2)$-β-D-glucosyl$]_n$	$[(1\rightarrow2)$-β-D-glucosyl$]_{n+1}$	(Nakajima et al. 2017)
C7(GPs-Glc-β-1,4-$(Glc)_n$_2)	BAB71818.1	cellodextrin phosphorylase (Cdp-ym4)	αGlc1P	$[(1\rightarrow4)$-β-D-glucosyl$]_n$	$[(1\rightarrow4)$-β-D-glucosyl$]_{n+1}$	(Sawano et al. 2013)
	ABN54185.1	cellodextrin-phosphorylase (Cdp;CtCDP;Cthe_2989)	αGlc1P	$[(1\rightarrow4)$-β-D-glucosyl$]_n$	$[(1\rightarrow4)$-β-D-glucosyl$]_{n+1}$	(Kawaguchi et al. 1998)
	ADZ85667.1	cellodextrin phosphorylase (Cdp;ClCDP;Clole_3989)	αGlc1P	$[(1\rightarrow4)$-β-D-glucosyl$]_n$	$[(1\rightarrow4)$-β-D-glucosyl$]_{n+1}$	(Ha et al. 2013)
C8(GPs-Glc-β-1,4-GlcOA)	AAM43298.1	cellobionic acid phosphorylase (CelAP;NdvB;XCC4077)	αGlc1P	D-gluconate	β-D-glucopyranosyl-(1→4)-D-gluconate	(Nihira, Saito, et al. 2013)
	ABD80168.1	cellobionic acid phosphorylase 94B (Cep94B;CBAP;SdCBAP;Sde_0906)	αGlc1P	D-gluconate	β-D-glucopyranosyl-(1→4)-D-gluconate	(Nihira, Saito, et al. 2013)
	EAA28929.1	cellobionic acid phosphorylase (CelAP;NdvB;NCU09425)	αGlc1P	D-gluconate	β-D-glucopyranosyl-(1→4)-D-gluconate	(Nihira, Saito, et al. 2013)

GH94_GPs	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
			Donor	Acceptor	Product	
Small meta-nodes	AAC45511.1	cellodextrin phosphorylase (CepB;CsCdP)	αGlc1P	[(1→4)-β-D-glucosyl]$_n$	[(1→4)-β-D-glucosyl]$_{n+1}$	(Reichenbecher, Lottspeich, and Bronnenmeier 1997)
Small meta-nodes	ABX81345.1	laminaribiose phosphorylase (ACL_0729;ACL0729)	αGlc1P	D-glucose	laminaribiose	(Nihira, Saito, Kitaoka, Nishimoto, et al. 2012)
GH94_GSs/ GPs	**GenBank/ UniProtKB accession**	**Protein Name**	**Reverse-phosphorolysis activity**			**Reference**
			Donor	**Acceptor**	**Product**	
C2(GSs-Glc-β-1,2-(Glc)$_n$)	ACD71661.1	cyclic β-1,2-glucan synthetase (Cgs)/ β-1,2-oligoglucan phosphorylase	αGlc1P	[(1→2)-β-D-glucosyl]$_n$	[(1→2)-β-D-glucosyl]$_{n+1}$	(Iñón de Iannino et al. 1998; Guidolin et al. 2009)

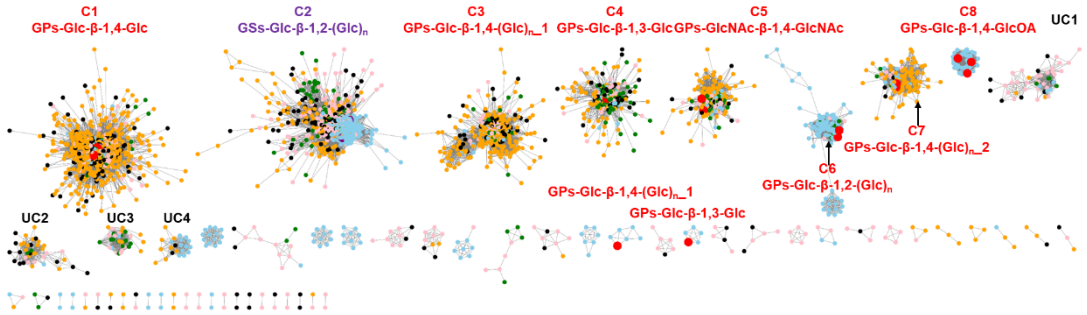


Figure 2: Protein sequence similarity network (SSN) of the 10,247 GH94 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-210}. Each of the 7,375 nodes contains sequences with more than 70% identity. Nodes are connected by an edge when the pairwise sequence identity is over 52%. In this figure, the 3,894 sequences appearing in 3,765 singletons have been omitted. The meta-nodes containing at least 20 sequences and with characterized members are labelled as C1 to C8. Red nodes contain biochemically characterized GPs; purple, biochemically characterized GSs. The meta-nodes containing at least 20 sequences and with no characterized members are labelled as UC1 to UC4. Nodes are colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; green, mouse gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome.

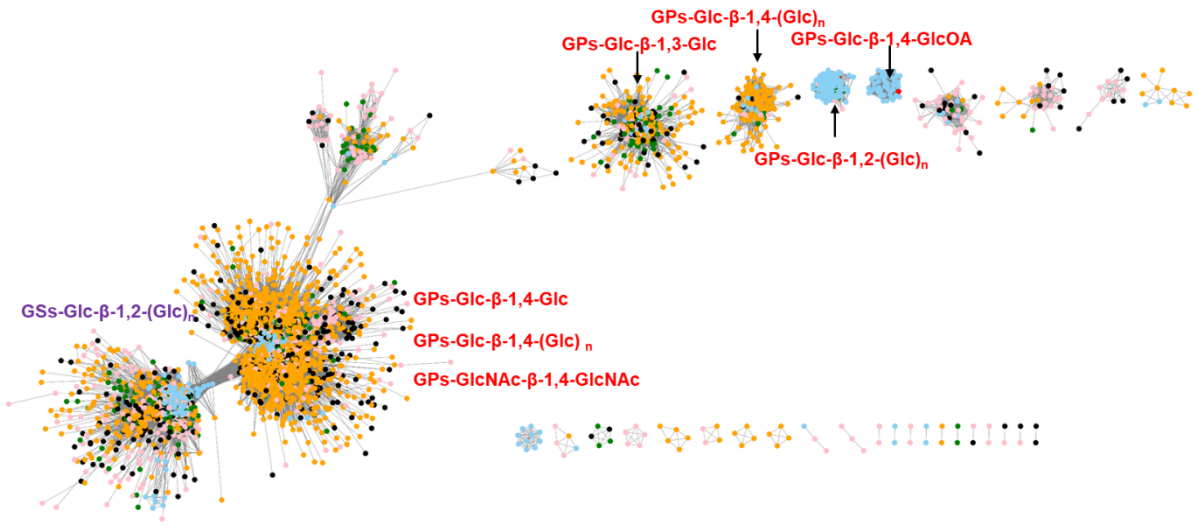


Figure 3: Protein sequence similarity network (SSN) of the 10,247 GH94 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-120}. Each of the 7,375 nodes contains sequences with more than 70% identity. Nodes are connected by an edge when the pairwise sequence identity is over 42%. In this figure, the 2,027 sequences appearing in 1965 singletons have been omitted. In red, biochemically characterized GPs; purple, biochemically characterized GSs. Nodes are colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; green, mouse gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome.

Table 7: Distribution of the predicted GH94 glycoside-phosphorylases and glucan synthases in the SSN meta-nodes, obtained with an E-value threshold of 10^{-210}. The number of sequences includes the full-length and truncated ones. Only the sequences belonging to meta-nodes with at least 20 sequences are mentioned.

GH94	**C1 (GPs-Glc-β-1,4-Glc)**	**C3 (GPs-Glc-β-1,4-(Glc)$_n$_1)**	**C4 (GPs-Glc-β-1,3-Glc)**	**C5 (GPs-GlcNAc-β-1,4-GlcNAc)**	**C6 (GPs-Glc-β-1,2-(Glc)n)**	**C7 (GPs-Glc-β-1,4-(Glc)$_n$_2)**	**C8 (GPs-Glc-β-1,4-GlcOA)**	**Total**	**C2 (GSs-Glc-β-1,2-(Glc)$_n$)**
Predicted mechanism	**GPs**	**GPs**	**GPs**	**GPs**	**GPs**	**GPs**	**GPs**	**GPs**	**GSs**
CAZy	204	60	67	276	409	21	265	1,302	1,236
Human	245	104	77	36	4	7	0	473	179
Mouse	103	88	107	57	12	4	0	371	57
Pig	305	127	59	24	0	7	0	522	106
Bovine	885	346	100	127	0	88	0	1,546	143
Total	1,742	725	410	520	425	127	265	4,214	1,721

GH194	UC1	UC2	UC3	UC4	Total
Predicted mechanism	GPs	GPs	GPs	GPs	GPs
CAZy	9	1	3	16	29
Human	35	6	14	0	55
Mouse	6	2	17	0	25
Pig	8	14	0	4	26
Bovine	3	20	10	8	41
Total	61	43	44	28	176

Family GH112

The CAZy database lists 326 sequences for the GH112 family, of which twelve are characterized. All of them are GPs acting on disaccharides (Table 8). In reverse phosphorolysis, the unique known donor is α-D-galactose 1-phosphate (αGal1P), and three different main acceptors have been reported: N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-galactosamine (GalNAc) and L-rhamnose (Rha). To date, two D-galactosyl-β-1,4-L-rhamnose phosphorylases are characterized (Nakajima, Nishimoto, and Kitaoka 2009b; 2010). In reverse-phosphorolysis, the two enzymes showed *in vitro* relaxed acceptor specificity, since both recognize also D-glucose, and one of them L-mannose, L-xylose, D-glucose, 2-deoxy-D-glucose and D-galactose (Nakajima, Nishimoto, and Kitaoka 2009b). Ten GH112 target β-1,3-galactosyl-N-acetylhexosamines (galacto-N-biose/lacto-N-biose phosphorylases) using GalNAc or GlcNAc acceptors in reverse-phosphorolysis (Nakajima et al. 2008; Nakajima and Kitaoka 2008; Nakajima, Nishimoto, and Kitaoka 2009a; Chao, Wim, and Tom 2011). Whatever is their preference for GlcNAc or GalNAc, these enzymes are relatively promiscuous and accept both acceptors. Based on the analysis of 1,984 sequences, optimal clusterization in family GH112 was obtained by setting an E-value threshold of 10^{-190}, corresponding to 48% pairwise sequence identity. We obtained two isofunctionnal meta-nodes, C1 and C2, in which no sequence has a signal peptide (Figure 4). The characterized GPs targeting Gal-β-1,3-GlcNAc and Gal-β-1,3-GalNAc appear in the same meta-nodes C1, these enzymes being promiscuous towards N-acetyl-hexosamine acceptors. When we set the E-value threshold at lower values, such as, for example, 10^{-220} (Figure 5), we scattered the characterized sequences into smaller meta-nodes and singletons without being able to separate the enzymes preferring GalNAc or GlcNAc acceptors. In the optimal SSN, the C2 meta-nodes contains the characterized GPs targeting Gal-β-1,4-Rha (Table 9). In this configuration, the sequences which are not classified in C1 or C2 appear in singletons or in three small meta-nodes, of size lower than the critical threshold of 20 sequences each. A total of 930 sequences (46.87%) are found in small meta-nodes and in singletons sharing less than 48% identity with a sequence in another meta-nodes (Table 14). Most of the sequences appearing in singletons are incomplete (95.98 %), while 95.1% of the full-length sequences cluster in meta-nodes. One sequence with a signal peptide (V1.FI14_GL0040405) was found in a small meta-nodes, indicating that it probably not codes for a GP (Figure 4, Table 9).

Table 8: List of the biochemically characterized GH112 sequences and their location in the SSN meta-nodes.

GH112_GPs	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
			Donor	Acceptor	Product	
C2 (GPs-β-1,4-Gal-Rha)	ACB74662.1	D-galactosyl-β-1,4-L-rhamnose phosphorylase (GalRhaP;Oter_1377)	αGal1P	L-rhamnose	β-D-galactosyl-(1→4)-L-rhamnose	(Nakajima, Nishimoto, and Kitaoka 2010)
	ABX42289.1	D-galactosyl-1,4-L-rhamnose phosphorylase (Cphy_1920)	αGal1P	L-rhamnose	β-D-galactosyl-(1→4)-L-rhamnose	(Nakajima, Nishimoto, and Kitaoka 2009b)
C1 (GPs-β-1,3-Gal-GlcNAc/ β-1,3-Gal-GalNAc)	AAO07997.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase (GalGlyNAcP;VV2_1091)	αGal1P	N-acetyl-D-glucosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine	(Nakajima and Kitaoka 2008)
	ABX40964.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase (Cphy_0577)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Nakajima, Nishimoto, and Kitaoka 2009b)
	ABX43387.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase (Cphy_3030)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Nakajima, Nishimoto, and Kitaoka 2009b)
	BAD80751.1	galacto-N-biose / lacto-N-biose phosphorylase (LnpA1;LnbP;GLNBP;BLLJ_1623)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Kitaoka, Tian, and Nishimoto 2005)
	BAD80752.1	lacto-N-biose phosphorylase (LnpA1;LnbP)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Kitaoka, Tian, and Nishimoto 2005)
	ACV29689.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase / galacto-N-biose/lacto-N-biose I phosphorylase (GLNBP;Apre_1669)	αGal1P	N-acetyl-D-glucosamine N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-glucosamine β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Chao, Wim, and Tom 2011)
	ZP_05748149.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase / galacto-N-biose phosphorylase (GNBP;HMPREF0357_1319)	αGal1P	N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Chao, Wim, and Tom 2011)
	ACZ00636.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase / galacto-N-biose phosphorylase (GNBP;Smon_0146)	αGal1P	N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Chao, Wim, and Tom 2011)
	ABG83511.1	galacto-N-biose phosphorylase (CPF_0553)	αGal1P	N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Nakajima et al. 2008)
	BAH10636.1	β-1,3-galactosyl-N-acetylhexosamine phosphorylase (GnpA)	αGal1P	N-acetyl-D-galactosamine	β-D-galactopyranosyl-(1→3)-N-acetyl-D-galactosamine	(Nakajima, Nishimoto, and Kitaoka 2009a)

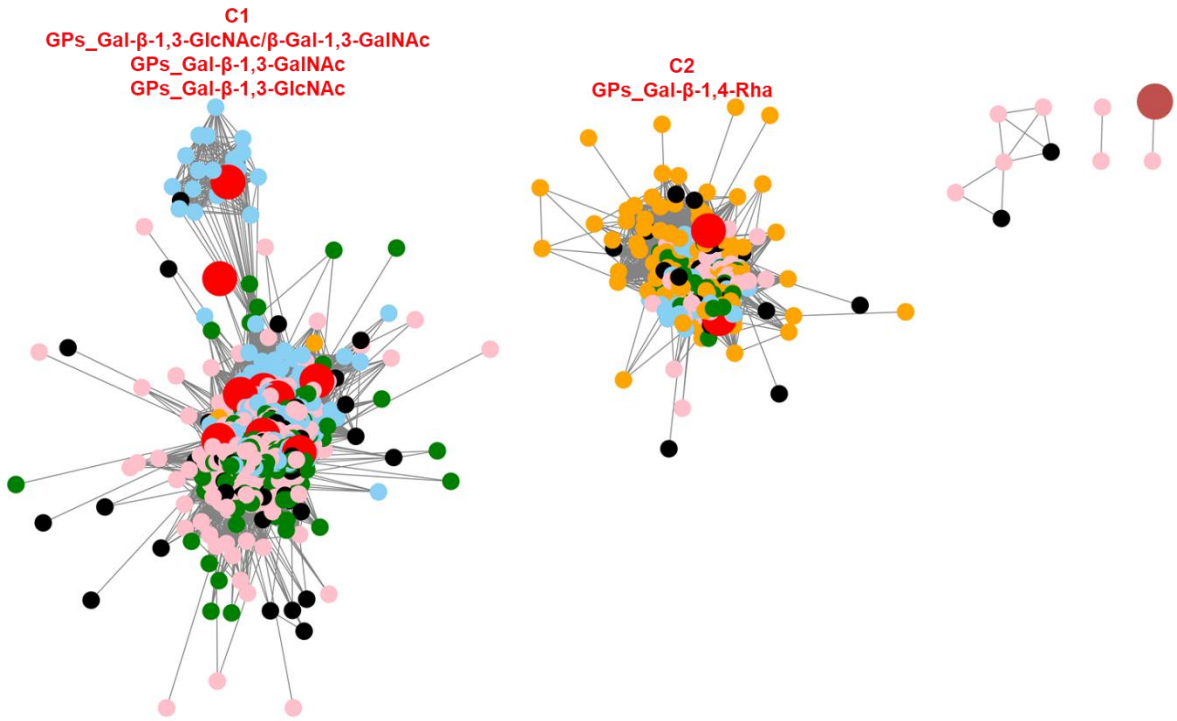


Figure 4: Protein sequence similarity network (SSN) of the 1,984 GH112 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-190}. Each of the 1,887 nodes contains sequences with more than 100% identity. Nodes are connected by an edge when the pairwise sequence identity is over 48%. In this figure, the 920 sequences appearing in 920 singletons have been omitted. The meta-nodes containing at least 20 sequences and with characterized members are labelled as C1 and C2. Red nodes contain biochemically characterized GPs. Brown node contains sequence with a signal peptide. Nodes are colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; green, mouse gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome.

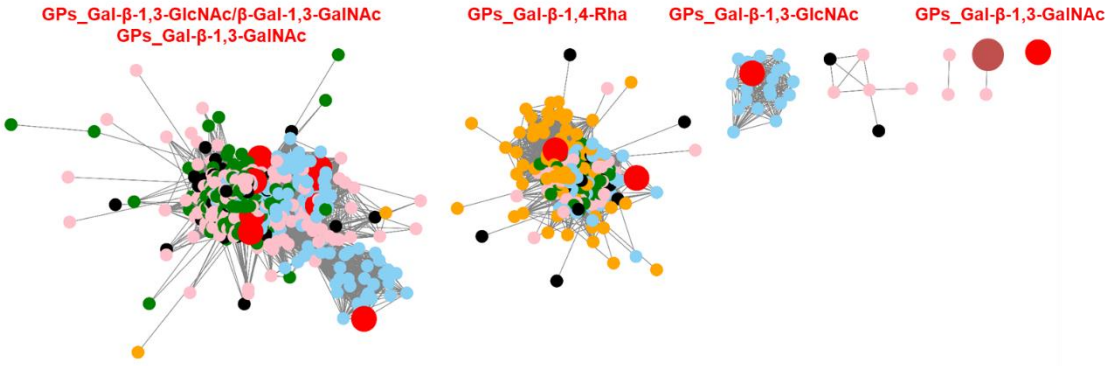


Figure 5: Protein sequence similarity network (SSN) of the 1,984 GH112 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-220}. Each of the 1,887 nodes contains sequences with more than 100% identity. Nodes are connected by an edge when the pairwise sequence identity is over 52%. In this figure, the 1009 sequences appearing in 1008 singletons have been omitted. In red, biochemically characterized GPs. Brown node contains sequence with a signal

peptide. Nodes are colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; green, mouse gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome.

Table 9: Distribution of the predicted GH112 glycoside-phosphorylases and hydrolases in the SSN meta-nodes, obtained with an E-value threshold of 10^{-190}. The number of sequences includes the full-length and truncated ones. Only the sequences belonging to meta-nodes with at least 20 sequences are mentioned.

GH112	C1(GPs-Gal-β-1,3 -GlcNAc/GalNAc)	C2(GPs-Gal-β-1,4-Rha)	Total
Prediction of mechanism	**GPs**	**GPs**	**GPs**
CAZy	276	40	316
Human	299	36	335
Mouse	189	33	222
Pig	63	21	84
Bovine	3	94	97
Total	830	224	1,054

Family GH149

The GH149 family is a small family, with only 122 sequences listed in the CAZy database. They are mostly issued from Gram-negative marine bacteria from the Proteobacteria and Bacteroidetes phyla. Two of them are characterized (Table 10). They target β-1,3-glucans (Kuhaudomlarp et al. 2018). They use glucose and longer β-1,3-linked gluco-oligosaccharides as acceptors in the reverse- phosphorolysis reaction, showing high flexibility towards β-1,3-glucan chain length and linkage regioselectivity (β-1,2; β-1,3; β-1,4 or β-1,6) (Kuhaudomlarp, Stevenson, et al. 2019). In the SSN analysis applied to the 526 sequences of family GH149, the optimal configuration was settled at the E-value threshold of 10^{-310}. The sequences are distributed in three meta-nodes, sharing each more than 48% pairwise sequence identity (Figure 6, Table 11). One (C1) contains the two characterized sequences. The remaining meta-nodes (UC1 and UC2) have no characterized member, but a signal peptide was detected in one sequence of UC2 (2009040_GL0562976). Others sequences are distributed in one small meta-nodes and in 316 singletons of which 303 (95,89 %) correspond to truncated sequences, while 93 % of the full-length sequences are found in meta-nodes (Table 14).

Table 10: List of the biochemically characterized GH149 sequences and their location in the SSN meta-nodes.

GH149_GPs	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
			Donor	Acceptor	Product	
C1(GPs-Glc-β-1,3-$(Glc)_n$)	/	β-1,3-glucan phosphorylase(Pro_7066)	αGlc1P	$[(1\rightarrow3)$-β-D-glucosyl$]_n$	$[(1\rightarrow3)$-β-D-glucosyl$]_{n+1}$	(Kuhaudomlarp et al. 2018)
	AUO30192.1	β-1,3-glucan phosphorylase(EgP1)	αGlc1P	$[(1\rightarrow3)$-β-D-glucosyl$]_n$	$[(1\rightarrow3)$-β-D-`glucosyl$]_{n+1}$	(Kuhaudomlarp et al. 2018)

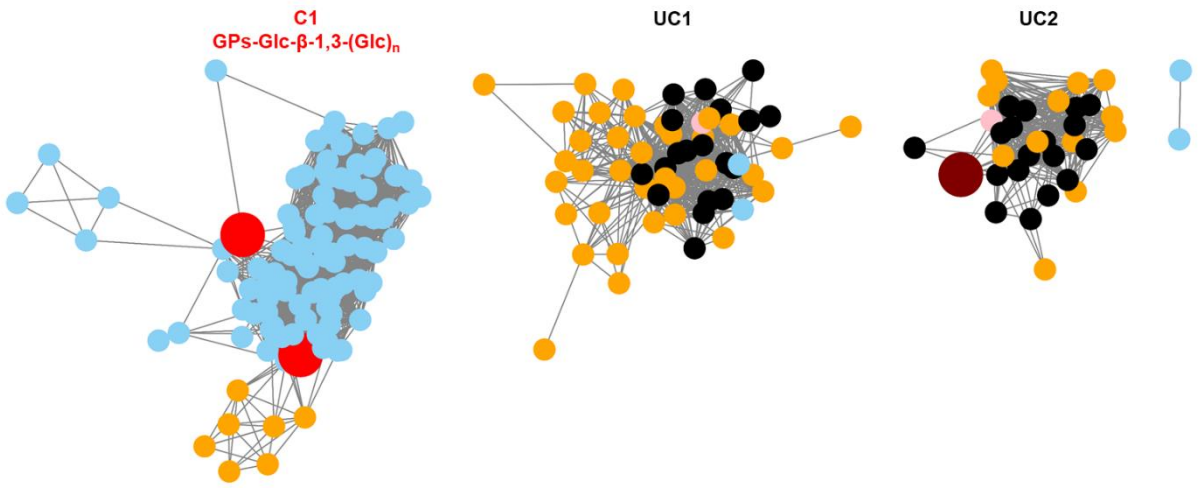


Figure 6: Protein sequence similarity network (SSN) of the 526 GH149 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-310}. Each of the 511 nodes contains sequences with more than 100% identity. Nodes are connected by an edge when the pairwise sequence identity is over 48%. In this figure, the 316 sequences appearing in 316 singletons have been omitted. The meta-nodes containing at least 20 sequences, with characterized members is labelled as C1 and with no characterized members are labelled UC1 and UC2. Red nodes contain biochemically characterized GPs. Brown node contains sequence with a signal peptide. Nodes are colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome.

Table 11: Distribution of the predicted GH149 glycoside-phosphorylases and hydrolases in the SSN meta-nodes, obtained with an E-value threshold of 10^{-310}. The number of sequences includes the full-length and truncated ones. Only the sequences belonging to meta-nodes with at least 20 sequences are mentioned.

GH149	**C1(GPs-Glc-β-1,3-Glc)$_n$)**	**UC1**	**Total**	**UC2**
Prediction of mechanism	**GPs**	**GPs**	**GPs**	**GHs**
CAZy	111	2	113	0
Human	0	1	1	1
Mouse	0	0	0	0
Pig	0	19	19	18
Bovine	7	35	42	14
Total	118	57	175	33

Family GH161

The CAZy database lists only 96 GH161 sequences, mainly from Gram positive bacteria, and, in a minor extend, from archae. One sequence also belongs to an eukaryote, the heterokont *Ochromonas danica* species. Two enzymes are characterized as β-1,3-glucan-phosphorylases (Table 12). In reverse-phosphorolysis, they use long gluco-oligosaccharides of various chain lengths as acceptors, but cannot use glucose. As the GH149 characterized members, they have high flexibility towards linkage position (Kuhaudomlarp, Pergolizzi, et al. 2019). For the 470 sequences of family GH161, we set the optimal configuration with an E-value threshold of 10^{-280}, corresponding to 44 % pairwise sequence identity (Figure 7). When the E-value threshold was settled at lower values, such as, for example, 10^{-290}, corresponding to 48% pairwise sequence identity (Figure 8), one characterized sequence was discarded in a singleton, without allowing the emergence of a new meta-nodes. In the selected SSN configuration, we obtained one single meta-nodes containing the two characterized β-1,3-glucan-phosphorylases (Table 13). In this unique meta-nodes, sequences are grouped in two entities, linked by a single node. The entity containing the two characterized enzymes is rich in sequences from the CAZy database, while the second is rich in metagenomic sequences. The remaining sequences are discarded in twelve small clusters and in singletons. Most of the 286 sequences appearing in singletons are incomplete (96,15%), while 93 % of full-length sequences are in meta-nodes (Table 14).

Table 12: List of the biochemically characterized GH161 sequences and their location in the SSN meta-nodes.

GH161_GPs	GenBank/ UniProtKB accession	Protein Name	Reverse-phosphorolysis activity			Reference
			Donor	Acceptor	Product	
C1(GPs-Glc-β-1,3-(Glc)$_n$	WP_019688419.1	β-1,3-glucan phosphorylase (PPT_RS0121460; PapP)	αGlc1P	[(1→3)-β-D-glucosyl]$_n$	[(1→3)-β-D-glucosyl]$_{n+1}$	(Kuhaudomlarp, Pergolizzi, et al. 2019)
	ACJ76363.1	β-1,3-glucan phosphorylase (THA_1941; TaCDP)	αGlc1P	[(1→3)-β-D-glucosyl]$_n$	[(1→3)-β-D-glucosyl]$_{n+1}$	(Kuhaudomlarp, Pergolizzi, et al. 2019)

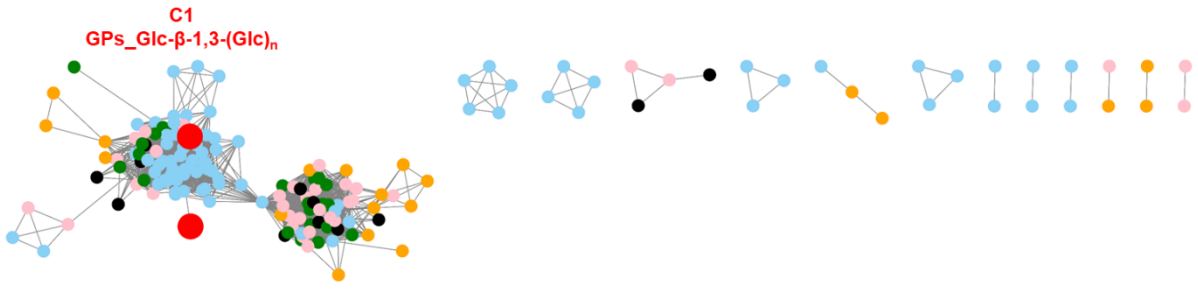


Figure 7: Protein sequence similarity network (SSN) of the 470 GH161 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-280}. Each of the 462 nodes contains sequences with more than 100% identity. Nodes are connected by an edge when the pairwise sequence identity is over 44%. In this figure, the 286 sequences appearing in 286 singletons have been omitted.

The meta-nodes containing at least 20 sequences, with characterized members is labelled as C1. Red nodes contain biochemically characterized GPs. Nodes are colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; green, mouse gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome.

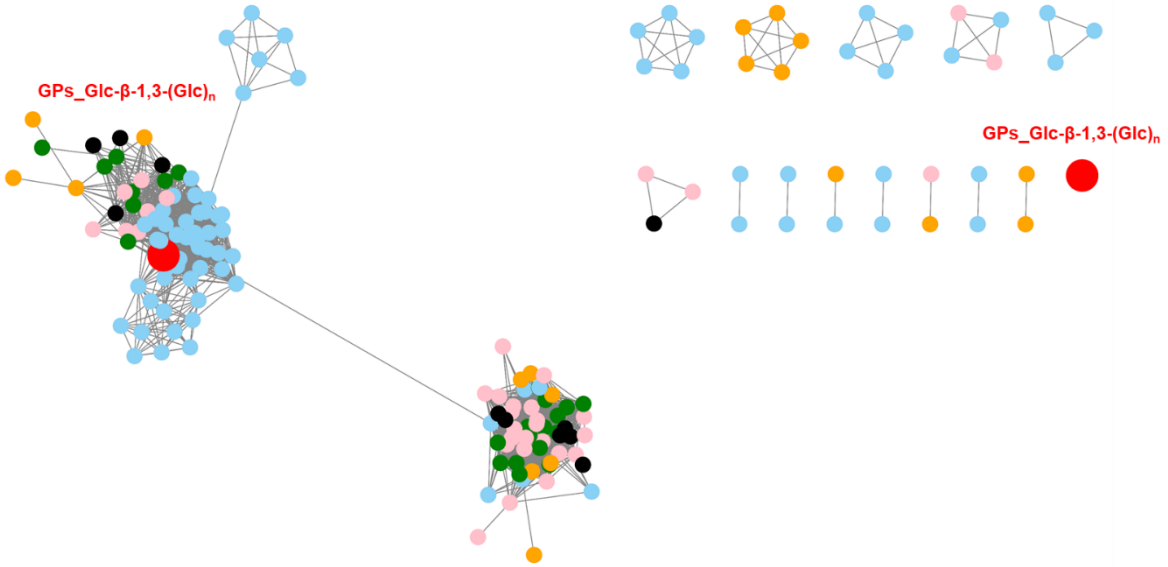


Figure 8: Protein sequence similarity network (SSN) of the 470 GH161 sequences from the CAZy database and from gut metagenomes, with an E-value threshold of 10^{-290}. Each of the 462 nodes contains sequences with more than 100% identity. Nodes are connected by an edge when the pairwise sequence identity is over 48%. In this figure, the 294 sequences appearing in 294 singletons have been omitted. In red, biochemically characterized GPs. Nodes are colored according to the origin of the sequences: sky blue, CAZy database; pink, human gut metagenome; green, mouse gut metagenome; black, pig gut metagenome; orange, bovine rumen metagenome.

Table 13: Distribution of the predicted GH161 glycoside-phosphorylases and hydrolases in the SSN meta-nodes, obtained with an E-value threshold of 10^{-280}. The number of sequences includes the full-length and truncated ones. Only the sequences belonging to meta-nodes with at least 20 sequences are mentioned.

GH161	**C1 (GPs-Glc-β-1,3-(Glc)$_n$)**
Prediction of mechanism	**GPs**
CAZy	69
Human	31
Mouse	23
Pig	10
Bovine	16
Total	149

Table 14: Characteristics of the SSNs. Characteristics of the GH130 sequences was previously analyzed (A. Li et al. 2020).

CAZy family	E-value threshold	Sequence identity threshold for connected nodes	Total number of sequences	Total number of full-length sequences	Number of full-length sequences in meta-nodes and proportion in the full-length sequence dataset (%)	Number of sequences in singletons/ small meta-nodes and proportion in the entire dataset (%)	Number of full-length sequences in singletons	Number of truncated sequences in singletons and proportion in the entire dataset (%)
GH65	10^{-180}	45%	7,255	6,432	6,264 97.4%	778/258 14,28%	168	610 (78.41%)
GH94	10^{-210}	52%	10,247	4,918	4,678 95.1%	3,894/236 40,3%	240	3,654 (93.84%)
GH112	10^{-190}	48%	1,984	755	718 95.1 %	920/10 46,87%	37	883 (95.98%)
GH130	10^{-70}	63%	6,308	3,692	3,649 98,8%	694/104 12,65%	43	651 (93.80%)
GH149	10^{-310}	46%	526	186	173 93%	316/2 60,45%	13	303 (95.89%)
GH161	10^{-280}	47%	470	176	161 93 %	286/35 68,3%	15	275 (96.15%)

Discussion

Representation of GPs-containing CAZy families in mammalian gut metagenomes

In this study, we increased the sequence diversity in the GH65, GH94, GH112, GH149 and GH161 families, by adding metagenomic sequences to those already listed in the CAZy database, of which at least 91 % come from genomes of cultured bacteria. Indeed, only 62 metagenomic sequences are listed in the CAZy database for these five families. Except for the highly multi-functional family GH65, which is well covered in the CAZy database with more than 6,500 sequences, we extracted from mammalian metagenomes between 2.7 and 5.1 times more sequences than the number listed in the CAZy database. Moreover, the sequence redundancy is low. With a threshold of 90% sequence identity, redundancy between each metagenome and the CAZy database does not exceed 11%, with nearly no redundancy in the families GH149 and GH161, of which many sequences of the CAZy database are issued from aquatic bacteria, in particular from the Flavobacteriales, Marinilabiliales or Cytophagales orders. Between the metagenomes, at the same threshold, redundancy ranges from 18.7% to 43.5%, highlighting the relative similarities between mammalian gut microbiomes.

All the targeted families except GH149 are similarly represented in mammalian gut metagenomes, with between 10^{-3} and 10^{-2} % of the non-redundant sequences in each metagenomic dataset. The GH130 sequences are in the same range (A. Li et al. 2020). These values are similar to the mean values obtained for GH families (between 2 and 17.10^{-3} %, depending on the metagenomic catalog) (J. Li et al. 2020). This significant abundance in these metagenomes is in accordance with the availability of the diverse substrates targeted by their members in mammalian guts ecosystems. Nevertheless, these values are nearly 10 times lower than for the most abundant families in these metagenomes, such as GH43 and GH13, which play major roles in hemicellulose and starch degradation, respectively.
For the GH65 and GH94 families, the SSN analysis allowed us to split the sequence diversity in many isofunctional meta-nodes, each related to one of the known substrates of the GH65 and GH94 enzymes. The GH65 GPs are active on disaccharides of various origins. Maltose is issued from starch, the plant energy storage polymer. Trehalose is synthetized in green algae, primitive plants, mushrooms, yeasts, bacteria, insects and nematodes, as a stress protectant. Kojibiose, nigerose, 2-O-α-d-glucopyranosyl-

glycerol and 3-O-α-d-glucopyranosyl-L-rhamnose are found in bacterial lipopolysaccharides or capsular polysaccharides, even though kojibiose is also found in the fruiting bodies of mushrooms. The GH94 family is also highly polyspecific. Most characterized GH94 GPs target the disaccharides chitobiose, cellobiose, laminaribiose and cellobionic acid (a disaccharide found in mycotic cellulose degradation), and cellodextrins (cello-oligosaccharides of DP>2). All these substrates are issued from the degradation of terrestrial and marine plants and algae, as well as from mushrooms and exo-skeleton of arthropods and insects for chitobiose. Other GH94 target cyclic β-1,2-glucooligosacccharides found in bacteria, or linear β-1,2-linked glucose oligomers found on bacterial cell wall glycans or in mushrooms fruiting body. In gut of mammals, these GPs thus probably mainly target bacterial and, to a lesser extent, mushroom glycosides.

In GH112, the galacto-N-biose/lacto-N-biose phosphorylases are involved in the degradation of the lacto- and galacto-N-biose structures found in milk oligosaccharides and in the core structure of mucin oligosaccharides which contribute to the intestinal colonization by bacteria (Kitaoka, Tian, and Nishimoto 2005). Our SSN analysis showed that very few sequences issued from the bovine rumen metagenome cluster with those of galacto-N-biose/lacto-N-biose phosphorylases. This is in accordance with the fact that no glycoprotein-rich mucus is secreted in the rumen, as opposed to the lower gastrointestinal tract of mouse, pig and human (Hoorens et al. 2011). In fact, nearly all the ruminal GH112 sequences cluster together with those of D-galactosyl-β-1,4-L-rhamnose phosphorylases. These sequences from rumen bacteria would thus target pectin, since the Gal-β-1,4-Rha motif is found in rhamnogalacturonan I. It is not excluded that these enzymes would also act on some bacterial exopolysaccharides containing this motif too (Nakajima, Nishimoto, and Kitaoka 2009b).

Finally, the only known function of GH161 enzymes is the degradation of β-1,3-linked gluco-oligosaccharides of DP ≥ 2 (Kuhaudomlarp, Pergolizzi, et al. 2019). β-1,3-glucosides are mainly described in marine polysaccharides, such as laminarin, which is an energy storage polysaccharide for brown algae. Nevertheless, they are also found in plant's callose produced in response to wounding, and in curdlan, a bacterial exopolysaccharide (Kuhaudomlarp et al. 2018). Except laminarin, these natural substrates are likely to be present in the intestinal microbiota of mammals, more particularly in the rumen, which would be richest in callose plants because of the herbivorous diet. Interestingly, even though GH161 sequences are not present in Bacteroidetes PULs, we observed that they clusterize with GH94 ones in other genomes, indicating that GPs could be involved in the complete breakdown of long linear β-1,3-glucoside chains, with GH161 β-1,3-glucan phosphorylases depolymerizing the longest chains, and GH94 laminaribiose-phosphorylases finalizing the catabolism of these substrates. However, interestingly, there is almost no redundancy between the GH161 sequences listed in the CAZy database and the gut metagenomic sequences. In addition, quite a few of the latter cluster in the SSN with those of characterized laminarin-phosphorylases, most of the metagenomic sequences appearing in a sub-meta-nodes. This segregation indicates likely novel GH161 functions in this sub-meta-nodes, mainly containing metagenomic sequences.

Contrary to the other ones, the family GH149 is absent or poorly represented (maximum 26.10^{-6} % of the sequences) in mammalian gut metagenomes. The only characterized GPs of the family GH149 are active on β-1,3-glucosides of DP ≥ 1. The chain length specificity of β-1,3-glucan phosphorylases in GH149 is more relaxed than that of the other β-1,3-glucan phosphorylases from the GH161 family. In Bacteroidetes PULs, GH149 sequences mainly co-localize with GH16, GH17, GH3 and GH30 sequences, which all contain β-glucanases and β-glucosidases, highlighting the potential contribution of the GH149 in the degradation of complex β-glucan structures. The fact that the GH149 sequences listed in the CAZy database are mostly issued from marine bacteria indicates that for most of them, the likely substrate target is laminarin. This also explains the low representation of GH149 sequences in the

mammal gut metagenomes. Interestingly, as for the GH161 family, very few of the metagenomic sequences cluster with the β-1,3-glucan phosphorylase sequences and the rest of the sequences from the CAZy database. The GH149 metagenomic sequences indeed cluster in the two independent uncharacterized meta-nodes UC1 and UC2, indicating that new functions that are still not described in the GH149 family could be found in these meta-nodes, or at least different linkage of chain length specificities. In particular, the UC2 meta-nodes probably contains glycoside-hydrolases, since one of its sequences contains a signal peptide.

The interest of SSNs to highlight uncharacterized sequence groups and to predict GP activity

The efficiency of SSNs for rapid clustering of a large number of CAZy sequences has already been shown (Viborg et al. 2019; A. Li et al. 2020). SSNs were recently applied to the creation of subfamilies within the large multi-functional CAZy family GH16 (Viborg et al. 2019), and to the discovery of a new GP function in the GH130 family (A. Li et al. 2020). In the present study, we describe the use of SSNs to analyze the diversity of five other CAZy families and to highlight groups of sequences which do not contain experimentally characterized member. These clusters are thus particularly interesting to mine for new glycoside modifying functions.

Of course, the sequences present in small meta-nodes (less than 20 sequences) would also be interesting to find new functions, although these functions would be rare in the ecosystem. However, a rare function is not necessarily without ecological importance, in particular since we did not assess in this study the gene abundance, prevalence and expression in the targeted microbiomes. Here, we chose a threshold of 20 sequences to consider meta-nodes, since it is one of the criteria used to define subfamilies in the CAZy database (Viborg et al. 2019). The percentage of sequences discarded in small meta-nodes and singletons varies from 12.6% in the GH130 family to 68.3% in GH161 family (Table 14). These values are particularly high for small families, such as GH149 and GH161, which raises questions about the representativeness of meta-nodes of more than 20 sequences in terms of functional diversity.

Most of the metagenomic sequences appearing as singletons are truncated, explaining their low similarity with any other sequence. The percentage of truncated sequences in singletons is nearly constant (between 96.1% and 93.8%), except in family GH65, for which it is only 78.4%. Anyway, when we consider all the full-length sequences, between 93 and 97.4 % of them are in meta-nodes (Table 14). These values indicate that our SSN analysis considers most of the diversity contained in the targeted databases and this, in very similar proportions for all the families.

In the optimal SSN configurations allowing the clusterization of sequences in iso-functional meta-nodes, we set an optimal E-value threshold corresponding to a pairwise sequence identity value under which the nodes are disconnected. The corresponding sequence identity values for the five studied families are in the same range, between 46% and 52% identity. This value was higher in the family GH130, where connections within a meta-nodes required at least 63% identity (A. Li et al. 2020). This difference is not due to the size of the family or its level of multi-functionality, since for the highly multi-functional GH65 and GH130 families, which gather respectively 7,265 and 6,308 sequences (similar numbers), the sequence identity thresholds are 45 % and 63 %, respectively. We suggest that this difference may rather be due to the level of diversity of quaternary structures. Indeed, with the exception of GH130 ones, GPs mainly act as dimers, including those of family GH161, in which no crystallographic structure has yet been solved (Kuhaudomlarp, Pergolizzi, et al. 2019). In contrast, almost all the GH130 members form higher homooligomeric complexes in solution and in crystals (Kawahara et al. 2012; Nakae et al. 2013; Nihira, Suzuki, et al. 2013; S. Ladevèze et al. 2015; Ye et al. 2016). The conformational stability of

these oligomeric complexes involves a high number of interaction surfaces between the different monomers, which affect the active site topology and, consequently, the enzyme donor and acceptor specificity (S. Ladevèze et al. 2015). For the GH130 family, sequence divergence, even minor, may thus result in differences in quaternary structures, and in substrate specificities.

Regarding the uncharacterized meta-nodes highlighted in this study, they represent highly variable proportions of the sequence diversity in each family. The GH65 and GH94 families already contain about 25 characterized members each, with various specificities. For these families, only 8 and 1.7 %, respectively, of the sequences are found in uncharacterized meta-nodes, which include both genomic and metagneomic sequences. The functional diversity of these families is thus probably already nearly fully characterized, and the sequences present in the CAZy database cover all these functions. In the GH112 family, no uncharacterized meta-nodes was found, indicating that except for the putative hydrolase identified in a small meta-nodes (V1.FI14_GL0040405), the functions already discovered in this family are probably exhaustive and cover well its sequence diversity. The newly created GH149 and GH161 families represent different cases. Indeed, in the GH149 family, 17 % of the sequences are found in uncharacterized meta-nodes, which contain nearly only metagenomic sequences, showing that the genomic sequences listed in the CAZy database probably do not cover the functionalities of this family. In the GH161 family, 68% of the sequences are in an uncharacterized sub-meta-nodes, which mostly contains metagenomic sequences, and a few sequences from the CAZy database too. As explained above, the uncharacterized sequence groups of these families thus represent interesting sources of enzymes to characterize.

Compared to what we carried out for the GH130 family, the prediction of glycoside-phosphorylase or hydrolase mechanism for the uncharacterized sequences of the GH65, GH94, GH112, GH149 and GH161 is less reliable. For these five families, we indeed just checked the presence or absence of peptide signals to predict hydrolase or phosphorylase functions, respectively. This is due to the fact that in these families (except GH65), no hydrolase has been discovered yet. This prevented us from identifying the presence of a carboxylic acid residue, which could act as a putative catalytic base in inverting hydrolases. Also, no CBM was found in the sequences of these five families, which could be the whiteness of extracellular or membrane bound hydrolases targeting polysaccharides (A. Li et al. 2020). Nevertheless, the biochemical characterization of the sequences V1.FI14_GL0040405 and 2009040_GL0562976, which have a signal peptide, should allow us to prove that some hydrolases do exist in the GH112 and GH149 families.

The same strategy has yet to be applied to the last family containing inverting GPs, GT108, and also to the GH3, GH13_18, GT4, and GT35 families, which contain retaining phosphorylases, if we can have enough computational resources to handle the huge amount of sequences (especially for the GT4 family), and after checking the modularity of these sequences.

Reference

Abe, Koichi, Masahiro Nakajima, Tetsuro Yamashita, Hiroki Matsunaga, Shinji Kamisuki, Takanori Nihira, Yuta Takahashi, et al. 2017. "Biochemical and Structural Analyses of a Bacterial Endo-β-1,2-Glucanase Reveal a New Glycoside Hydrolase Family." *Journal of Biological Chemistry* 292 (18): 7487–7506. https://doi.org/10.1074/jbc.M116.762724.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Andersen, Joakim Mark, Rodolphe Barrangou, Maher Abou Hachem, Sampo J. Lahtinen, Yong-Jun Goh, Birte Svensson, and Todd R. Klaenhammer. 2012. "Transcriptional Analysis of Prebiotic Uptake and Catabolism by Lactobacillus Acidophilus NCFM." *PLoS ONE* 7 (9). https://doi.org/10.1371/journal.pone.0044409.

Andersson, Ulrika, Fredrik Levander, and Peter Rådström. 2001. "Trehalose-6-Phosphate Phosphorylase Is Part of a Novel Metabolic Pathway for Trehalose Utilization in Lactococcus Lactis." *Journal of Biological Chemistry* 276 (46): 42707–13. https://doi.org/10.1074/jbc.M108279200.

Cate, James H. Doudna, Yong-Su Jin, Jonathan M. Galazka, and Suk-Jin Ha. 2014. Enhanced cellodextrin metabolism. United States US20140057323A1, filed February 7, 2012, and issued February 27, 2014. https://patents.google.com/patent/US20140057323A1/en.

Chaen, Hiroto, Takuo Yamamoto, Tomoyuki Nishimoto, Tetsuya Nakada, Shigeharu Fukuda, Toshiyuki Sugimoto, Masashi Kurimoto, and Yoshio Tsujisaka. 1999. "Purification and Characterization of a Novel Phosphorylase, Kojibiose Phosphorylase, from Thermoanaerobium Brockii." *Journal of Applied Glycoscience* 46 (4): 423–29. https://doi.org/10.5458/jag.46.423.

Chao, Chen, Soetaert Wim, and Desmet Tom. 2011. "Characterization of β-Galactoside Phosphorylases with Diverging Acceptor Specificities." *Enzyme and Microbial Technology* 49 (1): 59–65. https://doi.org/10.1016/j.enzmictec.2011.03.010.

Cho, Eunae, Daham Jeong, Youngjin Choi, and Seunho Jung. 2016. "Properties and Current Applications of Bacterial Cyclic β-Glucans and Their Derivatives." *Journal of Inclusion Phenomena and Macrocyclic Chemistry* 85 (3): 175–85. https://doi.org/10.1007/s10847-016-0630-3.

Ciocchini, Andrés E., L. Soledad Guidolin, Adriana C. Casabuono, Alicia S. Couto, Nora Iñón de Iannino, and Rodolfo A. Ugalde. 2007. "A Glycosyltransferase with a Length-Controlling Activity as a Mechanism to Regulate the Size of Polysaccharides." *Proceedings of the National Academy of Sciences* 104 (42): 16492–97. https://doi.org/10.1073/pnas.0708025104.

Daghlas, Salah A., and Shamim S. Mohiuddin. 2020. "Biochemistry, Glycogen." In *StatPearls*. Treasure Island (FL): StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK539802/.

Dakhova, O. N., N. E. Kurepina, V. V. Zverlov, V. A. Svetlichnyi, and G. A. Velikodvorskaya. 1993. "Cloning and Expression in Escherichia Coli of Thermotoga Neapolitana Genes Coding for Enzymes of Carbohydrate Substrate Degradation." *Biochemical and Biophysical Research Communications* 194 (3): 1359–64. https://doi.org/10.1006/bbrc.1993.1974.

D'Enfert, Christophe, and Thierry Fontaine. 1997. "Molecular Characterization of the Aspergillus Nidulans TreA Gene Encoding an Acid Trehalase Required for Growth on Trehalose." *Molecular Microbiology* 24 (1): 203–16. https://doi.org/10.1046/j.1365-2958.1997.3131693.x.

Destruelle, Monika, Helmut Holzer, and Daniel J. Klionsky. 1995. "Isolation and Characterization of a Novel Yeast Gene, ATH1, That Is Required for Vacuolar Acid Trehalase Activity." *Yeast* 11 (11): 1015–25. https://doi.org/10.1002/yea.320111103.

Devendran, Saravanan, Ahmed M. Abdel-Hamid, Anton F. Evans, Michael Iakiviak, In Hyuk Kwon, Roderick I. Mackie, and Isaac Cann. 2016. "Multiple Cellobiohydrolases and Cellobiose Phosphorylases Cooperate in the Ruminal Bacterium Ruminococcus Albus 8 to Degrade Cellooligosaccharides." *Scientific Reports* 6 (October). https://doi.org/10.1038/srep35342.

Eck, Raimund, Clemens Bergmann, Karl Ziegelbauer, Wolfgang Schönfeld, and Waldemer Künkel. 1997. "A Neutral Trehalase Gene from Candida Albicans: Molecular Cloning, Characterization

and Disruption." *Microbiology,* 143 (12): 3747–56. https://doi.org/10.1099/00221287-143-12-3747.

Ehrmann, Matthias A., and Rudi F. Vogel. 1998. "Maltose Metabolism of Lactobacillus Sanfranciscensis: Cloning and Heterologous Expression of the Key Enzymes, Maltose Phosphorylase and Phosphoglucomutase." *FEMS Microbiology Letters* 169 (1): 81–86. https://doi.org/10.1111/j.1574-6968.1998.tb13302.x.

Gerlt, John A., Jason T. Bouvier, Daniel B. Davidson, Heidi J. Imker, Boris Sadkhin, David R. Slater, and Katie L. Whalen. 2015. "Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1854 (8): 1019–37. https://doi.org/10.1016/j.bbapap.2015.04.015.

Goedl, Christiane, Thornthan Sawangwan, Patricia Wildberger, and Bernd Nidetzky. 2010. "Sucrose Phosphorylase: A Powerful Transglucosylation Catalyst for Synthesis of α-D-Glucosides as Industrial Fine Chemicals." *Biocatalysis and Biotransformation* 28 (1): 10–21. https://doi.org/10.3109/10242420903411595.

Guidolin, L. Soledad, Andrés E. Ciocchini, Nora Iñón de Iannino, and Rodolfo A. Ugalde. 2009. "Functional Mapping of Brucella Abortus Cyclic β-1,2-Glucan Synthase: Identification of the Protein Domain Required for Cyclization." *Journal of Bacteriology* 191 (4): 1230–38. https://doi.org/10.1128/JB.01108-08.

Ha, Suk-Jin, Jonathan M. Galazka, Eun Joong Oh, Vesna Kordić, Heejin Kim, Yong-Su Jin, and Jamie H. D. Cate. 2013. "Energetic Benefits and Rapid Cellobiose Fermentation by Saccharomyces Cerevisiae Expressing Cellobiose Phosphorylase and Mutant Cellodextrin Transporters." *Metabolic Engineering* 15 (January): 134–43. https://doi.org/10.1016/j.ymben.2012.11.005.

Hamazaki, H., and K. Hotta. 1979. "Purification and Characterization of an Alpha-Glucosidase Specific for Hydroxylysine-Linked Disaccharide of Collagen." *The Journal of Biological Chemistry* 254 (19): 9682–87.

Hidaka, Masafumi, Yuji Honda, Motomitsu Kitaoka, Satoru Nirasawa, Kiyoshi Hayashi, Takayoshi Wakagi, Hirofumi Shoun, and Shinya Fushinobu. 2004. "Chitobiose Phosphorylase from Vibrio Proteolyticus, a Member of Glycosyl Transferase Family 36, Has a Clan GH-L-like (Alpha/Alpha)(6) Barrel Fold." *Structure (London, England: 1993)* 12 (6): 937–47. https://doi.org/10.1016/j.str.2004.03.027.

Hidaka, Masafumi, Motomitsu Kitaoka, Kiyoshi Hayashi, Takayoshi Wakagi, Hirofumi Shoun, and Shinya Fushinobu. 2006. "Structural Dissection of the Reaction Mechanism of Cellobiose Phosphorylase." *The Biochemical Journal* 398 (1): 37–43. https://doi.org/10.1042/BJ20060274.

Hidaka, Y., Y. Hatada, M. Akita, M. Yoshida, N. Nakamura, M. Takada, T. Nakakuki, S. Ito, and K. Horikoshi. 2005. "Maltose Phosphorylase from a Deep-Sea Paenibacillus Sp.: Enzymatic Properties and Nucleotide and Amino-Acid Sequences." *Enzyme and Microbial Technology* 37 (2): 185–94. https://doi.org/10.1016/j.enzmictec.2005.02.010.

Honda, Yuji, Motomitsu Kitaoka, and Kiyoshi Hayashi. 2004. "Reaction Mechanism of Chitobiose Phosphorylase from Vibrio Proteolyticus: Identification of Family 36 Glycosyltransferase in Vibrio." *The Biochemical Journal* 377 (Pt 1): 225–32. https://doi.org/10.1042/BJ20031171.

Hoorens, Prisca R., Manuela Rinaldi, Robert W. Li, Bruno Goddeeris, Edwin Claerebout, Jozef Vercruysse, and Peter Geldhof. 2011. "Genome Wide Analysis of the Bovine Mucin Genes and Their Gastrointestinal Transcription Profile." *BMC Genomics* 12 (March): 140. https://doi.org/10.1186/1471-2164-12-140.

Huang, Ying, Beifang Niu, Ying Gao, Limin Fu, and Weizhong Li. 2010. "CD-HIT Suite: A Web Server for Clustering and Comparing Biological Sequences." *Bioinformatics* 26 (5): 680–82. https://doi.org/10.1093/bioinformatics/btq003.

Hüwel, Stephan, Lutz Haalck, Nikola Conrath, and Friedrich Spener. 1997. "Maltose Phosphorylase from Lactobacillus Brevis: Purification, Characterization, and Application in a Biosensor for Ortho-Phosphate." *Enzyme and Microbial Technology* 21 (6): 413–20. https://doi.org/10.1016/S0141-0229(97)00014-8.

Iñón de Iannino, Nora, Gabriel Briones, Marcelo Tolmasky, and Rodolfo A. Ugalde. 1998. "Molecular Cloning and Characterization of Cgs, the Brucella Abortus Cyclic β(1-2) Glucan Synthetase

Gene: Genetic Complementation of Rhizobium Meliloti NdvB and Agrobacterium Tumefaciens ChvB Mutants." *Journal of Bacteriology* 180 (17): 4392–4400.

Inoue, Yasushi, Keiko Ishii, Tetsuji Tomita, Tsuneya Yatake, and Fumio Fukui. 2002. "Characterization of Trehalose Phosphorylase from *Bacillus Stearothermophilus* SK-1 and Nucleotide Sequence of the Corresponding Gene." *Bioscience, Biotechnology, and Biochemistry* 66 (9): 1835–43. https://doi.org/10.1271/bbb.66.1835.

Inoue, Yasushi, Nozomu Yasutake, Yoshie Oshima, Yoshie Yamamoto, Tetsuji Tomita, Shinsuke Miyoshi, and Tsuneya Yatake. 2002. "Cloning of the Maltose Phosphorylase Gene from Bacillus Sp. Strain RK-1 and Efficient Production of the Cloned Gene and the Trehalose Phosphorylase Gene from Bacillus Stearothermophilus SK-1 in Bacillus Subtilis." *Bioscience, Biotechnology, and Biochemistry* 66 (12): 2594–99. https://doi.org/10.1271/bbb.66.2594.

Kawaguchi, Takashi, Yasuo Ikeuchi, Noriko Tsutsumi, Akihiko Kan, Jun-Ichi Sumitani, and Motoo Arai. 1998. "Cloning, Nucleotide Sequence, and Expression of the Clostridium Thermocellum Cellodextrin Phosphorylase Gene and Its Application to Synthesis of Cellulase Inhibitors." *Journal of Fermentation and Bioengineering* 85 (2): 144–49. https://doi.org/10.1016/S0922-338X(97)86758-X.

Kawahara, Ryosuke, Wataru Saburi, Rei Odaka, Hidenori Taguchi, Shigeaki Ito, Haruhide Mori, and Hirokazu Matsui. 2012. "Metabolic Mechanism of Mannan in a Ruminal Bacterium, Ruminococcus Albus, Involving Two Mannoside Phosphorylases and Cellobiose 2-Epimerase: Discovery of a New Carbohydrate Phosphorylase, β-1,4-Mannooligosaccharide Phosphorylase." *Journal of Biological Chemistry* 287 (50): 42389–99. https://doi.org/10.1074/jbc.M112.390336.

Keyhani, Nemat O., Xi-Bing Li, and Saul Roseman. 2000. "Chitin Catabolism in the Marine Bacterium Vibrio Furnissii Identification and Molecular Cloning of a Chitoporin." *Journal of Biological Chemistry* 275 (42): 33068–76. https://doi.org/10.1074/jbc.M001041200.

Kim, Sun-Ki, Daehwan Chung, Michael E. Himmel, Yannick J. Bomble, and Janet Westpheling. 2019. "Heterologous Co-Expression of Two β-Glucanases and a Cellobiose Phosphorylase Resulted in a Significant Increase in the Cellulolytic Activity of the Caldicellulosiruptor Bescii Exoproteome." *Journal of Industrial Microbiology & Biotechnology* 46 (5): 687–95. https://doi.org/10.1007/s10295-019-02150-0.

Kim, Sun-Ki, Michael E. Himmel, Yannick J. Bomble, and Janet Westpheling. 2018. "Expression of a Cellobiose Phosphorylase from Thermotoga Maritima in Caldicellulosiruptor Bescii Improves the Phosphorolytic Pathway and Results in a Dramatic Increase in Cellulolytic Activity." *Applied and Environmental Microbiology* 84 (3). https://doi.org/10.1128/AEM.02348-17.

Kim, Yeon-Kye, Motomitsu Kitaoka, Manem Krishnareddy, Yutaka Mori, and Kiyoshi Hayashi. 2002. "Kinetic Studies of a Recombinant Cellobiose Phosphorylase (CBP) of the Clostridium Thermocellum YM4 Strain Expressed in Escherichia Coli." *The Journal of Biochemistry* 132 (2): 197–203. https://doi.org/10.1093/oxfordjournals.jbchem.a003210.

Kitaoka, Motomitsu, Jiesheng Tian, and Mamoru Nishimoto. 2005. "Novel Putative Galactose Operon Involving Lacto-N-Biose Phosphorylase in Bifidobacterium Longum." *Applied and Environmental Microbiology* 71 (6): 3158–62. https://doi.org/10.1128/AEM.71.6.3158-3162.2005.

Koshland, D. E. 1953. "Stereochemistry and the Mechanism of Enzymatic Reactions." *Biological Reviews* 28 (4): 416–36. https://doi.org/10.1111/j.1469-185X.1953.tb01386.x.

Kuhaudomlarp, Sakonwan, Nicola J. Patron, Bernard Henrissat, Martin Rejzek, Gerhard Saalbach, and Robert A. Field. 2018. "Identification of Euglena Gracilis β-1,3-Glucan Phosphorylase and Establishment of a New Glycoside Hydrolase (GH) Family GH149." *Journal of Biological Chemistry* 293 (8): 2865–76. https://doi.org/10.1074/jbc.RA117.000936.

Kuhaudomlarp, Sakonwan, Giulia Pergolizzi, Nicola J. Patron, Bernard Henrissat, and Robert A. Field. 2019. "Unraveling the Subtleties of β-(1→3)-Glucan Phosphorylase Specificity in the GH94, GH149, and GH161 Glycoside Hydrolase Families." *Journal of Biological Chemistry* 294 (16): 6483–93. https://doi.org/10.1074/jbc.RA119.007712.

Kuhaudomlarp, Sakonwan, Clare E. M. Stevenson, David M. Lawson, and Robert A. Field. 2019. "The Structure of a GH149 β-(1 → 3) Glucan Phosphorylase Reveals a New Surface Oligosaccharide Binding Site and Additional Domains That Are Absent in the Disaccharide-Specific GH94

Glucose-β-(1 → 3)-Glucose (Laminaribiose) Phosphorylase." *Proteins: Structure, Function, and Bioinformatics* 87 (10): 885–92. https://doi.org/10.1002/prot.25745.

Kuhaudomlarp, Sakonwan, Samuel Walpole, Clare E. M. Stevenson, Sergey A. Nepogodiev, David M. Lawson, Jesus Angulo, and Robert A. Field. 2019. "Unravelling the Specificity of Laminaribiose Phosphorylase from Paenibacillus Sp. YM-1 towards Donor Substrates Glucose/Mannose 1-Phosphate by Using X-Ray Crystallography and Saturation Transfer Difference NMR Spectroscopy." *Chembiochem: A European Journal of Chemical Biology* 20 (2): 181–92. https://doi.org/10.1002/cbic.201800260.

Ladevèze, S., G. Cioci, P. Roblin, L. Mourey, S. Tranier, and G. Potocki-Véronèse. 2015. "Structural Bases for N-Glycan Processing by Mannoside Phosphorylase." *Acta Crystallographica Section D: Biological Crystallography* 71 (6): 1335–46. https://doi.org/10.1107/S1399004715006604.

Ladevèze, Simon, Elisabeth Laville, Jordane Despres, Pascale Mosoni, and Gabrielle Potocki-Véronèse. 2017. "Mannoside Recognition and Degradation by Bacteria." *Biological Reviews* 92 (4): 1969–90. https://doi.org/10.1111/brv.12316.

Li, Ao, Elisabeth Laville, Laurence Tarquis, Vincent Lombard, David Ropartz, Nicolas Terrapon, Bernard Henrissa, et al. 2020. "Analysis of the Diversity of the Glycoside Hydrolase Family 130 in Mammal Gut Microbiomes Reveals a Novel Mannoside-Phosphorylase Function." *Under Review*.

Li, Junhua, Huanzi Zhong, Yuliaxis Ramayo-Caldas, Nicolas Terrapon, Vincent Lombard, Gabrielle Potocki-Veronese, Jordi Estellé, et al. 2020. "A Catalog of Microbial Genes from the Bovine Rumen Unveils a Specialized and Diverse Biomass-Degrading Environment." *GigaScience Database*, May. https://doi.org/10.1101/272690.

Liu, Aimin, Hidehiro Tomita, Hebiao Li, Hiroshi Miyaki, Chika Aoyagi, Satoshi Kaneko, and Kiyoshi Hayashi. 1998. "Cloning, Sequencing and Expression of the Cellobiose Phosphorylase Gene of Cellvibrio Gilvus." *Journal of Fermentation and Bioengineering* 85 (5): 511–13. https://doi.org/10.1016/S0922-338X(98)80071-8.

Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. 2014. "The Carbohydrate-Active Enzymes Database (CAZy) in 2013." *Nucleic Acids Research* 42 (D1): D490–95. https://doi.org/10.1093/nar/gkt1178.

Meng, Dongdong, Xinlei Wei, Yi-Heng P. Job Zhang, Zhiguang Zhu, Chun You, and Yanhe Ma. 2018. "Stoichiometric Conversion of Cellulosic Biomass by in Vitro Synthetic Enzymatic Biosystems for Biomanufacturing." *ACS Catalysis* 8 (10): 9550–59. https://doi.org/10.1021/acscatal.8b02473.

Mukherjee, Keya, Tamari Narindoshvili, and Frank M. Raushel. 2018. "Discovery of a Kojibiose Phosphorylase in Escherichia Coli K-12." *Biochemistry* 57 (19): 2857–67. https://doi.org/10.1021/acs.biochem.8b00392.

Nakae, Setsu, Shigeaki Ito, Mariko Higa, Takeshi Senoura, Jun Wasaki, Atsushi Hijikata, Masafumi Shionyu, Susumu Ito, and Tsuyoshi Shirai. 2013. "Structure of Novel Enzyme in Mannan Biodegradation Process 4-O-β-d-Mannosyl-d-Glucose Phosphorylase MGP." *Journal of Molecular Biology* 425 (22): 4468–78. https://doi.org/10.1016/j.jmb.2013.08.002.

Nakajima, Masahiro, and Motomitsu Kitaoka. 2008. "Identification of Lacto-N-Biose I Phosphorylase from Vibrio Vulnificus CMCP6." *Applied and Environmental Microbiology* 74 (20): 6333–37. https://doi.org/10.1128/AEM.02846-07.

Nakajima, Masahiro, Takanori Nihira, Mamoru Nishimoto, and Motomitsu Kitaoka. 2008. "Identification of Galacto-N-Biose Phosphorylase from Clostridium Perfringens ATCC13124." *Applied Microbiology and Biotechnology* 78 (3): 465. https://doi.org/10.1007/s00253-007-1319-8.

Nakajima, Masahiro, Mamoru Nishimoto, and Motomitsu Kitaoka. 2009a. "Characterization of β-1,3-Galactosyl-N-Acetylhexosamine Phosphorylase from Propionibacterium Acnes." *Applied Microbiology and Biotechnology* 83 (1): 109. https://doi.org/10.1007/s00253-008-1838-y.

———. 2009b. "Characterization of Three β-Galactoside Phosphorylases from Clostridium Phytofermentans Discovery of d-Dalactosyl-β1→4-L-Rhamnose Phosphorylase." *Journal of Biological Chemistry* 284 (29): 19220–27. https://doi.org/10.1074/jbc.M109.007666.

———. 2010. “Characterization of D-Galactosyl-β1→4-l-Rhamnose Phosphorylase from Opitutus Terrae.” *Enzyme and Microbial Technology* 46 (3): 315–19. https://doi.org/10.1016/j.enzmictec.2009.12.007.

Nakajima, Masahiro, Nobukiyo Tanaka, Nayuta Furukawa, Takanori Nihira, Yuki Kodutsumi, Yuta Takahashi, Naohisa Sugimoto, et al. 2017. “Mechanistic Insight into the Substrate Specificity of 1,2-β-Oligoglucan Phosphorylase from Lachnoclostridium Phytofermentans.” *Scientific Reports* 7: 42671. https://doi.org/10.1038/srep42671.

Nakajima, Masahiro, Hiroyuki Toyoizumi, Koichi Abe, Hiroyuki Nakai, Hayao Taguchi, and Motomitsu Kitaoka. 2014. “1,2-β-Oligoglucan Phosphorylase from Listeria Innocua.” *PLOS ONE* 9 (3): e92353. https://doi.org/10.1371/journal.pone.0092353.

Nidetzky, Bernd, Richard Griessler, Alexandra Schwarz, and Barbara Splechtna. 2004. “Cellobiose Phosphorylase from Cellulomonas Uda: Gene Cloning and Expression in Escherichia Coli, and Application of the Recombinant Enzyme in a ‘Glycosynthase-Type’ Reaction.” *Journal of Molecular Catalysis B: Enzymatic*, Proceedings of the 6th International Symposium on Biocatalysis and Biotransformations - BIOTRANS’03, 29 (1): 241–48. https://doi.org/10.1016/j.molcatb.2003.11.014.

Nielsen, Henrik. 2017. “Predicting Secretory Proteins with SignalP.” In *Protein Function Prediction: Methods and Protocols*, edited by Daisuke Kihara, 59–73. Methods in Molecular Biology. https://doi.org/10.1007/978-1-4939-7015-5_6.

Nihira, Takanori, Hiroyuki Nakai, Kazuhiro Chiku, and Motomitsu Kitaoka. 2012. “Discovery of Nigerose Phosphorylase from Clostridium Phytofermentans.” *Applied Microbiology and Biotechnology* 93 (4): 1513–22. https://doi.org/10.1007/s00253-011-3515-9.

Nihira, Takanori, Hiroyuki Nakai, and Motomitsu Kitaoka. 2012. “3-O-α-d-Glucopyranosyl-l-Rhamnose Phosphorylase from Clostridium Phytofermentans.” *Carbohydrate Research* 350 (March): 94–97. https://doi.org/10.1016/j.carres.2011.12.019.

Nihira, Takanori, Mamoru Nishimoto, Hiroyuki Nakai, Ken’ichi Ohtsubo, and Motomitsu Kitaoka. 2014. “Characterization of Two α-1,3-Glucoside Phosphorylases from Clostridium Phytofermentans.” *Journal of Applied Glycoscience* advpub. https://doi.org/10.5458/jag.jag.JAG-2013_013.

Nihira, Takanori, Yuka Saito, Motomitsu Kitaoka, Mamoru Nishimoto, Ken’ichi Otsubo, and Hiroyuki Nakai. 2012. “Characterization of a Laminaribiose Phosphorylase from Acholeplasma Laidlawii PG-8A and Production of 1,3-β-d-Glucosyl Disaccharides.” *Carbohydrate Research* 361 (November): 49–54. https://doi.org/10.1016/j.carres.2012.08.006.

Nihira, Takanori, Yuka Saito, Motomitsu Kitaoka, Ken’ichi Otsubo, and Hiroyuki Nakai. 2012. “Identification of Bacillus Selenitireducens MLS10 Maltose Phosphorylase Possessing Synthetic Ability for Branched α-d-Glucosyl Trisaccharides.” *Carbohydrate Research* 360 (October): 25–30. https://doi.org/10.1016/j.carres.2012.07.014.

Nihira, Takanori, Yuka Saito, Mamoru Nishimoto, Motomitsu Kitaoka, Kiyohiko Igarashi, Ken’ichi Ohtsubo, and Hiroyuki Nakai. 2013. “Discovery of Cellobionic Acid Phosphorylase in Cellulolytic Bacteria and Fungi.” *FEBS Letters* 587 (21): 3556–61. https://doi.org/10.1016/j.febslet.2013.09.014.

Nihira, Takanori, Yuka Saito, Ken’ichi Ohtsubo, Hiroyuki Nakai, and Motomitsu Kitaoka. 2014. “2-O-α-D-Glucosylglycerol Phosphorylase from Bacillus Selenitireducens MLS10 Possessing Hydrolytic Activity on β-D-Glucose 1-Phosphate.” *PLOS ONE* 9 (1): e86548. https://doi.org/10.1371/journal.pone.0086548.

Nihira, Takanori, Erika Suzuki, Motomitsu Kitaoka, Mamoru Nishimoto, Ken’ichi Ohtsubo, and Hiroyuki Nakai. 2013. “Discovery of β-1,4-D-Mannosyl-N-Acetyl-D-Glucosamine Phosphorylase Involving in the Metabolism of N-Glycans.” *Journal of Biological Chemistry* 288 (38): 27366–74. https://doi.org/10.1074/jbc.M113.469080.

Nishimoto, Tomoyuki, Tetsuya Nakada, Hiroto Chaen, Shigeharu Fukuda, Toshiyuki Sugimoto, Masashi Kurimoto, and Yoshio Tsujisaka. 1996. “Purification and Characterization of a Thermostable Trehalose Synthase from Thermus Aquaticus.” *Bioscience, Biotechnology, and Biochemistry* 60 (5): 835–39. https://doi.org/10.1271/bbb.60.835.

O'Neill, Ellis. 2013. "An Exploration of Phosphorylases for the Synthesis of Carbohydrate Polymers." Doctoral, University of East Anglia. https://ueaeprints.uea.ac.uk/id/eprint/47480/.

O'Neill, Ellis C., Giulia Pergolizzi, Clare E. M. Stevenson, David M. Lawson, Sergey A. Nepogodiev, and Robert A. Field. 2017. "Cellodextrin Phosphorylase from Ruminiclostridium Thermocellum: X-Ray Crystal Structure and Substrate Specificity Analysis." *Carbohydrate Research* 451 (November): 118–32. https://doi.org/10.1016/j.carres.2017.07.005.

Papaspyridi, Lefki-Maria, Anastasia Zerva, and Evangelos Topakas. 2018. "Biocatalytic Synthesis of Fungal β-Glucans." *Catalysts* 8 (7): 274. https://doi.org/10.3390/catal8070274.

Park, Jae Kweon, Nemat O. Keyhani, and Saul Roseman. 2000. "Chitin Catabolism In The Marine Bacterium Vibrio Furnissii Identification, Molecular Cloning, And Characterization Of An,N′ -Diacetylchitobiose Phosphorylase" *Journal of Biological Chemistry* 275 (42): 33077–83. https://doi.org/10.1074/jbc.M001042200.

Puchart, Vladimír. 2015. "Glycoside Phosphorylases: Structure, Catalytic Properties and Biotechnological Potential." *Biotechnology Advances* 33 (2): 261–76. https://doi.org/10.1016/j.biotechadv.2015.02.002.

Reichenbecher, Marisa, Friedrich Lottspeich, and Karin Bronnenmeier. 1997. "Purification and Properties of a Cellobiose Phosphorylase (CepA) and a Cellodextrin Phosphorylase (CepB) from the Cellulolytic Thermophile Clostridium Stercorarium." *European Journal of Biochemistry* 247 (1): 262–67. https://doi.org/10.1111/j.1432-1033.1997.00262.x.

Sánchez-Fresneda, Ruth, José P. Guirao-Abad, María Martinez-Esparza, Sergi Maicas, Eulogio Valentín, and Juan-Carlos Argüelles. 2015. "Homozygous Deletion of ATC1 and NTC1 Genes in Candida Parapsilosis Abolishes Trehalase Activity and Affects Cell Growth, Sugar Metabolism, Stress Resistance, Infectivity and Biofilm Formation." *Fungal Genetics and Biology* 85 (December): 45–57. https://doi.org/10.1016/j.fgb.2015.10.007.

Sawano, Tatsuya, Wataru Saburi, Ken Hamura, Hirokazu Matsui, and Haruhide Mori. 2013. "Characterization of Ruminococcus Albus Cellodextrin Phosphorylase and Identification of a Key Phenylalanine Residue for Acceptor Specificity and Affinity to the Phosphate Group." *The FEBS Journal* 280 (18): 4463–73. https://doi.org/10.1111/febs.12408.

Schiraldi, Chiara, Isabella Di Lernia, and Mario De Rosa. 2002. "Trehalose Production: Exploiting Novel Approaches." *Trends in Biotechnology* 20 (10): 420–25. https://doi.org/10.1016/S0167-7799(02)02041-3.

Sernee, M. Fleur, Julie E. Ralton, Tracy L. Nero, Lukasz F. Sobala, Joachim Kloehn, Marcel A. Vieira-Lara, Simon A. Cobbold, et al. 2019. "A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in Leishmania Parasites." *Cell Host & Microbe* 26 (3): 385-399.e9. https://doi.org/10.1016/j.chom.2019.08.009.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13 (11): 2498–2504. https://doi.org/10.1101/gr.1239303.

Svartström, Olov, Johannes Alneberg, Nicolas Terrapon, Vincent Lombard, Ino de Bruijn, Jonas Malmsten, Ann-Marie Dalin, et al. 2017. "Ninety-Nine de Novo Assembled Genomes from the Moose (Alces Alces) Rumen Microbiome Provide New Insights into Microbial Plant Biomass Degradation." *The ISME Journal* 11 (11): 2538–51. https://doi.org/10.1038/ismej.2017.108.

Touhara, Kouki K., Takanori Nihira, Motomitsu Kitaoka, Hiroyuki Nakai, and Shinya Fushinobu. 2014. "Structural Basis for Reversible Phosphorolysis and Hydrolysis Reactions of 2-O-α-Glucosylglycerol Phosphorylase." *Journal of Biological Chemistry* 289 (26): 18067–75. https://doi.org/10.1074/jbc.M114.573212.

Viborg, Alexander Holm, Nicolas Terrapon, Vincent Lombard, Gurvan Michel, Mirjam Czjzek, Bernard Henrissat, and Harry Brumer. 2019. "A Subfamily Roadmap for Functional Glycogenomics of the Evolutionarily Diverse Glycoside Hydrolase Family 16 (GH16)." *Journal of Biological Chemistry*, September, jbc.RA119.010619. https://doi.org/10.1074/jbc.RA119.010619.

Wilson, Charlotte M., Miguel Rodriguez, Courtney M. Johnson, Stanton L. Martin, Tzu Ming Chu, Russ D. Wolfinger, Loren J. Hauser, et al. 2013. "Global Transcriptome Analysis of Clostridium Thermocellum ATCC 27405 during Growth on Dilute Acid Pretreated Populus and Switchgrass." *Biotechnology for Biofuels* 6 (1): 179. https://doi.org/10.1186/1754-6834-6-179.

Xu, Chenggang, Ranran Huang, Lin Teng, Dongmei Wang, Christopher L. Hemme, Ilya Borovok, Qiang He, et al. 2013. “Structure and Regulation of the Cellulose Degradome in Clostridium Cellulolyticum.” *Biotechnology for Biofuels* 6 (1): 73. https://doi.org/10.1186/1754-6834-6-73.

Yamamoto, Takuo, Kazuhiko Maruta, Kazuhisa Mukai, Hiroshi Yamashita, Tomoyuki Nishimoto, Michio Kubota, Shigeharu Fukuda, Masashi Kurimoto, and Yoshio Tsujisaka. 2004. “Cloning and Sequencing of Kojibiose Phosphorylase Gene from Thermoanaerobacter Brockii ATCC35047.” *Journal of Bioscience and Bioengineering* 98 (2): 99–106. https://doi.org/10.1016/S1389-1723(04)70249-2.

Yamamoto, Takuo, Mio Nishio-Kosaka, Seisuke Izawa, Hajime Aga, Tomoyuki Nishimoto, Hiroto Chaen, and Shigeharu Fukuda. 2011. “Enzymatic Properties of Recombinant Kojibiose Phosphorylase from Caldicellulosiruptor Saccharolyticus ATCC43494.” *Bioscience, Biotechnology, and Biochemistry* 75 (6): 1208–10. https://doi.org/10.1271/bbb.110116.

Ye, Yuxin, Wataru Saburi, Rei Odaka, Koji Kato, Naofumi Sakurai, Keisuke Komoda, Mamoru Nishimoto, Motomitsu Kitaoka, Haruhide Mori, and Min Yao. 2016. “Structural Insights into the Difference in Substrate Recognition of Two Mannoside Phosphorylases from Two GH130 Subfamilies.” *FEBS Letters* 590 (6): 828–37. https://doi.org/10.1002/1873-3468.12105.

Yernool, Dinesh A., James K. McCarthy, Douglas E. Eveleigh, and Jin-Duck Bok. 2000. “Cloning and Characterization of the Glucooligosaccharide Catabolic Pathway β-Glucan Glucohydrolase and Cellobiose Phosphorylase in the Marine HyperthermophileThermotoga Neapolitana.” *Journal of Bacteriology* 182 (18): 5172–79. https://doi.org/10.1128/JB.182.18.5172-5179.2000.

Zhao, Hua, Zhong-Kang Wang, You-Ping Yin, Yan-Ling Li, Zhen-Lun Li, Guo-Xiong Peng, and Yu-Xian Xia. 2007. “Trehalose and Trehalose-Hydrolyzing Enzyme in the Haemolymph of Locusta Migratoria Infected with Metarhizium Anisopliae Strain CQMa102.” *Insect Science* 14 (4): 277–82. https://doi.org/10.1111/j.1744-7917.2007.00153.x.

Zilli, D. M. W., R. G. Lopes, S. L. Alves, L. M. Barros, L. C. Miletti, and B. U. Stambuk. 2015. “Secretion of the Acid Trehalase Encoded by the CgATH1 Gene Allows Trehalose Fermentation by Candida Glabrata.” *Microbiological Research* 179 (October): 12–19. https://doi.org/10.1016/j.micres.2015.06.008.

Conclusion and perspectives

This PhD project aimed at discovering new GPs, in order to identify new functions of glycoside breakdown in microbial ecosystems, and to design innovative enzymatic routes for the synthesis of high-added value oligosaccharides.

A sequence-based approach for GP discovery

To boost the discovery of new GPs, I developed an approach combining sequence-based mining of genomes and metagenomes, and activity-based screening of GP activity. We targeted the CAZy families already containing characterized GP members (eight families in 2016, eleven in 2019), but, potentially, with new activities.

In order to analyze the representation of each CAZy family containing GPs in microbiomes, and intra-family sequence diversity, I tested the efficiency of Sequence Similarity Networks (SSNs), that were never tested in the lab before my PhD work. The generic strategy I established to mine (meta)genomes for GPs is based on the following steps:

- annotation of the gut microbial gene catalogs of human, mouse, pig and cow (totalizing nearly 34 million genes). V. Lombard performed this task, in the framework of the collaboration between G. Potocki-Veronese and the CAZy team. CAZy mining was based on Blast analysis against the CAZy database, and detection of Hidden Markov models (HMMs) built for each CAZy family. These sequences, as for all CAZy families, had been annotated in the framework of a collaborative work directed by Diego Morgavi and involving, among many other contributors, the CAZy team and G. Potocki-Veronese, to establish and analyze a catalog of microbial genes from the bovine rumen (Li et al. 2020)

- construction of SSNs from all the identified metagenomic and genomic sequences (extracted from the CAZy database), family per family.

- optimization of the clustering parameters to obtain meta-nodes containing functionally homogeneous members, and identification of the meta-nodes that do not contain any characterized member

- for each meta-node, prediction of the hydrolytic or phosphorolytic mechanism, based on i) presence/absence of signal peptide, GPs requiring, *in cellulo*, intracellular phosphate to catalyze phosphorolysis; ii) conservation of the catalytic GP/GH machinery and of the GP's phosphate binding residues; iii) presence/absence of carbohydrate-binding modules (since most GPs act on short oligosaccharides and thus don't need CBMs, which help some CAZymes to bind polysaccharides)

To establish the proof of concept of this strategy, we first targeted the CAZy family GH130, which contains both GHs and GPs acting on mannosides, a very diverse class of glycosides involved in the interactions between eukaryotes and prokaryotes, in particular in the human gut. This family is thus, with the GH65 one, the most diversified CAZy family containing GPs, in terms of enzymatic mechanisms, substrate and linkage specificities. With the 4,714 metagenomic sequences and the 1,594 sequences already listed in the CAZy database (November 2018), of which 98% are issued from cultured bacteria or archaea, our GH130 dataset totalizes 6,308 sequences. Using SSNs, we divided the diversity of sequences into 15 mostly isofunctional meta-nodes; of these, nine contained no experimentally characterized member. By examining the multiple sequence alignments in each meta-node, we predicted the determinants of the phosphorolytic mechanism and linkage specificity. We thus hypothesized that eight uncharacterized meta-nodes would be phosphorylases. These sequences are indeed characterized by the absence of signal peptides and of the catalytic catalytic base. Those sequences with the conserved

E/K, E/R and Y/R pairs of residues involved in substrate binding, would target β-1,2, β-1,3 and β-1,4-linked mannosyl residues, respectively. These predictions were tested by characterizing members from the uncharacterized meta-nodes, chosen among the most prevalent and abundant GH130 sequences of the human gut microbiome. These target sequences were synthetized, cloned and expressed in *E. coli*. The GP function of three of the most original GH130 targets was biochemically proven by using chromogenic assays, HPAEC-PAD, NMR, and MS-MS analysis of the phosphorolysis and reverse-phosphorolysis products. We discovered the first known β-1,4-mannosyl-glucuronicacid phosphorylase (the U1 enzyme), which targets a motif of the *Shigella* lipopolysaccharide O-antigen. Furthermore, the in-depth analysis of the prevalence and abundance of all known GH130 sequences in the human gut microbiome, which had been previously initiated with 65 genomic sequences from human gut bacteria (Ladevèze et al. 2013), allowed us to identify a biomarker of inflammatory bowel diseases (IBD). This sequence encodes the very original dual mannoside-phosphorylase/transmannosylase IBD_P1, which targets the *Man*-β1,2-Man-α1,2-Man trisaccharidic motif found in *Candida albicans* and *Salmonella sp.*, which are IBD-associated microorganisms.

Overall, with this work we demonstrated the power of this novel strategy for the discovery of original GPs, novel CAZyme functions and pathology biomarkers from huge sequence datasets, and for the exploitation of the catalogs of microbial genes from various ecosystems. We also showed the interest of this approach to analyze the diversity of predicted GPs in microbiomes. For example, we found that GH130 GPs targeting plant mannans and mammal N-glycans are the most abundant GH130 enzymes in gut microbiomes and, most generally, in the bacterial world, while highlighting that some of the enzymes from this family are involved in interactions between gut bacteria.

Then, I applied the strategy validated with the GH130 family to the GH65, GH94, GH112, GH149, GH161 families, which contain inverting GPs. These families are well represented in the mammalian gut microbiomes, except the GH149 one, which is nearly absent from these ecosystems, while being highly abundant in marine bacteria.

I also showed that the sequence and functional diversity of the GH65 and the GH94 families are already well covered by the presently available genomic and biochemical data. This is also the case of the GH112 family, even though we identified a putative hydrolase in a small meta-node. It indicates that the GH112 family probably contains rare glycoside-hydrolases, as previously shown by the S. Withers's group. for the GH3 one (Macdonald, Blaukopf, and Withers 2015), This hypothesis will be confirmed soon with the biochemical characterization of the protein identified from this dataset, which totalizes 1,984 GH112 sequences. The case of the GH149 and GH161 families is different. For them, I identified several uncharacterized meta-nodes (or sub-meta-nodes for the GH161 family) mainly containing gut metagenomic sequences. Novel activities (likely GPs for GH161, and GHs and GPs for GH149) could thus be found in these meta-nodes, even though the diversity of sequences in these families is low, with around 500 sequences each, including the metagenomic ones. I also analyzed the GH13_18 family, which contains retaining GPs, and identified a meta-node with a likely novel function. Unfortunately, this function (6'-P-Sucrose-phosphorylase), was identified from two genomic sequences (Franceus et al. 2019; Tauzin et al. 2019), before I could characterize the enzymes that I identified. In total, in addition to the GH130 sequences, I have identified 11 targets from the families analyzed during this PhD work. These sequences are particularly abundant and/or prevalent in the human gut metagenome. They represent eleven uncharacterized meta-nodes, and thus, potentially, eleven novel functions, or at least functions which are not described for these families. They will be biochemically characterized soon in the DiscOmics group. Furthermore, this strategy has not? yet to be applied to the last families known to contain GPs, the GT108, and the GH3, GT4, and GT35 families.

Advantages and limits of SSNs

From my point of view, the main interest of SSNs is, of course, to highlight groups of uncharacterized sequences, but also to split the diversity of each family into several smaller sequence groups, facilitating multiple sequence alignments and search for consensus motifs. This rendered possible the accurate prediction of mechanism and linkage specificities for the enzymes of the GH130 family. Multiple sequence alignments are indeed not optimal for large datasets of thousands sequences, since they generate shifts which could affect the analysis of residue or sequence motif conservation. I faced these hurdles during my PhD work, as I will explain further in this chapter.

Splitting large datasets in sub-groups of similar sequences with SSNs is also an interesting strategy to decrease the computational time and to improve structural alignments, based either on crystallographic protein structures or on tridimensional models. This is currently being assessed by Jeremy Esque and Isabelle André, from the Molecular Modelisation group of the TBI Enzymology pole. Indeed, SSNs alone cannot be used to predict enzyme mechanism, nor substrate or linkage specificities, which can vary due to differences in one single amino acid. We showed with the GH130 family, that primary sequence analysis can be sufficient to predict the GP or GH mechanism, and osidic linkage position, at least for families in which a sufficient number of enzymes with different mechanisms and linkage specificities have been biochemically characterized. However, this approach was insufficient to predict substrate specificities, which were identified, in our studies, by activity-based screening. In addition, the mannoside-phosphorylase/transmannosylase IBD_P1 belongs to a meta-node containing the previously characterized strict GP acting on β-1,2-mannosides (the Lin0857 enzyme from *Listeria innocua* Clip11262, (Tsuda et al. 2015). The very original catalytic properties of IBD_P1 were unpredictable by using SSNs and multiple sequence alignment analysis, since they are probably due to discrete differences of the active site topologies between enzymes belonging to this meta-node. This will have to be further explored by determining the IBD_P1 crystallographic structure, which will be compared to that of Lin0857 (PDB accession number 5B0P).

SSNs also have other limits. First, they cannot be directly used to analyze CAZy families containing many multimodular sequences. In this case, the multimodularity must be analyzed and the sequences truncated to analyze with SSNs only the CAZy module of interest. This is what the CAZy team did with the GH16 family (Viborg et al. 2019). Nevertheless, we showed in this PhD work that SSNs could be directly used for the families mainly containing GPs members, which are, in a large majority, monomodular. The other limit of SSNs is their instability, when the number and diversity of the analyzed sequences vary. I experimented that when I re-analyzed the diversity in families GH65, GH94, GH112 and GH130 at one year of interval, to update the dataset for the sequences listed in the CAZy database. That is why the creation of stable CAZy sub-families cannot be based only on SSNs-based clustering, but on phylogenetic analyses, which are much more stable and also have the advantage to provide one with evolutionary distances, contrary to SSNs. Nevertheless, again, SSNs can be used to facilitate the creation of the multiple sequence alignments required for the generation of phylogenetic trees, by selecting a small number of sequences from each meta-node. This is the strategy used to create GH16 sub-families (Viborg et al. 2019), and to analyze the diversity of the GH130 family in this PhD work.

Challenges in the discovery of GPs (or other CAZymes)

As showed in this study, and in many other ones targeting other functions, the discovery of new CAZymes requires robust and rapid tools to:

- predict enzyme mechanisms and substrate specificities (types of glycosyl units, osidic linkage anomery and position)
- biochemically confirm these functions, or to screen them experimentally if prediction failed or is impossible because of a lack of data on the sequence/structure/function relationships for some families.

As introduced above, the limitation of primary sequence analysis in predicting substrate specificities could be compensated by tridimensional structure comparison, SSN meta-node per SSN meta-node. This has not been tested in the present study, in which we circumvented these limitations by *in vitro* screening of the donor and acceptor specificities, by quantification of the concentration of inorganic phosphate released during reverse-phosphorolysis. This strategy was efficient both to prove the GP mechanism of the targeted enzymes, and to rapidly profile their preference towards glycosyl-phosphates and acceptors. Nevertheless, this approach cannot be used for GHs. For these latter, the discovery of functions requires either high-throughput screening assays with large libraries of glycosides, using, for example, carbohydrate arrays, or the development of efficient *in silico* tools to predict the target substrates.

In silico analysis of PUL composition, and thus, of the combinations of CAZymes involved in glycoside catabolism, can be a good way to predict the substrate targeted by Bacteroidetes PUL (Lapébie et al. 2019), and by each of its CAZyme component. Analysis of Bacteroidetes and Firmicutes PULs was previously used to predict the substrate specificity of GPs from the GH130 (Ladevèze et al. 2013; Cuskin et al. 2015) and GH13_18 families(Tauzin et al. 2019). In this thesis work, I tested this strategy to predict the specificity and physiologic function of the members of the GH130 uncharacterized meta-nodes.

We hypothesized that the GH130 enzyme functions could be predicted thanks to the presence, in a same locus, of genes encoding CAZymes belonging to different families containing characterized members, and to which we could assign specific functions. Metagenomic sequences from mammal guts represent the majority of our GH130 sequence set. These metagenomic sequences being short contigs of few kbp only, we had no access to full-length loci. In order to analyze the genomic context of the GH130 sequences, we thus only targeted the sequences listed in the CAZy database, issued from cultivated bacteria of which the genome has been sequenced. The strategy of genomic context analysis was the following:

- all the sequences from the CAZy database were extracted, meta-node by meta-node

- their taxonomical origin was retrieved. If the sequence came from a Bacteroidetes, the PUL database (Terrapon et al. 2018) was searched for the PUL containing the GH130 target. If not, we checked the 20 genes neighboring the GH130 target, and retrieved their CAZy annotation from the CAZy database. For each CAZy family, the frequency of co-occurrence in the PUL-like system with a GH130 was calculated, and used as edge attributes in a Cytoscape representation, as performed by (Ladevèze et al. 2013). For each meta-node, we deduced the glycosyl residues and linkages of the substrates targeted by the PUL-like system from the functions assigned to the most abundant CAZy families found in the loci. Then, the glycoside structures containing these motives were searched in three glycoside structure databases: the Carbohydrate Structure Database (CSDB, http://csdb.glycoscience.ru), which lists natural carbohydrate structures and the related literature, the Glycosciences database (http://www.glycosciences.de), which contains a variety of resources available in the field of glycoinformatics, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) GLYCAN website (https://www.kegg.jp/kegg/glycan), which provides KEGG pathway for glycan biosynthesis and metabolism. However, the results of this time-consuming analysis were disappointing. Firstly, the prediction of the non-Bacteroidetes PUL-like systems was highly speculative, and resulted in a large variability of the loci composition, however analyzed meta-node by meta-node. When considering only the Bacteroidetes PULs, which are the sole CAZy containing bacterial operons referenced to date, the results were not exhaustive. Indeed, as shown in Chapter 2, there are meta-nodes which have no known Bacteroidetes member to date, including the one which contains the novel U1 enzyme targeting Man-β-1,4-GlcA. In addition, it was difficult or even impossible to predict the substrates of GH130 enzymes when the PULs contain other GH130 sequences belonging to uncharacterized meta-nodes. We thus

considered that the results of this analysis were too speculative to be presented in this thesis and published.

In order to identify the substrate targeted by the enzymes characterized in this study, we thus used a panel of analytic tools to characterize the structure of the products obtained by either reverse-phosphorolysis, phosphorolysis or transglycosylation. The chromogenic assays allowed us to identify the best donors and acceptors, but not the osidic linkage type. HPAEC-PAD combined to NMR analyses were then used to identify the linkage type in the synthesized glycosides, but only those for which standards do exist. For those yet not described or which are not commercially available, we had to produce them using known GPs (in collaboration with Julien Durand and Laurence Tarquis, from the DiscOmics group), and to use highly resolutive MS-MS to determine the linkage position. Finally, for the most complex glycoside structures, such as those produced by the IBD_P1 enzyme, a very innovative method combining MS-MS with high-resolution IMS/IMS was used by David Ropartz and Simon Ollivier, our collaborators from the BIBS platform at INRAE Nantes, to determine intra-chain anomerism of oligosaccharides. This analysis was performed from only 1 µg of carbohydrates, directly from reactional mixtures. This demonstration paves the way for carbohydrate 'sequencing' using cyclic ion mobility mass spectrometry.

In this thesis, I showed that novel GP functions can be discovered from CAZy families known to contain GPs, by rationalizing the mining of large sequence spaces with the approach described above. However, we do not exclude the probability to find GPs from other CAZy families. At the beginning of the PhD work, I indeed tested another strategy for GP discovery, based on the analysis of the conservation of the catalytic machinery of putative enzymes belonging to CAZy families in which no GP has yet been discovered. The work was focused on families of which the characterized members use an inverting mechanism of hydrolysis. The inverting hydrolytic mechanism indeed requires two catalytic residues, while the inverting phosphorolytic mechanism only requires one catalytic acid/base, phosphate acting as catalytic base. The identification of putative GPs was thus based on the following steps:

- retrieving of all the CAZy sequences of the targeted families from the CAZy database

- analysis of multiple sequence alignments to identify the sequences in which i) the catalytic acid is conserved; ii) the catalytic base conserved in characterized GHs is missing

- prediction of the cellular location of the potential targets by signal peptide detection.

In this study, I did not use SSNs to split the targeted families into several meta-nodes. I thus faced many difficulties to analyze the conservation of catalytic residues in multiple sequence alignments, in which shifts had to be manually corrected, as previously done in other studies (Viborg et al. 2019). I thus stopped these analyses to concentrate my efforts onto the development of the SSN-based strategy described in this thesis for families GH65, GH94, GH112, GH130, GH149 and GH161. Nevertheless, I identified one putative GP sequence, belonging to the GH6 family. This target, which might be an inverting GP acting on β-linked-glucosides, has yet not been characterized to validate this strategy. Moreover, this approach targeting all CAZy families with inverting GH members could be optimized by using SSNs.

Discovery of GPs involved in microbial interactions, and their interest for the synthesis of antigenic oligosaccharides

In this study, we highlighted the specificity of some GH130 enzymes from human gut bacteria towards mannosides found in some pathogenic yeasts and bacteria (*Pseudomonas aeruginosa* for the Man-β-1,3-Glc motif targeted by the U7 enzyme, *Shigella boydii for the* Man-β-1,4-GlcA motif targeted by the U1 enzyme, *Candida albicans* and *Salmonella* species for the Man-β1,2-Man-α1,2-Man motif targeted by the IBD_P1 enzyme). These results revealed possible novel interactions between gut bacteria,

and that GH130 enzymes could target several antigenic motifs of the *Candida albicans* mannan. To experimentally validate these hypotheses, co-culture and transcriptomic studies targeting the commensals that produce such enzymes will have to be performed. Since the enzymes characterized in this work are issued from uncultured bacteria, homologs from cultured species will have to be searched in the SSN meta-nodes they belong to, by carefully examining the conservation of the amino acids involved in the active site topology in structural alignments.

Whatever are their physiological roles *in vivo*, the U1, U7 and IBD_P1 enzymes identified in this thesis work are very interesting biotechnological tools for the synthesis of antigenic mannosides. *In vitro* synthesis of β-mannosides is indeed challenging, due to the steric hindrance conferred by the configuration of mannosyl residues and the thermodynamic instability of the β-anomer (Stork and La Clair 1996; Hayes and Pietruszka 2017). To my knowledge, the U1 enzyme is the first known enzyme to synthesize *in vitro* the Man-β-1,4-GlcA disaccharide. Besides, the successive synthesis of β-1,2 / α-1,2 linkages in the Man-β1,2-Man-α1,2-Man motif was already showed to be feasible. However, the described synthesis pathway is complex, involving several yeast mannosyltransferases initiating and elongating α- and β-mannosylated chains (Mora-Montes et al. 2010; Fabre et al. 2014)(. The synthesis of the Man-β-1,3-Glc motif could also be performed by the zobellia_231 GH130 enzyme (Awad et al. 2017), even though the interest of this enzyme for the synthesis of this oligosaccharide was not mentioned in the paper of Awad et al.

The Man-β-1,4-GlcA and Man-β1,2-Man-α1,2-Man oligosaccharides could be used, for example, as dietary supplements to prevent or help treating shigellosis, salmonellosis and candidiasis in fragile persons, such as those suffering from IBD or immunosuppressed patients. Indeed, this could result in the increase of the abundance in the gut microbiota of commensal bacteria that are able to breakdown these oligosaccharides, and thus to attack the pathogenic species harboring these epitopes. Of course, for such applications, it will be necessary to study the impact of oral administration on the diversity and functioning of the microbiota, in order to avoid promoting bacteria able to metabolize these compounds, but which could be deleterious for the gut health. The three GPs discovered in this thesis work might also be used in chemo-enzymatic pathways for the synthesis of vaccine candidates. Both these strategies have already been tested in mice with β-1,2-linked mannosides (Nitz et al. 2002; Dalle et al. 2003; Sendid et al. 2004; Han et al. 2012; Johnson and Bundle 2013), targeting candidiasis.

Of course, many studies will have to be carried out to reach these objectives, in particular to assess and optimize the mannoside synthesis and purification processes. The U1, U7 and IBD_P1 enzymes do not synthesize mannosides of polymerization degree higher than three. Nevertheless, IBD P1 catalyzes a side-reaction, the synthesis of β-1,2-mannobiose and mannotriose by reverse-phosphorolysis. These compounds will be difficult to separate from the Man-β1,2-Man-α1,2-Man trisaccharide. Kinetic analyses with the various substrates targeted by IBD_P1 should allow us to define the optimal substrate concentrations and enzyme/substrate ratio to use, to minimize reverse-phosphorolysis, and Man-β1,2-Man-α1,2-Man phosphorolysis if this reaction occurs. In addition, the in-depth analysis of the IBD_P1 active site conformation could also allow us to identify the mutations to perform to favor the targeted reactions and thus, to optimize this biocatalyst.

Résumé de la thèse

Ce projet de thèse visait à découvrir de nouvelles glycoside-phosphorylases (GPs), afin d'identifier de nouvelles fonctions de dégradation des oligosaccharides dans les écosystèmes microbiens, et de concevoir des voies enzymatiques innovantes pour la synthèse d'oligosaccharides à haute valeur ajoutée.

Une nouvelle approche basée sur l'analyse de séquences pour la découverte de GPs

Pour accélérer le processus de découverte de nouvelles GPs, j'ai développé une approche combinant analyse *in silico* de grands jeux de données génomiques et métagénomiques, et criblage de l'activité GP. Nous avons ciblé les familles CAZy contenant déjà des GPs (huit familles en 2016, onze en 2019), mais potentiellement de nouvelles activités.

Afin d'analyser la représentation de chaque famille CAZy contenant des GPs dans les microbiomes, et la diversité des séquences au sein de chaque famille, j'ai testé l'efficacité des réseaux de similarité de séquences (SSNs), qui n'avaient jamais été testés au laboratoire avant ce travail de thèse. La stratégie générique que j'ai établie pour identifier les GPs dans les (méta)génomes est basée sur les étapes suivantes:

- annotation des catalogues de gènes microbiens (totalisant presque 34 millions de gènes) dans le système digestif de l'homme, de la souris, de la vache et du porc. Cette tâche a été réalisée par V. Lombard, dans le cadre de la collaboration entre G. Potocki-Veronese et l'équipe CAZy. L'annotation des CAZymes a été réalisée par analyse Blast contre la base de données CAZy, et détection de modèles de Markov cachés (HMMs) construits pour chaque famille CAZy. Ces séquences, comme celles de toutes les familles CAZy, avaient été annotées dans le cadre d'un travail collaboratif dirigé par Diego Morgavi et impliquant, parmi de nombreux autres contributeurs, l'équipe CAZy et G. Potocki-Veronese, pour établir et analyser un catalogue de gènes microbiens du rumen bovin(Li et al. 2020)

- construction de SSNs à partir de toutes les séquences métagénomiques et génomiques (extraites de la base de données CAZy), famille par famille.

- optimisation des paramètres de clusterisation pour obtenir des groupes de séquences contenant des membres caractérisés aux fonctions identiques, et identification des groupes qui ne contiennent aucun membre caractérisé

- pour chaque groupe de séquences, prédiction du mécanisme hydrolytique ou phosphorolytique, basée sur i) la présence / absence de peptide signal, les GPs nécessitant, *in cellulo*, le phosphate intracellulaire pour catalyser la phosphorolyse; ii) conservation de la machinerie catalytique des GPs ou GHs et des résidus impliqués dans la liaison au phosphate pour les GPs; iii) présence / absence de modules de liaison aux glucides (car la plupart des GPs agissent sur des oligosaccharides courts, et n'ont donc pas besoin de ces modules 'CBMs' impliqués dans la reconnaissance des polysaccharides par certaines CAZymes)

Pour établir la preuve de concept de cette stratégie, nous avons d'abord ciblé la famille CAZy GH130, qui contient à la fois des GHs et des GPs agissant sur les mannosides, une classe très diversifiée d' oligosaccharides impliqués dans les interactions entre eucaryotes et procaryotes, en particulier dans l'intestin humain. Avec GH65, cette famille est ainsi la famille CAZy contenant des GPs la plus diversifiée, en termes de mécanismes enzymatiques, de spécificité vis- à-vis du substrat et de la liaison osidique. Avec 4 714 séquences métagénomiques et les 1 594 séquences déjà répertoriées dans la base de données CAZy (Novembre 2018), dont 98% sont issues de bactéries ou d'archaebactéries cultivées, notre jeu de données GH130 totalise 6 308 séquences. À l'aide des SSNs, nous avons divisé la diversité

des séquences en 15 groupes principalement isofonctionnels; parmi ceux-ci, neuf ne contenaient aucun membre caractérisé expérimentalement. En examinant les alignements de séquences dans chaque groupe, nous avons prédit les déterminants du mécanisme phosphorolytique et la spécificité de liaison. Nous avons donc émis l'hypothèse que huit groupes non caractérisés contiennent des phosphorylases. Ces séquences sont en effet caractérisées par l'absence de peptides signaux et du nucléophile catalytique. Les séquences avec les paires de résidus conservées E / K, E / R et Y / R, impliquées dans la liaison du substrat, cibleraient respectivement les résidus mannosyl liés en β-1,2, β-1,3 et β-1,4. Ces prédictions ont été testées en caractérisant des membres des groupes non caractérisés, choisis parmi les séquences GH130 les plus répandues et les plus abondantes du microbiome intestinal humain, qui ont été synthétisées, clonées et exprimées dans *E. coli*. La fonction GP de trois des cibles GH130 les plus originales a été prouvée biochimiquement en utilisant des analyses chromogéniques, et par HPAEC-PAD, RMN et MS-MS des produits de phosphorolyse et de phosphorolyse inverse. Nous avons découvert la première β1,4-mannosyl-glucuronicacide phosphorylase (l'enzyme U1), qui cible un motif de l'antigène O du lipopolysaccharide des espèces *Shigella*. De plus, l'analyse approfondie de la prévalence et de l'abondance de toutes les séquences connues de GH130 dans le microbiome intestinal humain, qui avait été précédemment initiée avec 65 séquences génomiques de bactéries intestinales (Ladevèze et al. 2013), nous a permis d'identifier un biomarqueur des maladies inflammatoires de l'intestin (MICI). Cette séquence code pour la très originale enzyme IBD_P1, de type mannoside-phosphorylase / transmannosylase, qui cible le motif trisaccharidique Man-β1,2-Man-α1,2-Man trouvé dans *Candida albicans* et les espèces *Salmonella*, qui sont des microorganismes associés aux MICI.

Avec ce travail, nous avons démontré la puissance de cette nouvelle stratégie pour la découverte de GPs originales, de nouvelles fonctions enzymatiques et de biomarqueurs de pathologie, à partir de grands ensembles de séquences, permettant d'exploiter les métagenomes de divers écosystèmes. Nous avons également montré l'intérêt de cette approche pour analyser la diversité des GPs prédites dans les microbiomes. Par exemple, nous avons constaté que les GPs de la famille GH130 ciblant les mannanes d'origine végétale et les N-glycanes de mammifères sont les enzymes GH130 les plus abondantes dans les microbiomes intestinaux et, plus généralement, dans le monde bactérien, tout en mettant en évidence que certaines des enzymes de cette famille sont impliquées dans les interactions entre bactéries intestinales.

Par la suite, j'ai appliqué la stratégie validée avec la famille GH130 aux familles GH65, GH94, GH112, GH149, et GH161, qui contiennent des GPs 'inverting', qui inversent la configuration de l'anomérie de substrat. Ces familles sont bien représentées dans les microbiomes intestinaux des mammifères, à l'exception de la famille GH149, qui est presque absente de ces écosystèmes, tout en étant très abondante dans les bactéries marines.

J'ai également montré que la diversité de séquences et de fonctions des familles GH65 et GH94 sont déjà bien couvertes par les données génomiques et biochimiques actuellement disponibles. C'est également le cas de la famille GH112, même si nous avons identifié une hydrolase putative dans un petit groupe de séquences. Ceci indique que la famille GH112, comme précédemment montré par le groupe S. Withers pour la famille GH3, contient probablement de rares glycoside-hydrolases(Macdonald, Blaukopf, and Withers 2015). Ceci devra être confirmé grâce à la caractérisation biochimique de la protéine identifiée à partir de cet ensemble de données, qui totalise 1 984 séquences. Le cas des familles GH149 et GH161 est différent. Pour elles, j'ai identifié plusieurs groupes de séquences non caractérisés (ou sous-groupes pour la famille GH161), contenant principalement des séquences métagénomiques. De nouvelles activités (probablement GPs pour GH161, et GHs et GPs pour GH149) pourraient donc y être trouvées, même si la diversité des séquences dans ces familles est faible, avec environ 500 séquences chacune, incluant les séquences métagénomiques. J'ai également analysé la famille GH13_18, qui

contient des GPs de type ‘retaining’, ne changeant pas la configuration de l’anomérie de substrat. J’ai identifié un groupe de séquences originales, indiquant une fonction putative nouvelle. Malheureusement, cette fonction (6’-P-sucrose-phosphorylase) a été identifiée à partir de deux séquences génomiques (Franceus et al. 2019; Tauzin et al. 2019) , avant que j’ai pu caractériser les enzymes que j’avaient identifiées. Au total, en plus des séquences GH130, j'ai identifié 11 cibles dans les familles analysées au cours de cette thèse. Ces séquences sont particulièrement abondantes et / ou prévalentes dans le métagénome intestinal humain. Elles représentent onze groupes non caractérisés, et donc potentiellement onze nouvelles fonctions, ou du moins des fonctions qui ne sont pas décrites pour ces familles. Elles seront bientôt caractérisées biochimiquement dans le groupe DiscOmics. De plus, cette stratégie doit encore être appliquée aux dernières familles connues pour contenir des GPs, les GT108 et les familles GH3, GT4 et GT35.

Avantages et limites des SSNs

De mon point de vue, le principal intérêt des SSN est, bien sûr, de mettre en évidence des groupes de séquences non caractérisées, mais aussi de diviser la diversité de chaque famille en plusieurs groupes de séquences plus petits, facilitant les alignements de séquences et la recherche de motifs consensus. Cela a rendu possible la prédiction du mécanisme et de la spécificité de liaison pour les enzymes de la famille GH130. Les alignements de séquences ne sont en effet pas optimaux pour de grands jeux de données, couvrant des milliers de séquences, car ils génèrent des décalages qui peuvent affecter l'analyse de la conservation des résidus ou motifs. Je me suis heurté à cette difficulté pendant mon travail de doctorat, comme je l'expliquerai plus loin dans ce chapitre.

La division de grands ensembles de séquences en sous-groupes de séquences similaires grâce aux SSNs est probablement aussi une stratégie intéressante pour diminuer le temps de calcul et améliorer les alignements structuraux, basés soit sur des structures cristallographiques, soit sur des modèles tridimensionnels. Cette stratégie est en cours d'évaluation par Jeremy Esque et Isabelle André, du groupe Modélisation Moléculaire du pôle TBI Enzymologie. En effet, les SSNs seuls ne peuvent pas être utilisés pour prédire le mécanisme enzymatique, ni les spécificités de substrat ou de liaison, qui peuvent varier en raison de différences d’un seul acide aminé. Nous avons montré avec la famille GH130, que l'analyse de séquences primaires peut être suffisante pour prédire le mécanisme GP ou GH, et la position de la liaison osidique, au moins pour les familles dans lesquelles un nombre suffisant d'enzymes avec différents mécanismes et spécificités de liaison ont été caractérisées biochimiquement. Cependant, cette approche était insuffisante pour prédire les spécificités du substrat, qui ont été identifiées, dans nos études, par criblage de l'activité. En outre, la mannoside-phosphorylase / transmannosylase IBD_P1 appartient à un groupe de séquences contenant aussi une GP stricte caractérisée, agissant sur les β1,2-mannosides (l'enzyme Lin0857 de *Listeria innocua* Clip11262,(Tsuda et al. 2015). Les propriétés catalytiques très originales d'IBD_P1 n’ont pas été prédites en utilisant la stratégie basée à la fois sur les SSNs et les alignements de séquences, car elles sont probablement dues à de fines différences de topologie du site actif entre les enzymes appartenant à ce même groupe. Cela devra être approfondi en déterminant la structure cristallographique d’IBD_P1, qui sera comparée à celle de Lin0857 (numéro d'accession PDB 5B0P).

Les SSNs ont également d'autres limites. Premièrement, ils ne peuvent pas être utilisés directement pour analyser des familles CAZy contenant beaucoup de séquences multimodulaires. Dans ce cas, la multimodularité doit être analysée et les séquences tronquées pour n'analyser avec les SSNs que le

module CAZy d'intérêt. C'est ce qu'a réalisé l'équipe CAZy avec la famille GH16 (Viborg et al. 2019). Néanmoins, nous avons montré dans ce travail de thèse que les SSNs pouvaient être utilisés directement pour les familles contenant principalement des GPs, qui sont, dans une large majorité, monomodulaires. L'autre limite des SSNs est leur instabilité, lorsque le nombre et la diversité des séquences analysées varient. J'ai expérimenté cela lorsque j'ai ré-analysé la diversité des familles GH65, GH94, GH112 et GH130 à un an d'intervalle, pour mettre à jour le jeu de séquences répertoriées dans la base de données CAZy. C'est pourquoi la création de sous-familles CAZy stables ne peut pas être basée uniquement sur un clustering SSNs, mais sur des analyses phylogénétiques, qui sont beaucoup plus stables et ont également l'avantage de fournir des distances évolutives, contrairement aux SSNs. Néanmoins, encore une fois, les SSNs peuvent être utilisés pour faciliter la génération des alignements de séquences multiples nécessaires à la création d'arbres phylogénétiques, en sélectionnant un petit nombre de séquences de chaque groupe. C'est la stratégie qui a été utilisée par l'équipe CAZy pour créer des sous-familles GH16 (Viborg et al. 2019), et, dans ce travail de thèse, pour analyser la diversité de la famille GH130.

Défis pour la découverte de GPs (ou d'autres CAZymes)

Comme le montre cette étude, et de nombreuses autres ciblant d'autres fonctions, la découverte de nouvelles CAZymes nécessite des outils robustes et rapides pour:

- prédire les mécanismes enzymatiques et les spécificités du substrat (types d'unités glycosyle, anomérie et position de la liaison osidique)

- confirmer biochimiquement ces fonctions, ou les tester expérimentalement si la prédiction a échoué, ou est impossible en raison du manque de données sur les relations séquence / structure / fonction pour certaines familles.

Comme présenté ci-dessus, les limites de l'analyse de séquences primaires pour la prédiction des spécificités du substrat pourraient être compensées par la comparaison de structures tridimensionnelles, pour chaque groupe de séquences généré par les SSNs. Cela n'a pas été testé dans la présente étude, dans laquelle nous avons contourné ces limites en criblant *in vitro* les spécificités vis-à-vis du donneur et de l'accepteur, en quantifiant la concentration de phosphate inorganique libéré lors de la phosphorolyse inverse. Cette stratégie a été efficace à la fois pour prouver le mécanisme GP des enzymes ciblées, et pour profiler rapidement leur préférence envers les glycosyl-phosphates donneurs et les accepteurs. Néanmoins, cette approche ne peut pas être utilisée pour les GHs. Pour ces dernières, la découverte de fonctions nécessite soit du criblage à haut débit, avec de grandes banques d'oligosaccharides, en utilisant, par exemple, des puces à sucres, soit le développement d'outils *in silico* efficaces pour prédire les substrats cibles.

L'analyse *in silico* de la composition des *loci* bactériens impliqués dans la dégradation des polysaccharides (PULs) peut être un bon moyen de prédire le substrat ciblé par les PULs des Bacteroidetes (Lapébie et al. 2019), et par chacun de leurs composants enzymatiques. L'analyse des PULs de Bacteroidetes et de Firmicutes a été précédemment utilisée pour prédire la spécificité de substrat de GPs des familles GH130 (Ladevèze et al. 2013; Cuskin et al. 2015) et GH13_18 (Tauzin et al. 2019). Dans ce travail de thèse, j'ai testé cette stratégie pour prédire la spécificité et la fonction physiologique des membres des groupes de séquences non caractérisés de la famille GH130.

Ainsi, nous avons émis l'hypothèse que les fonctions des enzymes GH130 pouvaient être prédites grâce à la présence, dans un même locus, de gènes codant pour des CAZymes appartenant à différentes familles contenant des membres caractérisés, et auxquels nous pouvions attribuer des fonctions spécifiques. Les séquences métagénomiques du système digestif des mammifères représentent la majorité de notre jeu de données GH130. Ces séquences métagénomiques étant de courts contigs de quelques kbp seulement, nous n'avions pas accès aux loci entiers. Afin d'analyser le contexte génomique des séquences GH130, nous n'avons donc ciblé que les séquences répertoriées dans la base de données CAZy, issues de bactéries cultivées dont le génome a été séquencé. La stratégie d'analyse du contexte génomique était la suivante:

- toutes les séquences de la base de données CAZy ont été extraites, groupe SSN par groupe SSN

- leur origine taxonomique a été analysée. Si la séquence provenait d'un Bacteroidetes, le PUL contenant potentiellement la cible GH130 était recherché dans la base de données PULDB (Terrapon et al. 2018). Sinon, nous avons analysé les 20 gènes voisins de la cible GH130, et récupéré leur annotation CAZy dans la base de données CAZy. Pour chaque famille CAZy, la fréquence de co-occurrence dans les loci avec une GH130 a été calculée, et utilisée pour représenter avec Cytoscape ces 'réseaux' de CAZymes, comme réalisé par Ladeveze et al (Ladevèze et al. 2013). Pour chaque groupe de séquences généré avec les SSNs, nous en avons déduit les résidus glycosyle et les liaisons des substrats ciblés par les enzymes des familles CAZy les plus abondantes dans les loci. Ensuite, les structures glycosidiques contenant ces motifs ont été recherchées dans trois bases de données: CSDB (http://csdb.glycoscience.ru), qui répertorie les structures glycosidiques naturelles et la littérature connexe, la base de données Glycosciences (http: //www.glycosciences.de), qui contient une grande variété de ressources glycoinformatiques, et le site Web de l'Encyclopédie des gènes et génomes de Kyoto (KEGG) GLYCAN (https://www.kegg.jp/kegg/glycan) , qui fournit la voie KEGG pour la biosynthèse et le métabolisme des glycanes. Cependant, les résultats de cette vaste analyse ont été décevants. Premièrement, la prédiction des *loci* des non-Bacteroidetes était hautement spéculative, et a entraîné une grande variabilité dans la composition des *loci*, pour un même groupe de séquences SSNs. En considérant uniquement les PULs des Bacteroidetes, qui sont les seuls opérons bactériens codant pour des CAZymes référencés à ce jour, les résultats n'étaient pas exhaustifs. En effet, comme montré dans le chapitre 2, il y a des groupes de séquences GH130 qui n'ont pas de membre Bacteroidetes connu, dont celui qui contient la nouvelle enzyme U1 ciblant le motif Man-β-1,4-GlcA. De plus, il était difficile, voire impossible, de prédire les substrats des enzymes GH130 lorsque les PULs contiennent d'autres séquences GH130 appartenant à des groupes non caractérisés. Nous avons donc considéré que les résultats de cette analyse étaient trop spéculatifs pour être présentés dans cette thèse et publiés.

Afin d'identifier les substrats ciblés par les enzymes caractérisées dans cette étude, nous avons donc utilisé un panel d'outils analytiques pour caractériser la structure des produits obtenus soit par phosphorolyse inverse, soit par phosphorolyse ou transglycosylation. Les tests chromogéniques nous ont permis d'identifier les meilleurs donneurs et accepteurs, mais pas le type de liaison osidique. Les analyses HPAEC-PAD et RMN ont ensuite été utilisées pour identifier le type de liaison des oligosaccharides synthétisés, mais uniquement ceux pour lesquels des standards existent. Pour ceux qui ne sont pas encore décrits ou qui ne sont pas disponibles dans le commerce, nous avons dû les produire en utilisant des GPs connues (en collaboration avec Julien Durand et Laurence Tarquis, du groupe DiscOmics), et utiliser une technique MS-MS hautement résolutive pour déterminer la position de la liaison. Enfin, pour les structures glycosidiques les plus complexes, comme celles produites par l'enzyme IBD_P1, une méthode très innovante combinant MS-MS et IMS / IMS haute résolution a été utilisée par David Ropartz et Simon Ollivier, nos collaborateurs de la plateforme BIBS de l'INRAE Nantes, pour déterminer l'anomérie intra-chaîne des oligosaccharides. Cela a été réalisé à partir de

seulement 1 µg de glucides, directement à partir des mélanges réactionnels. Cette démonstration ouvre la voie au «séquençage» des glucides par spectrométrie de masse à mobilité ionique cyclique.

Dans cette thèse, j'ai montré que de nouvelles fonctions GPs peuvent être découvertes à partir de familles CAZy déjà connues pour contenir des GPs, en rationalisant l'analyse de larges espaces de séquences avec l'approche décrite ci-dessus. Cependant, nous n'excluons pas la probabilité de trouver des GPs dans d'autres familles CAZy. Au début de ces travaux de thèse, j'ai en effet testé une autre stratégie de découverte de GPs, basée sur l'analyse de la conservation des résidus impliqués dans le mécanisme catalytique des enzymes appartenant aux familles CAZy dans lesquelles aucune GPs n'a encore été découverte. Le travail a été concentré sur les familles dont les membres caractérisés utilisent un mécanisme 'inverting'. Dans ce cas, le mécanisme d'hydrolyse nécessite en effet deux résidus catalytiques, tandis que le mécanisme de phosphorolyse ne nécessite qu'un résidu catalytique acide / base, le phosphate jouant le rôle de nucléophile. L'identification de GPs putatives a donc été basée sur les étapes suivantes:

- récupération de toutes les séquences CAZy des familles ciblées dans la base de données CAZy

- analyse des alignements de séquences primaires pour identifier les séquences dans lesquelles i) l'acide catalytique est conservé; ii) le nucléophile conservé dans les GHs caractérisées est manquant

- prédiction de la localisation cellulaire des cibles potentielles par détection de peptide signal.

Dans cette étude, je n'ai pas utilisé les SSNs pour diviser la diversité des familles ciblées. J'ai donc rencontré des difficultés pour analyser la conservation des résidus catalytiques dans des alignements de séquences, dans lesquels les décalages devaient être corrigés manuellement, comme cela a été fait dans d'autres études (Viborg et al. 2019). J'ai donc arrêté ces analyses pour concentrer mes efforts sur le développement de la stratégie SSNs décrite dans cette thèse pour les familles GH65, GH94, GH112, GH130, GH149 et GH161. Néanmoins, j'ai identifié une séquence de GPs putative, appartenant à la famille GH6. Cette cible, qui pourrait être une GP 'inverting' ciblant des β-glucosides, n'a pas encore été caractérisée pour valider cette stratégie. De plus, cette approche ciblant toutes les familles CAZy avec des membres GH 'inverting' pourrait être optimisée en utilisant des SSNs.

Découverte de GPs impliqués dans les interactions microbiennes et intérêt pour la synthèse d'oligosaccharides antigéniques

Dans cette étude, nous avons mis en évidence la spécificité de certaines enzymes GH130 de bactéries intestinales humaines vis-à-vis de mannosides trouvés dans certaines levures et bactéries pathogènes (*Pseudomonas aeruginosa* pour le motif Man-β-1,3-Glc ciblé par l'enzyme U7, *Shigella boydii* pour la motif Man-β-1,4-GlcA ciblé par l'enzyme U1, *Candida albicans* et espèces du genre *Salmonella* pour le motif Man-β1,2-Man-α1,2-Man ciblé par l'enzyme IBD_P1). Ces résultats ont révélé de nouvelles interactions possibles entre bactéries intestinales, et que les enzymes GH130 pourraient cibler plusieurs motifs antigéniques du mannane de *Candida albicans*. Pour valider expérimentalement ces hypothèses, des études de co-cultures et de transcriptomique ciblant les bactéries commensales produisant de telles GPs devront être réalisées. Étant donné que les enzymes caractérisées dans ce travail sont issues de bactéries non cultivées, leurs homologues d'espèces cultivées devront être recherchés dans les groupes SSN auxquels elles appartiennent, en examinant attentivement la conservation des acides aminés impliqués dans la topologie du site actif dans des alignements structuraux.

Quels que soient leurs rôles physiologiques *in vivo*, les enzymes U1, U7 et IBD_P1 identifiées dans ce travail de thèse sont des outils biotechnologiques très intéressants pour la synthèse de mannosides antigéniques. La synthèse *in vitro* de β-mannosides est en effet difficile, en raison de l'encombrement stérique conféré par la configuration des résidus mannosyl et de l'instabilité thermodynamique de l'anomère β (Stork and La Clair 1996; Hayes and Pietruszka 2017). À ma connaissance, l'enzyme U1 est la première enzyme connue à synthétiser *in vitro* le disaccharide Man-β-1,4-GlcA. Par ailleurs, la synthèse successive des liaisons β-1,2 / α-1,2 dans le motif Man-β1,2-Man-α1,2-Man s'est déjà avérée réalisable. Cependant, la voie de synthèse décrite est complexe, impliquant plusieurs mannosyltransférases de levure initiant et allongeant des chaînes α et β mannosylées (Mora-Montes et al. 2010; Fabre et al. 2014). La synthèse du motif Man-β-1,3-Glc pourrait, quant à elle, également être réalisée par la GH130 zobellia_231 (Awad et al. 2017), même si l'intérêt de cette enzyme pour la synthèse de cet oligosaccharide n'a pas été mentionné dans l'article d'Awad et al.

Les oligosaccharides Man-β-1,4-GlcA et Man-β1,2-Man-α1,2-Man pourraient être utilisés, par exemple, comme compléments alimentaires pour prévenir ou aider à traiter les shigelloses, salmonelloses et candidoses chez les personnes fragiles, notamment celles souffrant de MICI ou les patients immunodéprimés. En effet, cela pourrait entraîner une augmentation de l'abondance dans le microbiote intestinal de bactéries commensales capables de cataboliser ces oligosaccharides, et donc d'attaquer les espèces pathogènes présentant ces épitopes. Bien entendu, pour de telles applications, il sera nécessaire d'étudier l'impact de l'administration orale sur la diversité et le fonctionnement du microbiote, afin d'éviter de favoriser les bactéries capables de métaboliser ces composés, mais qui pourraient nuire à la santé digestive. Les trois GPs découvertes dans ce travail de thèse pourraient également être utilisées dans les voies chimio-enzymatiques de synthèse de vaccins. Ces deux stratégies ont déjà été testées chez la souris avec des β-1,2-mannosides(Nitz et al. 2002; Dalle et al. 2003; Sendid et al. 2004; Han et al. 2012; Johnson and Bundle 2013), ciblant les candidoses.

De nombreuses études devront être menées pour atteindre ces objectifs, notamment pour évaluer et optimiser les processus de synthèse et de purification des mannosides. Les enzymes U1, U7 et IBD_P1 ne synthétisent pas de mannosides de degré de polymérisation supérieur à trois. Néanmoins, IBD P1 catalyse une réaction secondaire, la synthèse de β-1,2-mannobiose et mannotriose par phosphorolyse inverse. Ces composés seront difficiles à séparer du trisaccharide Man-β1,2-Man-α1,2-Man. Des analyses cinétiques avec les différents substrats ciblés par IBD_P1 devraient nous permettre de définir les concentrations de substrat optimales et les rapports enzyme / substrat à utiliser, pour minimiser la phosphorolyse inverse, et la phosphorolyse du Man-β1,2-Man-α1,2-Man si cette réaction est possible. Par ailleurs, l'analyse approfondie de la conformation du site actif de cette enzyme pourrait également nous permettre d'identifier les mutations à effectuer pour favoriser les réactions ciblées, et ainsi optimiser ce biocatalyseur.

Summary

Glycoside-phosphorylases are particular carbohydrate active enzymes (CAZymes), able to catalyze both glycoside degradation by using inorganic phosphate to breakdown osidic linkages, and synthesis by using sugar-phosphates as glycosyl donors. Very few glycoside-phosphorylases have been characterized to date compared to the other CAZymes, despite their involvement in important biological processes, in particular in the gut of mammals, and their potential for the synthesis of high-added value glycosides. In genomes and metagenomes, their sequences are indeed difficult to discriminate from those of glycoside-hydrolases and transferases, with which they share many structural and mechanistic similarities. Many glycoside-phosphorylases are thus probably still hidden in the uncultured fraction of microbial ecosystems.

In this thesis work, we developed a novel approach to boost the discovery of glycoside-phosphorylases, and to analyze their diversity in microbiomes. This generic approach combines sequence-based mining of (meta)genomes and activity-based screening of GP activity. To establish the proof of concept, we first targeted the CAZy family GH130, which contains both mannoside-phosphorylases and mannosidases targeting mannosides of various structures and origins. We analyzed 6,308 GH130 sequences, including 4,714 from the human, bovine, porcine and murine microbiomes. Using sequence similarity networks, we divided the diversity of sequences into 15 mostly isofunctional meta-nodes; of these, nine contained no experimentally characterized member. By examining the multiple sequence alignments in each meta-node, we predicted the determinants of the phosphorolytic mechanism and linkage specificity. These predictions were tested by characterizing four members of this family, of which the sequences are among the most prevalent and abundant GH130 sequences of the human gut microbiome. Their functions were proven by using chromogenic assays, high-performance anion exchange chromatography, nuclear magnetic resonance and cyclic ion mobility mass spectrometry. We discovered the first known β1,4-mannosyl-glucuronicacid phosphorylase, and a very original dual mannoside-phosphorylase/transmannosylase. Both of them target glycosidic motifs found in pathogenic yeasts and bacteria. This approach was then applied to the analysis of families GH65, GH94, GH112, GH149 and GH161. We showed that the sequence and functional diversity of the GH65, GH94 and GH112 families is already well covered by the presently available genomic and biochemical data. In contrast, the GH149 and GH161 sequences, in particular the metagenomic ones, are probably a source of functional novelty. In total, eleven targets were selected from uncharacterized meta-nodes, representing, potentially, eleven novel functions, or at least functions which are not described for these families. In this thesis work, we thus developed an efficient strategy for the discovery of glycoside-phosphorylases and assessment of their diversity in microbiomes. It also revealed possible interactions between gut bacteria, and allowed us to identify novel biotechnological tools for the synthesis of antigenic oligosaccharides.

Résumé

Les glycoside-phosphorylases sont des enzymes actives sur les hydrates de carbone (CAZymes), capables de catalyser à la fois la dégradation des glycosides en utilisant le phosphate inorganique pour rompre les liaisons osidiques, et leur synthèse en utilisant les sucre-phosphates comme donneurs de glycosyles. Très peu de glycoside-phosphorylases ont été caractérisées à ce jour par rapport aux autres CAZymes, malgré leur implication dans d'importants processus biologiques, en particulier dans le système digestif des mammifères, et leur potentiel pour la synthèse de glycosides à haute valeur ajoutée. Dans les génomes et les métagénomes, leurs séquences sont en effet difficiles à distinguer de celles des glycosides-hydrolases et des transférases, avec lesquelles elles partagent de nombreuses similitudes structurales et mécanistiques. De nombreuses glycoside-phosphorylases sont donc probablement encore cachées dans la fraction non cultivée des écosystèmes microbiens.

Dans ce travail de thèse, nous avons développé une nouvelle approche pour accélérer la découverte de glycoside-phosphorylases, et pour analyser leur diversité dans les microbiomes. Cette approche générique combine l'exploration de larges espaces de séquences (méta)génomiques, et le criblage d'activités glycoside-phosphorylases. Pour établir la preuve du concept, nous avons d'abord ciblé la famille CAZy GH130, qui contient à la fois des mannoside-phosphorylases et des mannosidases ciblant des mannosides de structures et d'origines diverses. Nous avons analysé 6 308 séquences GH130, dont 4 714 provenant des microbiomes humains, bovins, porcins et murins. En utilisant des réseaux de similarité de séquences, nous avons divisé la diversité des séquences en 15 groupes principalement isofonctionnels ; parmi ceux-ci, neuf ne contenaient aucun membre caractérisé expérimentalement. En analysant les alignements de séquences pour chaque groupe, nous avons pu prédire les déterminants du mécanisme phosphorolytique et de la spécificité de liaison osidique. Ces prédictions ont été testées en caractérisant quatre membres de cette famille, dont les séquences sont parmi les séquences GH130 les plus répandues et les plus abondantes dans le métagénome intestinal humain. Leurs fonctions précises ont été identifiées grâce à une combinaison d'analyses chromogéniques, chromatographiques, de résonance magnétique nucléaire, de spectrométrie de masse et mobilité ionique. Nous avons découvert la première β-1,4-mannosyl-acide glucuronique-phosphorylase connue, et une mannoside-phosphorylase/transmannosylase très originale. Toutes deux ciblent des motifs glycosidiques trouvés dans des levures et bactéries pathogènes. Cette approche a ensuite été appliquée à l'analyse des familles GH65, GH94, GH112, GH149 et GH161. Nous avons montré que la diversité de séquences et de fonctions des familles GH65, GH94 et GH112 est déjà bien couverte par les données génomiques et biochimiques actuellement disponibles. En revanche, les séquences GH149 et GH161, en particulier issues des métagénomes, sont probablement une source de nouveauté fonctionnelle. Au total, onze cibles ont été sélectionnées à partir de groupes de séquences non caractérisés, représentant, potentiellement, onze nouvelles fonctions, ou du moins des fonctions qui ne sont pas décrites pour ces familles. Ce travail de thèse a ainsi permis de développer une stratégie efficace pour la découverte de nouvelles glycoside-phosphorylases et l'évaluation de leur diversité dans les microbiomes. Il a également révélé des interactions possibles entre les bactéries intestinales, et permis d'identifier de nouveaux outils biotechnologiques pour la synthèse d'oligosaccharides antigéniques.