



**HAL**  
open science

# Discovery and characterization of glycoside-transporters from uncultured bacteria

Zhi Wang

► **To cite this version:**

Zhi Wang. Discovery and characterization of glycoside-transporters from uncultured bacteria. Biochemistry, Molecular Biology. INSA de Toulouse, 2020. English. NNT: 2020ISAT0037. tel-04213512

**HAL Id: tel-04213512**

**<https://theses.hal.science/tel-04213512>**

Submitted on 21 Sep 2023

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

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



# THÈSE

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

Délivré par l'Institut National des Sciences Appliquées de  
Toulouse

---

Présentée et soutenue par

**Zhi WANG**

Le 23 avril 2020

**Découverte et caractérisation des transporteurs de glycosides**

---

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 Catherine SCHOULER**, Rapporteur

**M. Roland MARMEISSE**, Rapporteur

**M. Matthieu ARLAT**, Examinateur

**M. Nicolas TERRAPON**, Examinateur

**Mme Alexandra TAUZIN**, 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 **Zhi Wang**

**Discovery and characterization of  
glycoside-transporters from uncultured bacteria**

*Directrices de thèse*

Mmes Gabrielle POTOCKI-VERONESE et Elisabeth LAVILLE



**NOM:** WANG

**PRÉNOM:** Zhi

**TITRE:** Découverte et caractérisation des transporteurs de glycosides.

**SPÉCIALITÉ:** Ingénieries Microbienne et Enzymatique

**ANNÉE:** 2020

**LIEU:** INSA, TOULOUSE

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

---

### **Résumé**

La métabolisation des oligosaccharides et polysaccharides est une fonction capitale pour les bactéries, en particulier celles des microbiotes digestifs des mammifères. Pour s'en nourrir, elles ont développé une machinerie protéique très diversifiée pour reconnaître, transporter et dégrader les sucres de structures complexes, notamment ceux composant les fibres alimentaires d'origine végétale. Malgré le rôle crucial des transporteurs, leur identification et leur caractérisation restent limitées du fait de plusieurs verrous techniques, liés au fait que ces protéines membranaires sont difficiles à étudier sous forme recombinante, et que la majorité d'entre elles appartiennent à des bactéries non-cultivées. Dans le cadre de cette thèse, nous avons criblé une banque de clones fosmidiques d'*Escherichia coli*, contenant des loci métagénomiques impliqués dans la métabolisation d'oligo- et polysaccharides, pour des fonctions de transports. Treize clones métagénomiques issus des microbiomes digestifs humain et bovin ont été identifiés comme pouvant utiliser des oligosaccharides comme seules sources de carbone, grâce à la co-expression de gènes codant pour des enzymes actives sur les hydrates de carbone et pour des transporteurs de différentes familles. Nous avons ensuite développé une stratégie générique pour analyser, *in vitro* et *in cellulo*, la spécificité de transporteurs recombinants chez *Escherichia coli*, combinant ingénierie de loci et du génome de l'hôte, criblage et quantification de l'efficacité de liaison, de translocation et de croissance, à partir d'oligosaccharides non marqués. Grâce à ces développements technologiques, nous avons caractérisé deux loci identifiés après l'étape de criblage comme étant impliqués dans la métabolisation des fructooligosaccharides par des Clostridiales non cultivées, dont l'un constitue un biomarqueur du régime alimentaire et de pathologies. Nous avons décrypté le mécanisme de deux systèmes phosphotransférases (PTS) de compositions différentes, spécifiques du kestose, et avons mis en évidence la promiscuité des constituants des PTS des bactéries Gram-positives et négatives. De plus, l'intégration de données biochimiques et (m)tagénomiques nous a permis de revisiter la classification et la distribution taxonomique des PTS bactériens, et de mieux comprendre l'interaction entre les prébiotiques et les bactéries intestinales. Enfin, la capacité de liaison d'une protéine de liaison aux glycanes, SusD, impliquée dans le transport de xylooligosaccharides par des bactéries du phylum Bacteroidetes, a été étudiée avec un large panel de techniques, et sa structure cristallographique résolue. L'ensemble de ces résultats démontre l'efficacité de notre stratégie pour identifier rapidement des transporteurs d'oligosaccharides de familles et d'origines taxonomiques variées, et pour caractériser les relations entre leur structure et leurs spécificités de liaison et de translocation.

---

**Mots-clés:** Transporteurs d'oligosaccharides, Métagénomique, Microbiote, Prébiotiques, Relations structure-fonction, Systèmes multi-géniques

---

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

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

**SURNAME:** WANG

**FIRST NAME:** Zhi

**TITLE:** Discovery and characterization of glycoside-transporters from uncultured bacteria.

**SPECIALITY:** Enzymatic and Microbial Engineering

**YEAR:** 2020

**PLACE:** INSA, TOULOUSE

**THEIS DIRECTOR:** Mrs Gabrielle POTOCKI-VERONESE and Elisabeth LAVILLE

---

**Summary:**

Glycoside metabolism is a crucial function for bacteria, in particular those residing in the gut of mammals. To feed on the dietary fibers of plant origin, they have developed very diverse proteic machineries to recognize, transport and degrade glycosides of complex structures. Despite the crucial role of transporters, their identification and characterization remain limited due to the fact that these membrane proteins are difficult to study in recombinant form, and that the majority of them belong to uncultured bacteria. In this thesis work, we screened a library of *Escherichia coli* fosmid clones, containing metagenomic loci involved in the metabolism of oligo- and polysaccharides, for transport functions. Thirteen metagenomic clones issued from the human and the bovine gut microbiomes were identified as being able to utilize oligosaccharides as sole carbon sources, thanks to the co-expression of genes encoding carbohydrate active enzymes and transporters from different families. Then, we developed a generic strategy to analyze, *in vitro* and *in cellulo*, the specificity of recombinant transporters in *Escherichia coli*, combining carbohydrate utilization locus and host genome engineering, screening and quantification of the binding, translocation, and growth rates, based on the use of unlabeled glycosides. Thanks to these technological developments, we characterized two loci identified after the screening step to be involved in the metabolism of fructooligosaccharides by uncultured Clostridiales, of which one constitutes a dietary and pathology biomarker in the human gut microbiome. We deciphered the mechanism of two phosphotransferase systems (PTS) of different structures, that we proved to be specific for kestose, and highlighted the promiscuity of the PTS components of Gram-positive and negative bacteria. Furthermore, integration of biochemical and (meta)genomic data allowed us to revisit the classification and taxonomical distribution of bacterial PTS and to better understand the interaction between prebiotics and human gut bacteria. Finally, the binding ability of a surface glycan binding protein, SusD, involved in xylooligosaccharide uptake by Bacteroidetes, was investigated with a large panel of techniques, and its crystallographic structure solved. These results demonstrate the efficiency of our strategy to rapidly identify glycoside transporters from various families and taxonomical origins, and to characterize the relationships between their structure and their binding and translocation specificities.

---

**Keywords:** Oligosaccharide transporters, Metagenomics, Microbiota, Prebiotics, Structure-function relationships, Multi-genic systems

---

**DOCTORAL SCHOOL:** SEVAB - Ecological, Veterinary, Agronomic Sciences and Bioengineering

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

## Acknowledgements

First and foremost, I would like to express my special appreciation and sincere gratitude to my supervisor Research Director Gabrielle Potocki-Veronese who fearlessly accepted me as PhD student for this project and was the main creator of the great ideas, techniques and whole background of this thesis. I would also like to acknowledge her immense knowledge, continuous scientific guidance and not least her inspiration. It would never have been possible for me to take this work to completion without their incredible support and encouragement.

My sincere thanks also go to Dr. Alexandra S. Tauzin, who helped me greatly in all the time of my research and writing of this thesis. I feel so sorry and also be grateful for the countless time she gave to me so that I could ask her 'hundreds questions every day', in which her patience and direction benefited me a lot.

I am also like to thank my co-supervisor Elisabeth Laville who helped me a lot in the experiment process. I would also like to thank all the collaborators: Fabien L  tisse, Pietro Tedesco, Gianluca Cioci, Guy Lippens, Pascale Lepercq and Myriam Mercade, for their help in the experiment and colorful discussion to smooth my research.

I would like to thanks, Dr. Catherine Schouler and Dr. Rolland Marmeisse, who honored me to be part of my jury and took the time to judge my work.

I would like to thank my committee members, Prof. Matthieu Arlat and Assistant Professor Nicolas Terrapon, for their serving as my committee members and for their fruitful discussion and meaningful suggestions for my project.

I would also like to thank all current and past members of the Enzyme molecular engineering and catalysis group for a nice working atmosphere.

Thanks are also due to the financial support from China Scholarship Council (CSC) in the frame of CSC-UT/INSA otherwise I would not have been able to develop my scientific discoveries.

A warm thanks to my family, who have supported me through the three years, otherwise, I would not have had the courage to embark on this journey in the first place.

天行健，君子以自强不息；地势坤，君子以厚德载物。

-- 《易经》

"The Heavens are in motion ceaselessly; The enlight'ned exert themselves constantly. While the Earth is supportive and natural, only the virtuous can bear the utmost."

-- The Book of Change

# Table of Contents

|   |     |
|---|-----|
| Acknowledgements .....  | 1   |
| Table of Contents .....   | 3   |
| List of figures and tables .....  | 5   |
| Abbreviations list .....  | 9   |
| General introduction.....   | 11  |
| Scientific production .....   | 15  |
| Introduction .....  | 17  |
| 1. Diversity of glycosides.....   | 19  |
| 1.1 Glycosides from plants and algae.....   | 20  |
| 1.2 Mammalian glycosides.....   | 20  |
| 1.3 Microbial glycosides .....  | 21  |
| 1.4 Arthropoda glycosides.....  | 22  |
| 2. Carbohydrate-active enzymes .....  | 22  |
| 3. Functional and structural diversity of glycoside transporters in bacteria.....     | 23  |
| 3.1. Techniques for functional characterization of glycoside-transporters .....       | 24  |
| 3.2. TonB-dependent transporters and associated surface-glycan binding proteins ..... | 28  |
| 3.3. ABC transporters of glycosides.....  | 36  |
| 3.4. PTS transporters of glycosides .....   | 47  |
| 3.5. MFS transporters of glycosides.....  | 57  |
| 3.6. Conclusion and thesis objectives.....  | 65  |
| Bibliography.....   | 67  |
| Result.....   | 103 |
| First article.....  | 105 |
| Second article .....  | 145 |
| Third article .....   | 195 |
| Fourth article .....  | 221 |
| Conclusion and perspectives .....   | 255 |
| R ésum é de la th èse .....   | 263 |



# List of figures and tables

## Introduction

|   |    |
|---|----|
| Figure 1. Schematic representation of the starch utilization system in <i>Bacteroides thetaiotaomicron</i> 5482. ....   | 29 |
| Figure 2. Transport of maltooligosaccharides by ABC in Gram-negative bacteria. ....   | 46 |
| Figure 3. Schematic illustration of the sugar transport and phosphorylation process by PTS. ....  | 47 |
| Figure 4. Carbohydrate transport of through the transmembrane PTS_EIIC domain.....  | 56 |
| Figure 5. Schematic representation of the three types of MFS transporters. ....   | 57 |
| Figure 6. Alternating access mechanism for an ion/carbohydrate symport. ....  | 64 |
| Table 1. Summary of the SusD and SusC/D pairs involved in polysaccharide utilization by Bacteroidetes, characterized either by biochemical, or by gene deletion or mutation approaches. ....  | 34 |
| Table 2. List of the bacterial ABC transporters involved in glycoside utilization, characterized either by binding or transport assays, gene deletion and/or transcriptional approaches. .... | 38 |
| Table 3. Phylum and order distribution of the characterized glycoside-targeting ABC transporters in bacteria. ....  | 44 |
| Table 4. List of the bacterial PTS transporters involved in glycoside utilization, characterized either by binding or transport assays, gene deletion and/or transcriptional approaches. .... | 50 |
| Table 5. Phylum an order distribution of the characterized glycoside-targeting PTS transporters in bacteria. ....   | 54 |
| Table 6. List of the bacterial MFS transporters involved in glycoside utilization, characterized either by binding or transport assays, gene deletion and/or transcriptional approaches. .... | 60 |
| Table 7. Phylum and order distribution of characterized glycoside-targeting MFS transporters in bacteria. ....  | 62 |

## First article

|   |     |
|---|-----|
| Figure 1. Growth of metagenomic clones on various oligosaccharides used as sole carbon sources: (A) FOS, (B) lactulose, (C) XOS, (D) cellotriose, (E) mixture of cellobiosyl-cellobiose and glucosyl-cellotriose, (F) mannotriose, (G) mixture of galactosyl-mannobiose and mannotriose, and (H) lactulose..... | 121 |
| Figure 2. Gene organization of hit clones that are able to grow on different glycosides.....  | 125 |

|  |     |
|--|-----|
| Figure 3. Growth of (A) clone F3 on aldouronic acids and (B) clones 5 and 41O1 on laminatriose. ....   | 126 |
| Table 1. Predicted substrate specificity of glycoside transporters of metagenomic clones, and corresponding transporters and CAZyme families identified in the clone sequences. ....   | 112 |
| Table 2. Number of hit clones obtained from different libraries and the hit yield in the growth screening.....   | 121 |
| Table 3. Summary of hit clones with the substrate used for the growth screening, their contig size and taxonomical assignation at the level of phylum, order or species, depending on the similarity of their sequence with referenced genomes. .... | 122 |
| Table S1. Known enzymatic activity of glycoside hydrolase (GH) families in the selected clones based on CAZy database and the substrates selected (based on activities in bold) for growth tests. ....   | 139 |

## Second article

|   |     |
|---|-----|
| Figure 1. Representation of FOS and sucrose utilization loci.....   | 153 |
| Figure 2. I9 FOS utilization locus abundance and prevalence in the human gut microbiome. ....   | 154 |
| Figure 3. Functional analysis of the I9 FOS utilization locus.....  | 156 |
| Figure 4. Oligosaccharide uptake and binding by I9min_PTS cells.....  | 160 |
| Figure 5. Saturation experiments on the <i>Dorea</i> PTS. ....  | 161 |
| Figure 6. Functionality of the <i>Dorea</i> PTS in <i>E. coli</i> . ....  | 163 |
| Figure 7. Functional, structural and taxonomical diversity of PTS. ....   | 166 |
| Figure S1. Relative expression level of genes conferring the FOS metabolization phenotype to the I9 clone, vs that of the endogenous <i>E. coli</i> housekeeping gene ( <i>ihfB</i> ). .... | 190 |
| Figure S2. HPAEC-PAD analysis of the reaction products resulting from (a) FOS and (b) Inulotriose hydrolysis by the I9min_GH32 cytoplasmic extracts, after 24 h incubation at 37 °C.....    | 191 |
| Figure S3. Analysis of the uptake of each component of the FOS mixture by the (a) clone I9, and its variants (b) I9min, (c) I9min_PTS/GH32, (d) I9min_ABC/GH32 and (e) I9min_GH32. ....     | 192 |
| Figure S4. Growth curves of knock-out mutants ( <i>ptsI</i> , <i>ptsH</i> and <i>fruB</i> ) of <i>E. coli</i> K-12 BW25113 (KEIO strain) on (a) glucose and (b) fructose. ....              | 193 |
| Table S1. List of primers used in this study.....   | 188 |

|   |     |
|---|-----|
| Table S2. Ability of intracellular extracts of I9min_GH32 cells to hydrolyze various FOS and $\beta$ -fructans..... | 189 |
|---|-----|

### Third article

|  |     |
|--|-----|
| Figure 1. Organization of the F1 FOS utilization locus, comparison with the I9 one, and synteny within other Clostridiales strains. ....   | 201 |
| Figure 2. (a) Classification of the F1 PTS based on the sequence similarity network (SSN) of PTS-EIIC domains from 126 proteins, retrieved from the Transporter Classification database, from the literature, and from the present study, constructed with an initial score of $10^{-5}$ . The nodes are connected by an edge if the pairwise sequence identity is $\geq 21.3\%$ . Unfilled nodes indicate sequences for which the functions were inferred by transcriptomics, while filled ones correspond to the functions validated biochemically or by gene deletion. (b) Multiple sequence alignment of the EIIB domains of the F1 and I9 PTSs, and their homologous belonging to the Glc, Fru and LacI families of the TCDB. The most conserved amino acids are indicated with colors. The conserved phosphorylation site (Cys residue) of EIIB is indicated by red asterisks above the alignment..... | 202 |
| Figure 3. Functional analysis of the F1 FOS utilization locus. ....  | 204 |
| Figure 4. Functionality of the F1 PTS transporter in <i>E. coli</i> . ....   | 207 |
| Table S1. Analysis of the consumption of the carbohydrates by HPAEC-PAD by clone F1 grown for 48 hours in M9 supplemented with FOS as sole carbon source.....  | 219 |
| Table S2. List of primers used for the construction of clone F1 variants. ....   | 220 |

### Fourth article

|  |     |
|--|-----|
| Figure 1. Affinity gel electrophoresis of the F5_SusD on a gel containing no polysaccharide (negative control) and on gels with 0.1% (w/v) beechwood xylan and wheat arabinoxylan, respectively..... | 226 |
| Figure 2. Characterization of the ligand binding abilities of F5-SusD using STD-NMR.....   | 227 |
| Figure 3. Complementation of XOS locus with recombinant F5_SusD. ....  | 229 |
| Figure 4. F5_SusD crystal structure. ....  | 232 |
| Figure 5. Comparison between F5_SusD putative binding area and the homologous structures (same as Figure 4). ....  | 233 |
| Table 1. Data collection and refinement statistics for the crystal structure of F5_SusD .....  | 231 |
| Figure S1. 1D proton spectrum at 298K of 500 $\mu$ l of P5_SusD protein $\sim 3.8$ mg/ml ( $=54 \mu$ M) in Tris buffer.....  | 248 |

|   |     |
|---|-----|
| Figure S2. (left) 1D proton spectrum at 298K of 500µl of xylan@500µM with P5-SusD at 0 (blue), 15 (red) or 27µM (magenta) in Tris buffer. (right) STD spectra of the different samples with saturation at 0ppm for the isolated xylan (blue), in presence of 15 (red) or 27 µM (magenta) of P5-SusD. The reference xylan spectrum is shown in black.....    | 249 |
| Figure S3. (left) 1D proton spectrum of 500µl of xylan@500µM with P5-SusD at 0 (blue), 15 (red) or 27µM (magenta) in Tris buffer Te= 298K (right) STD spectra of the different samples with saturation at 10.1ppm for the isolated xylan (blue), in presence of 15 (red) or 27 µM (magenta) of P5-SusD. The reference xylan spectrum is shown in black..... | 250 |
| Figure S4. Close-up view of the interaction between the loop 292-315 and a symmetry related molecule. ....  | 251 |
| Figure S5. Overall structure of the F5_SusD with the charged area colored according to a gradient from positive red to negative blue.....   | 252 |
| Figure S6. Study of the possible interaction between SusC and SusD. ....  | 253 |
| Table S1. List of primers for F5 used in this study. ....   | 247 |

## Abbreviations list

|              |   |
|--------------|---|
| AAs:         | Auxiliary Activities  |
| ABC:         | ATP-binding cassettes   |
| AGPs :       | Pectin and arabinogalactan proteins   |
| AGE:         | Affinity gel electrophoresis  |
| AXOS:        | Arabino-xylo-oligosaccharides   |
| CAZymes:     | Carbohydrate-active enzymes   |
| CBMs:        | Carbohydrate binding modules  |
| CEs:         | Carbohydrate Esterases  |
| DSF:         | Differential scanning fluorimetry   |
| EI:          | Enzyme I  |
| EII:         | Enzyme II   |
| FODMAPs      | Fermentable oligo-, di-, and monosaccharides and polyols                                      |
| FOS:         | Fructooligosaccharides  |
| GAGs:        | Glycosaminoglycans  |
| GalNAc:      | <i>N</i> -Acetylgalactosamine   |
| GlcNAc:      | <i>N</i> -Acetylglucosamine   |
| GHs:         | Glycoside Hydrolases  |
| GOLD:        | Genomes Online database   |
| GTs:         | GlycosylTransferases  |
| HG:          | Homogalacturonan  |
| HPAEC-PAD:   | High performance anion exchange chromatography with pulsed amperometric detection             |
| HMOs:        | Human milk oligosaccharides   |
| HPr:         | Histidine protein   |
| HTCS:        | Hybrid two-component systems  |
| IBD:         | Inflammatory bowel diseases   |
| IBS:         | Irritable bowel syndrome  |
| IC-ESI-HRMS: | Ion chromatography coupled with electrospray ion ionization high-resolution mass spectrometry |

|        |                                       |
|--------|---------------------------------------|
| LPMOs: | Lytic polysaccharide mono-oxygenases  |
| MFS:   | Major facilitator superfamily         |
| NBD:   | Nucleotide-binding domain             |
| PEP:   | Phosphoenolpyruvate                   |
| PLs:   | Polysaccharide Lyases                 |
| PTS:   | Phosphotransferase systems            |
| PUL:   | Polysaccharide utilization locus      |
| RGI:   | Rhamnogalacturonan I                  |
| RGII:  | Rhamnogalacturonan II                 |
| STD    | Saturation transfer difference        |
| SBP:   | Substrate-binding protein             |
| SCFAs: | Short-chain fatty acids               |
| SD:    | Scaffold domain                       |
| SGBPs: | Surface glycan-binding proteins       |
| SPR:   | Surface plasmon resonance             |
| SSN:   | Sequence similarity networks          |
| SUS:   | Starch utilization system             |
| TBDT:  | TonB-dependent transporters           |
| TCDB:  | Transporter Classification Database   |
| TCID : | Transporter Classification Identifier |
| TD:    | Transport domain                      |
| TMD:   | Transmembrane domain                  |
| TPR:   | Tetratricopeptide repeat              |
| XIC:   | Extracted Ion Chromatogram            |
| XOS:   | Xylooligosaccharides                  |

## General introduction

Glycosides are very abundant and structurally diverse molecules. They are found in all kinds of living organisms, in which they ensure important roles, such as energy storage, cell structuration, signaling and differentiation. They represent the largest repository of metabolically accessible carbon, and thus, an important energy source for most bacteria in many environments. In particular, in the gut of mammals, glycosides of vegetal, mammalian and bacterial origins are very abundant. Their utilization by bacteria implies multiple effects on the host, positively influencing the digestive comfort, but also the progression of various chronic gut diseases and potentially, bacterial infections. Besides, for biotechnological applications, the control of glycoside utilization is important to develop microbial cell factories and microbial consortia able to efficiently convert non-refined biomass to high-value products.

To feed on structurally complex glycosides, bacteria have developed specific polysaccharide utilization loci (PULs), which encode the transporters and carbohydrate-active enzymes (CAZymes) that are necessary to import these substrates into the cells and to degrade them into monosaccharides. Transport is thus a crucial step in the metabolism of glycosides, which is itself a key of microbiota functioning and equilibrium. However, the identification and characterization of transporters face several technological locks. Despite the major advances realized for some cultured bacteria in the last decade, the difficulties of genetic manipulation, often combined with functional redundancy, and the fact that transmembrane proteins cannot be studied as easily as soluble proteins, indeed make the functional characterization of transport systems challenging. This is even impossible in the case of uncultured microorganisms, which dominate in microbial ecosystems. In order to directly explore the taxonomical and functional diversity of uncultured microbes, the metagenomic approach has been developed in the late 1980s. According to the definition of Jo Handelsman, metagenomics (also called community genomics, environmental genomics, or population genomics) is “the genomic analysis of a population of microorganisms”.

In the last decade, a high-throughput activity-based functional metagenomic approach has been developed in the ‘DiscOmics’ group of the Molecular Catalysis and Engineering team at the Toulouse Biotechnology Institute. This approach, applied to the functional exploration of the human and the bovine gut microbiomes, allowed the identification of several hundreds *Escherichia coli* clones which include large metagenomic DNA fragments (30 to 40 kbp)

containing entire bacterial operon-like gene clusters encoding for the synergistic CAZymes targeting complex glycosides. In these metagenomic loci, the CAZyme-encoding genes are very often clustered with transporter encoding ones. Furthermore, it was shown that glycoside transporters from certain glycoside foraging bacteria can be produced in functional forms, from a PUL cloned in a fosmid and expressed in *E. coli*.

In this context, my thesis work was initiated in 2017, focusing on the identification and characterization of transporters involved in key glycoside metabolism pathways. We targeted the human intestinal and bovine rumen microbiomes, which are both rich in glycoside catabolic pathways. In this manuscript, I present the results of three years of work, dedicated to the understanding of certain mechanisms of glycoside transport in uncultured bacteria, and to technological developments to address this complex research field.

This manuscript includes five chapters:

The First chapter constitutes the bibliographic introduction, of which a part will be published as a review paper. This chapter focuses on the diversity of glycosides in Nature, on the CAZymes involved in their degradation, and presents the different glycoside transporter families in bacteria.

The following four chapters consist in the results obtained during this thesis, which are presented in the form of four articles:

Chapter Two is an article to be submitted, describing the identification of functional glycoside transporters from uncultured bacteria.

Chapter Three is an article published in the journal *mSphere* in 2020. It describes the mechanism of prebiotic fructooligosaccharide (FOS) harvesting by a non-beneficial human gut bacterium.

Chapter Four is an article in preparation. It focuses on the mechanistic study of a FOS transporter from an uncultured human gut bacterium, which presents an original structure lacking a functional domain.

Chapter Five is also an article in preparation, focusing on the functional and structural characterization of a surface glycan binding protein involved in xylooligosaccharides uptake by an uncultured Bacteroidetes.

The conclusions and perspectives at the end of this manuscript highlight the efficiency of our strategy to rapidly identify functional glycoside transporters from various families and taxonomical origins, and to characterize the relationships between their structure and their binding and translocation specificities.



# Scientific production

## Research papers

- Wang Z., Tauzin A.S., Laville E., Tedesco P., L'église F., Terrapon N., Lepercq P., Mercade M. and Potocki-Veronese G. 2020. Harvesting of prebiotic fructooligosaccharides by nonbeneficial human gut bacteria. *mSphere* 5:e00771-19.
- Wang Z., Tauzin A.S. and Potocki-Veronese G. Identification of functional glycoside transporters from uncultured bacteria. To be submitted.
- Wang Z., Tauzin A.S., Tedesco P., L'église F. and Potocki-Veronese G. Characterization of a fructooligosaccharide transporter from an uncultured human gut Clostridiales. In preparation.
- Tauzin A.S., Wang Z., Cioci G., Lippens G. and Potocki-Veronese G. Non-binding SusD protein is essential for xylooligosaccharide utilization in a human gut uncultured *Bacteroides*. In preparation.

## Review paper

- Wang Z., Tauzin A.S. and Potocki-Veronese G. Functional and structural diversity of glycoside transporters in bacteria. In preparation.

## Oral and poster communications

- Wang Z., Tauzin A.S., Laville E. and Potocki-Veronese G. Functional metagenomics for the discovery of glycoside transporters. International summer school on biocatalysis as a key enabling technology, 3-6 October, 2017, Siena, Italy (flash presentation and poster).
- Wang Z., Tauzin A.S., Laville E., Tedesco P., L'église F. and Potocki-Veronese G. Functional characterization of a FOS transporter system from an uncultured human gut *Dorea* species. CBM13 - Carbohydrate Bioengineering Meeting. May 19-22, 2019, Toulouse, France (flash presentation and poster rewarded as "best poster award").

- Wang Z., Tauzin A.S., Laville E., Tedesco P., L  isse F. and Potocki-Veronese G. Prebiotics harvesting by non-beneficial human gut bacteria. Functional Metagenomics 2019, 16-19 June, 2019, Trondheim, Norway (poster).

# **Introduction**



Carbohydrates are one of the major molecules in living systems. They seldom occur in Nature as monosaccharides (Prestegard et al., 2017). The glycosyl residues are indeed frequently linked via glycosidic linkages to other glycosyl or non-glycosyl groups, to form glycosides (Cummings and Stephen, 2007). Glycosides represent the main energy sources of bacteria, whatever their habitats are (Flint et al., 2008; Grondin et al., 2017; Ladevèze et al., 2017). To face their structural diversity, bacteria have evolved a large panel of highly specific enzymes and transporters to efficiently capture and metabolize them. The knowledge of the specificities and mechanisms of transporters is an important prerequisite to understand how bacteria metabolize glycosides, and thus, colonize their habitats. This is an important field for human and animal nutrition and health, since the type of bacteria which selectively ferment glycosides in the gut, influences the well-being and health of the host (Koropatkin et al., 2012). Besides, from a biotechnological point of view, robust processes of plant biomass transformation into high-value products rely on the use of performant chassis strains with engineered transporters and metabolic networks (Lee et al., 2013; Lian et al., 2014). In this chapter, we will briefly present the diversity of glycosides and of carbohydrate-active enzymes, and report the state of the art on glycoside-transporters. This last section is the first draft of a review paper on the functional, structural and taxonomical diversity of glycoside-transporters, and on the methods to characterize their specificities.

## **1. Diversity of glycosides**

As introduced upper, the structural diversity of glycosides on our planet is huge. Considering only a hexasaccharide, the theoretical number of possible isomers exceeds  $10^{12}$  (Laine, 1994). However, a recent study indicated that the breakdown of natural glycans requires several thousand enzyme combinations, suggesting that the real diversity of glycans (glycosides only made of glycosyl units) is probably a tiny fraction of the astronomical number based on theoretical combinations of monosaccharides and possible glycosidic bonds (Lapôtre et al., 2019). The types of glycosides accessible for the heterotrophic bacteria also depends on the habitats. Here, I briefly present the main characteristics of the glycosides from plants, mammals, microbes and arthropods, with a specific highlight on those present in mammal gut microbiomes, on which my thesis work was focused. The term ‘microbiome’ refers to the entire habitat, including the microorganisms (bacteria, archaea, lower and higher eukaryotes, and viruses), their genomes (i.e., genes), and the surrounding environmental conditions (Marchesi and Ravel, 2015).

## **1.1 Glycosides from plants and algae**

Plants are mainly composed of polysaccharides and glycosides, in the form of carbon storage carbohydrates (e.g. starch and sucrose) and cell structuring ones (e.g. cellulose, hemicelluloses, pectin and arabinogalactan proteins (AGPs)) (McCann and Knox, 2011; Selvendran, 1984). Cellulose consists of a linear chain of  $\beta$ -1,4-linked glucosyl residues, while hemicelluloses include a larger diversity of polysaccharides with more complex structures, such as xyloglucans, mannan and heteromannans, xylan and heteroxylans. Pectin structures are even more diverse and complex, consisting of large galacturonic acid-containing complex polysaccharides containing three major structural units: homogalacturonan (HG), rhamnogalacturonan I (RGI) and rhamnogalacturonan II (RGII) (Ridley et al., 2001). A key feature of AGPs is the presence of protein backbone attached with polysaccharide structural components (Tan et al., 2012). Besides, in the marine and hypersaline biotopes, plants and algae cell walls and mucus contain major amounts of anionic (sulfated) polysaccharides (Patil et al., 2018; Xu et al., 2017), due to a specialization for processes particularly accentuated in the aquatic environments (Colombo et al., 1983), like ion exchanges. Algae also contain high amount of species-specific storage polysaccharides, such as laminarin in brown algae (Patil et al., 2018). These algal glycosides represent an important carbon source for marine bacteria (Kappelmann et al., 2019). In the land environment, these diverse plant glycosides (mostly those from land plants) make up a major part of the diet for herbivorous and omnivorous mammals. Being not or incompletely utilized by the host enzymes, they, thus, constitute a vast array of carbon sources that is exploited by gut bacteria (Cockburn and Koropatkin, 2016; Hemsworth et al., 2016). Many other microorganisms, like soil bacteria (Zhu and McBride, 2017) and fungi (Barbi et al., 2016) can also utilize plant glycosides, which represents a key step in the carbon cycle on Earth (Wilson, 2008). Finally, the most efficient plant degraders and their enzymes constitute a valuable battery of biotechnological tools for plant biomass biorefinery (Lee et al., 2019).

## **1.2 Mammalian glycosides**

Mammals produce structurally diverse and complex glycosides, mainly glycosaminoglycans (GAGs) in the joints and connective tissues, epithelial cell-derived N-glycans and mucin-O-glycans (proteoglycans), glycogen as energy reserve polymer and human milk oligosaccharides (HMOs). GAGs are attached to proteins on cell surfaces and extracellular matrices. Based on the difference in the core disaccharide structure, GAGs are classified into four groups: hyaluronic acids, chondroitin sulphate or dermatan sulfate, heparin

or heparan sulphate, and keratan sulfate (Prydz, 2015). N-glycans derived from the epithelial cells are composed of a conserved pentasaccharide core capped with a diverse range of glycosyl residues, consisting of three main types: complex, high mannose and hybrid (Parry et al., 2006). O-glycans, such as those present in mucin, involve an *N*-acetylgalactosamine (GalNAc) residue linked to the hydroxyl group of either a serine or a threonine, to form the Tn antigen (GalNAc $\alpha$ 1-S/T) (Brockhausen and Stanley, 2017; Li and Chai, 2019). The O-GalNAc motif, named core 1, is the common core structure of many longer and more complex O-glycans. Another common core structure of O-glycans (core 2) contains a branched *N*-acetylglucosamine (GlcNAc) residue attached to core 1. There are also some other core structures extended with one or two GlcNAc and/or GalNAc residues based on core 1, although less common. In addition, human milk contains diverse oligosaccharides of which the structure is similar to that of the mucin O-glycans. Glycogen is a branched polymer consisting of linear chains of glucosyl residues linked by  $\alpha$  (1, 4) glycosidic bonds, with an average chain length of approximately 8 to 12 glucose units. Branches are linked to the chains by  $\alpha$  (1, 6) glycosidic bonds (Deng et al., 2016). HMOs contain a lactose motif at the reducing end, which can be sialylated and/or fucosylated, and elongated and branched with repeats of lacto-*N*-biose I (Gal $\beta$ 1–3GlcNAc) or *N*-acetyl-lactosamine (Gal $\beta$ 1–4GlcNAc) (Ayechu-Muruzabal et al., 2018). This backbone can also be modified with one or more fucosyl or sialic acid residues (Boehm and Stahl, 2007). In infants, HMOs play a crucial role in the installation of the gut microbiota (Ayechu-Muruzabal et al., 2018). Nevertheless, in the latter stages of life, the diverse mammalian glycans found in the human gut, coming from a range of sources, including dietary animal products and host molecules, represent an alternative carbon source to plant fibers for gut bacteria. Other body sites rich in human glycosides can also be accessible for bacteria, such as the skin (Reichenbach et al., 2018), the oral cavity (Renzi et al., 2011), the genito-urinary tract (Alteri et al., 2009), and the respiratory system (Hobbs et al., 2018).

### **1.3 Microbial glycosides**

Microbes (fungi and bacteria)-derived glycosides are mainly cell wall peptidoglycans and lipopolysaccharides, and exopolysaccharides. They are structurally extremely diverse, with over 100 different monosaccharide constituents identified (Kamerling and Gerwig, 2007). Even if, with this strategy of structural diversification of cell wall and extracellular glycosides, microbes have developed numerous ways of protecting themselves from degradation by other species, some members of the human gut microbiota can break down the glycoside outer layer

of other microbes, for example some bacteria (van Bueren et al., 2015) and yeast (Cuskin et al., 2015). Besides, in all fermented dietary products, microbial glycosides are also present, and can be utilized by the human gut bacteria (Cockburn and Koropatkin, 2016).

#### **1.4 Arthropoda glycosides**

The exoskeleton of many arthropods such as insects, spiders, crabs, shrimps and crustaceans is made of chitin, a polymer of  $\beta$ -1,4-linked GlcNAc (Beier and Bertilsson, 2013), which is also present in the cell wall of fungi (Arroyo et al., 2016). Many bacteria, especially in the soil and marine environments, are efficient chitin degraders (Meiborn et al., 2004; Wieczorek et al., 2019).

## **2. Carbohydrate-active enzymes**

The enzymes which are able to modify, synthesize or degrade glycosides are named carbohydrate-active enzymes (CAZymes). They are listed and classified in the CAZy database created in 1991 (<http://www.cazy.org/>) (Lombard et al., 2014). The CAZy classification is based on sequence similarities. It reflects similarities in the fold and in the catalytic machinery between related sequences (Davies and Henrissat, 1995). Importantly, a CAZy family can contain members with different functions.

CAZymes contain one or several catalytic domains. Non-catalytic, carbohydrate binding modules are also often associated with catalytic modules on a single polypeptidic chain. They enhance the associated module's activity by facilitating the association with its substrate (Boraston et al., 2004). In the CAZy database, the catalytic modules are classified into different families of:

- Glycoside Hydrolases (GHs), which perform hydrolysis of glycosidic bonds between two carbohydrate monomers, or between a carbohydrate and another molecule. Some GHs also catalyse transglycosylation reactions (Henrissat and Davies, 1997).
- GlycosylTransferases (GTs) are involved in the synthesis of glycosidic linkages, using various sugar-1-phosphate derivatives as glycosyl donor and an acceptor, which can be a carbohydrate or another class of hydroxylated molecules (Coutinho et al., 2003).

- Polysaccharide Lyases (PLs) cleave glycosidic bonds in uronic acid-containing polysaccharides, by using a  $\beta$ -elimination mechanism, which does not require a water molecule (Lombard et al., 2010).
- Carbohydrate Esterases (CEs) perform O- or N- deacetylation reaction on esterified sugars (Jayani et al., 2005).
- Auxiliary Activities (AAs) are modules performing synergetic redox reactions with other CAZymes. AA families contain the lytic polysaccharide mono-oxygenases (LPMOs), which degrade polysaccharides by an oxidative mechanism (Levasseur et al., 2013).

### **3. Functional and structural diversity of glycoside transporters in bacteria**

In addition to CAZymes, transporters play a critical role in glycoside utilization by bacteria. The advances in these last years in the development of culturing, sequencing and computing technologies results in an increasing number of glycoside-transporters sequences (Gelfand and Rodionov, 2008), of which the function can be annotated with more or less reliability, based on sequence similarity. Predicted and characterized glycoside transporters are listed and classified in several databases, in particular the recently created Transporter Classification Database (TCDB, <http://www.tcdb.org/>) (Saier, 2006). Based on sequence similarity, transporter sequences can easily be assigned to one of the four different superfamilies of glycoside transporter proteins, which present different mechanisms and architectures: the ATP-binding cassettes (ABC), the major facilitator superfamily (MFS) transporters, the phosphotransferase systems (PTS) and the TonB-dependent transporters (TBDT). Nevertheless, the prediction of their substrate specificity remains a challenging task. Their specificity towards glycosides may indeed vary dramatically among members of the same family (Diallinas, 2014). The transport of glycosides in bacteria was reviewed several times these last years, each time with a specific focus on (i) the human gut (Cockburn and Koropatkin, 2016; Schwalm and Groisman, 2017); (ii) plant polysaccharides and prebiotics (Goh and Klaenhammer, 2015); or (iii) the TBDT family of transporters (Bolam and van den Berg, 2018; Grondin et al., 2017). Here, we present the structural, taxonomical and functional diversity of glycoside transporters in bacteria, with a particular focus on the different methods used to characterize their functions, with their advantages and their limits. The systems involved in the transport of monosaccharides, such as the carbohydrate utilization loci

containing TonB-dependent transporters (CUT loci) (Boulangier et al., 2010), are not included in this chapter.

### **3.1. Techniques for functional characterization of glycoside-transporters**

Transporter sequences can be assigned to a specific superfamily, family or sub-family thanks to the presence of family-specific motives, and by sequence similarity to already characterized transporters, of which the diversity will be presented hereafter.

However, as members of the same family can transport various glycosides, it is difficult and often impossible to infer transporter specificity from sequence analysis alone, without experimental evidence. Moreover, there could also be transporters with unassigned function because of their low sequence similarity with characterized transporters, such as the 2.A.1.85 family of MFS transporters listed in TCDB. Thus, a set of methods and techniques have emerged to identify the correct substrate for a given predicted glycoside-transporter. We classified them into three categories: analysis of growth phenotypes, determination of binding specificity, and determination of uptake specificity.

#### **3.1.1. Analysis of growth phenotypes**

The most trivial method to screen the ability of a strain to utilize a glycoside, and thus, to internalize its substrate or one of its components, is to perform growth assays by monitoring the turbidity of the culture medium. Glycoside utilization can also be easily detected by using MacConkey indicator plates, which is a solid differential media used to selectively isolate enteric Gram-negative bacteria (such as *Escherichia coli*) and differentiate lactose fermenting Gram-negative rods as the positive colonies turn red when they could utilize the lactose (Peng et al., 2009).

The choice of glycosides to be tested can be rationalized thanks to genome sequencing, to identify putative operons coding for complete machineries of glycoside sensing, binding, translocation through the cell membrane and degradation by CAZymes (Lombard et al., 2014). The type of substrate (for example pectins, xylans, mannans, cellulose, starch, O- and N-glycans, etc.) targeted by an operon can be predicted from the putative specificities of each of the CAZymes it encodes. Enzyme specificities are inferred by sequence homology with the numerous biochemically characterized members from the same family, which are referred in the CAZy database (<http://www.cazy.org>) (Lombard et al., 2014). Such genomic environment analyses recently allowed the prediction of the specificity of three ABC transporters from

*Thermus thermophilus* HB8 towards oligosaccharides composing starch and  $\beta$ -glucans, the polysaccharides targeted by the CAZymes surrounding the transporter in each locus (Chandravanshi et al., 2019).

The up-regulation of the transporter encoding gene (or of the entire operon) during growth on a particular glycoside is a good indication that this transporter is involved in the utilization of this substrate. Quantification of gene expression levels requires transcriptomic analyses, either by quantitative real-time PCR (qPCR) (Tauzin et al., 2016b) or RNA chips (Martens et al., 2011), DNA microarrays (Andersen et al., 2012) or more recently, RNA sequencing (Despres et al., 2016). With the improvement in the sensitivity and decrease of cost, large-scale transcriptomic studies can be undertaken to compare the transcriptomes of thousands of bacteria (Lowe et al., 2017), giving indications on the specificity of putative transporters towards glycosides. However, transcriptomic analysis does not provide the experimental proof of the specificity of a transporter, for three reasons. First, the transcript levels might not correlate well with protein levels (Liu et al., 2016). In this case, the coupling of transcriptomics and proteomics strengthens the hypotheses regarding the function of the glycoside transporters (McNulty et al., 2013). Secondly, the glycoside detected by the transcription regulator differs from the one which is translocated through the cell membrane (Schwalm and Groisman, 2017). Finally, functional redundancy is frequent in bacteria, in which genes encoding several transporters can be upregulated and even overproduced during utilization of a complex glycoside (Kabisch et al., 2014). In any case, the involvement of a transporter in the utilization of a specific glycoside can be proven only after deletion or inactivation of the target gene (Andersen et al., 2012; Chen et al., 2015; Webb et al., 2008), to abolish or reduce the growth.

As described hereafter for each glycoside-transporter family, growth assays, often coupled to transcriptomic analyses, are mostly performed with native strains. However, in order to avoid functional redundancy and to facilitate the genetic manipulations to inactivate or delete genes, the function of transporters can also be characterized in recombinant strains.

For example, heterologous expression of the SusC/D gene pair from the inulin using type strain *Bacteroides thetaiotaomicron* 8736 in the non-inulin using type strain *B. thetaiotaomicron* VPI-5482 resulted in inulin use by the recipient strain, proving that the SusC/D from *Bacteroides thetaiotaomicron* 8736 is involved in the utilization of inulin (Joglekar et al., 2018). Also, a locus encoding an ABC transporter and CAZymes from

*Streptococcus mutans* LT11 conferred the ability of *E. coli* to ferment melibiose (Tao et al., 1993). And recently, in our lab, Tauzin et al. extended this approach to the study of transporters from uncultured bacteria (Tauzin et al., 2016b). By cloning various truncated forms of an entire metagenomic locus encoding an MFS, a SusC/D transporter and at least one xylosidase in a fosmid, Tauzin et al. demonstrated the specificity of each of these transporters towards linear xylooligosaccharides of various length (Tauzin et al., 2016b). Furthermore, in our lab Dagkesamanskaya et al. also developed an assay to perform recombinant *E. coli* growth assays in pico-liter droplets thanks to millifluidics, in order to decrease the amount of substrate required to screen for transporter functionality and specificity by several orders of magnitudes (Dagkesamanskaya et al., 2018).

Whether in native or recombinant strains, in any cases, the ability and efficiency of a bacterium to grow on glycosides are related to the levels of production and to the combined efficiency of several proteins to sense, bind, transport, degrade and metabolize the substrate. By analyzing the growth phenotype only, it is thus impossible to quantify the efficiency of a transporter and thus, to precisely profile its specificity towards glycosides of defined structures. This requires other methods, which are described hereafter.

### **3.1.2. Determination of binding specificity**

For transporters involving binding proteins, named substrate-binding protein (SBP) and surface glycan-binding proteins (SGBPs) in ABC and SusCD transporters, respectively, the investigation of the binding protein characteristics is easier than studying those of the transmembrane protein. Indeed, binding proteins are not transmembrane proteins, which facilitates their recombinant production and analysis. After having successfully purified the recombinant binding proteins, their binding specificity towards substrates can easily be screened with affinity gel electrophoresis (AGE) (Mystkowska et al., 2018) or differential scanning fluorimetry (DSF) (Boucher and Noll, 2011), and quantified by isothermal titration calorimetry (ITC) (Leth et al., 2018), surface plasmon resonance (SPR) (Theilmann et al., 2019), fluorescence spectroscopy (Phansopa et al., 2014) or saturation transfer difference nuclear magnetic resonance (STD-NMR) (Bell et al., 2019). As the principles of these techniques are different, each has its own strengths and weaknesses. Affinity gel electrophoresis is a simple and rapid method for the analysis of protein-binding affinities to soluble polysaccharides, based on reduced mobility of the protein interacting with substrates (Tomme et al., 2000). However, it cannot detect the binding with oligosaccharides which are

often the target substrates of the transporter. The ITC, DSF, SPR, STD-NMR and spectroscopic approaches all rely on the detection of a change in a physicochemical property of the ligand or the protein because of the binding (Vuignier et al., 2010). While the ITC, DSF, STD-NMR and spectroscopy necessitate protein in solution, SPR experiments imply surface immobilization of biomolecules. Thus, binding of the protein to the sensor chip surface may involve loss of structural integrity and function impacting negatively the interactions between the protein and its ligand (Nguyen et al., 2015). DSF rarely provides a dissociation constant at physiological temperatures (Grøftehaug et al., 2015). STD-NMR can detect the reduction of free ligand signal from the ligand exchanging between free and bound state when there is binding between ligand and protein (Viegas et al., 2011). It has rarely been used to characterize glycoside binding proteins (Bell et al., 2019). ITC is one of the most common methods to study the binding specificity of glycoside binding proteins, since it allows direct determination of the association constant by measuring the heat exchange during the complex formation at a constant temperature (Bjelić and Jelesarov, 2008). However, the large amount of soluble protein required for accurate measurements is an important issue of ITC, especially for certain proteins that could be hard to prepare in large amounts. SPR, which is also widely used for the investigation of binding kinetics and strengths, consists of the measurement of the change in the refractive index near a sensor surface. It only requires a few micrograms of protein. The equipment and maintenance costs of these techniques might be discouraging (Jecklin et al., 2009). The steady-state tryptophan spectroscopic methods allow measurement of binding affinity constants using relatively low amounts of protein, and less expensive equipment compared to SPR and ITC (Niesen et al., 2007). However, the spectroscopic methods are mainly for high-affinity binding sites, in the nanomolar range (Vuignier et al., 2010). Finally, all these techniques require purified proteins, and thus, optimization of the recombinant protein production. It is noteworthy that the binding specificity of these transporters doesn't lead to the direct determination of the uptake specificity. For example, MalE, the SBP of the maltose ABC from *E. coli* K-12 binds  $\alpha$ - and  $\beta$ -cyclodextrins, while these substrates cannot be up taken by the maltose ABC transporter (Mächtel et al., 2019; Quioco et al., 1997).

### **3.1.3. Determination of uptake specificity**

The most generic method to determine transport specificity for all transporter families (including those without binding proteins) and without any bias due, for example, to CAZyme activity, is based on the analysis of the accumulation of radiolabeled compounds in cells via

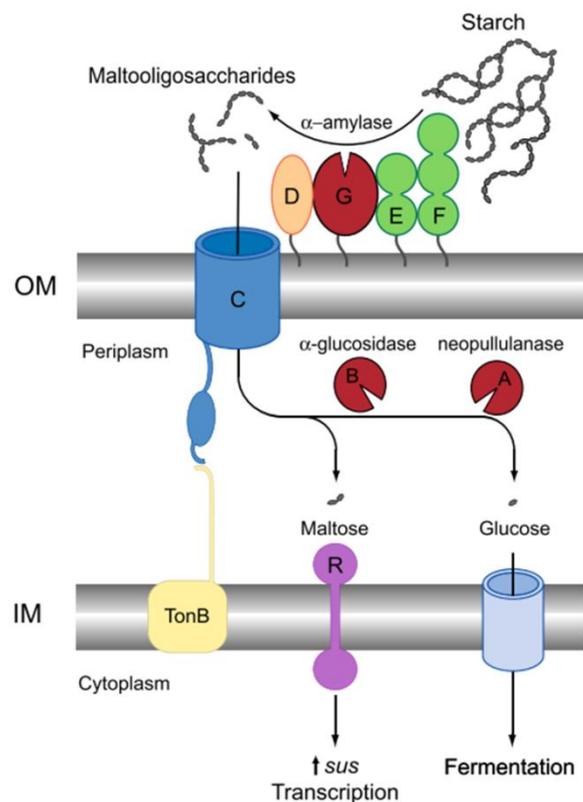
uptake assays (Martin and Russell, 1987). This quantitative method is also highly sensitive. It can indeed detect transport events in the micromolar range in less than 10 minutes. However, it requires a specific secured environment to manipulate radiolabeled compounds. In addition, this method is mainly used for monitoring monosaccharide uptake (Blanvillain et al., 2007), since most of glycosides are not available in the radioactive form or are prohibitively expensive (for example, 12,000 USD per gram for D-Lactose-1-<sup>13</sup>C monohydrate), which prevents large-scale screening campaigns. As explained before, in native strains, the multiplicity of carbohydrate transporters can bias the results of the screening of specificity. To solve this problem, the targeted transporter can be produced in recombinant strains deficient in the function of interest (Wei et al., 2012), or in membrane vesicles or proteoliposomes (Cao et al., 2011; McCoy et al., 2016). Nevertheless, these approaches can easily fail when the transport proteins are not active due to targeting to the membrane, insertion or folding issues, during the purification or reconstitution process (Majd et al., 2018).

Less generic methods also exist, focusing on particular transport mechanisms involving energy-coupling or substrate modification. The phosphoenolpyruvate (PEP-dependent) glycoside phosphorylation, which is a specific feature of PTS transporters, can be tracked by analyzing the amount of pyruvate in cell-free systems, and to deduce the amount of phosphorylated glycosides (Martin and Russell, 1987). The pyruvate amount can be monitored by quantifying the product of NADH oxidation by the lactate. For ABC transporters, the uptake of glycoside requires the energy of ATP binding and hydrolysis. Thus, the specificity of ABC transporters can be determined by testing growth on the target substrate when the synthesis of ATP is inhibited with specific inhibitor such as iodoacetate (Kaplan and Hutkins, 2003).

### **3.2. TonB-dependent transporters and associated surface-glycan binding proteins**

From late 1980s, the Salyers' group identified and characterized a locus involved in the utilization of starch in *Bacteroides thetaiotaomicron* 5482 (Tancula et al., 1992) and later named starch utilization system (SUS) (D'Elia and Salyers, 1996). In this strain, the SUS locus, named polysaccharide utilization locus (PUL (Bjursell et al., 2006)) is composed of eight genes named *susRABCDEFG* (**Figure 1**). First, the recognition and the binding of starch at the cell surface are mediated by outer membrane proteins consisting of SusD to F. Then, starch is initially degraded into malto-oligosaccharides by the surface lipoprotein SusG

corresponding to a GH13  $\alpha$ -amylase (Shipman et al., 1999). Then, the oligosaccharides are internalized via the TonB-dependent transporter (SusC) (Reeves et al., 1996) before their subsequent degradation by periplasmic GHs (SusA and SusB) (D'Elia and Salyers, 1996; Smith and Salyers, 1991). Finally, SusR, which is an inner membrane-spanning sensor/regulator protein, recognizes periplasmic maltose and triggers the upregulation of the SUS genes (D'Elia and Salyers, 1996). However, generally speaking, PULs are mostly regulated by three kinds of transcription factors including the SusR-like regulators, the AraC-like hybrid two-component systems (HTCSs), and the extracytoplasmic function sigma ( $\sigma$ ) factor/anti-sigma factors (Ravcheev et al., 2013). Both SusC and SusD have been identified as the main contributors of substrate uptake, while SusE and SusF, which are not present in all PULs, have auxiliary roles in substrate recognition and binding.



**Figure 1. Schematic representation of the starch utilization system in *Bacteroides thetaiotaomicron* 5482.**

This figure shows that the inner membrane-spanning protein SusR senses the disaccharide inducer, maltose, in the periplasm and subsequently drives the transcription of *susABCDEFG*. Starch is bound to the surface of the cell by the starch-binding outer membrane lipoproteins SusDEF. Subsequent hydrolysis by a similarly membrane-tethered  $\alpha$ -amylase, SusG,

generates oligosaccharides small enough to transit through the TonB-dependent transporter. Once in the periplasm, SusA and SusB, a neopullulanase and  $\alpha$ -glucosidase, respectively, process oligosaccharides into glucose. The monosaccharide is then shuttled into the cytoplasm by an unknown transporter (from Foley et al., 2016).

In the past decade, several *Bacteroides* SUS-like systems have been deeply characterized (**Table 1**), targeting large diversity of polysaccharides including xyloglucan, xylans, mannans,  $\beta$ -glucans, arabinogalactans, pectins, laminarin, levan, porphyran and heparin sulphate among others (Cuskin et al., 2015; Larsbrink et al., 2014; Luis et al., 2018; Mystkowska et al., 2018; Ndeh et al., 2017; Rogowski et al., 2015; Sonnenburg et al., 2010; Tamura et al., 2017). Some cellulolytic activities have been highlighted in PULs of rumen uncultured Bacteroidetes (Mackenzie et al., 2015; Naas et al., 2014; Pope et al., 2012), and the SusD-like proteins of these PULs were shown to strongly bind to cellulose (Mackenzie et al., 2012; Naas et al., 2014). Growth studies and transcriptomic analysis of these metagenomic species being impossible to perform, to date, there is no proof that the native strains containing these PULs are able to utilize cellulose.

The complexity of the PUL organization reflects the structural complexity of their cognate substrate, with the presence of genes encoding enzymes which are not CAZymes, such as sulfatases and peptidases, epimerase and members of the ROK (Repressors, ORFs and Kinases) family, and also additional transporters such as ATP-binding Cassette and Major Facilitator Superfamily transporters (Terrapon et al., 2018). For example, the utilization of yeast  $\alpha$ -mannan by *Bacteroides ovatus* is mediated by three PULs working in synergy and encoding more than a dozen of glycoside hydrolases, few phosphatases and glycosyltransferases (Cuskin et al., 2015). By contrast, the xyloglucan PUL from *B. ovatus* consists of 'only' a twelve genes PUL including eight GHs and one unique SusC/D pair (Larsbrink et al., 2014). However, characterization of syntenic xyloglucan PULs from other Bacteroidetes strains revealed slight differences in gene organization and enzyme specificity, conferring the ability to access to distinct plant-specific xyloglucan side chains (D'Éjean et al., 2019). Until now, the biochemical characterization of the PULs have been focused mainly on GHs and SGBPs proteins. As explained in the section 3.1. of this chapter, the functional characterization of the transmembrane component of SUS systems, the TBDT, is more difficult to perform.

SusC is a TBDT structurally characterized by a N-terminal plug domain inserted into a 22-stranded  $\beta$ -barrel. TBDTs are active transport systems mediating substrate-specific transport across the outer membrane. The binding of the substrate induces an allosteric rearrangement of the plug domain containing a conserved binding motif (Ton box) that is then released into the periplasmic space. There, it interacts with the C-terminal domain of TonB, while the N-terminal of TonB forms a complex with the inner membrane proteins ExbB and ExbD (TonB complex). The active transport of nutrients into the periplasm by TBDTs requires the energy derived from the proton motive force transmitted from the TonB complex (Larsen et al., 1999).

SusD-like proteins are cell surface glycan-binding proteins (SGBPs) involved in the capture of extracellular glycans. Interestingly, SusD-like proteins share no sequence similarity with known carbohydrate binding modules (CBMs) and are notably around twice larger (around 65 kDa) than CBMs (Foley et al., 2016). In 2008, the first crystal structure has been solved for a SusD involved in starch utilization from *Bacteroides thetaiotaomicron* VPI-5482 (Koropatkin et al., 2008). Up to date, the overall structure of SusD-like proteins consists of a region of  $\alpha$  helices organized into four tetratricopeptide repeat (TPR) units and a more variable region including many loops and short  $\alpha$  helices. The variable region is unique for each SusD-like protein. It depends on their cognate substrate and includes an optimized platform for specific substrate binding, mostly provided by at least two tryptophan residues. SusD-like proteins are attached to the outer membrane via a lipidated N-terminal cysteine projecting the protein above the surface of the membrane to facilitate glycan capture (Cameron et al., 2014; Koropatkin et al., 2008). Rarely, some SusD-like proteins do not bind efficiently the cognate substrate of the transporter. In these cases, substrate binding is ensured by additional SGBPs (D éjean et al., 2020; Luis et al., 2018).

In SUS, SusC physically interacts with SusD to form a complex which represents a minimal system when complemented with SusG for starch utilization by *B. thetaiotaomicron* VPI-5482 (Cho and Salyers, 2001). However, by using non-binding SusD-like mutants, substrate binding to SusD-like proteins was proved not to be a prerequisite for growth, since the expression of other SGBPs (SusE and SusF-like proteins from the same PUL) tends to supplement this loss (Cameron et al., 2014; Tauzin et al., 2016a). Deletion of the SusD-like proteins impairs growth on the cognate substrate, implying an essential structural role of the proteins in the SusC/D pair formation in some systems for which the binding ability is also not a prerequisite for transport of substrate (Koropatkin et al., 2008; Tauzin et al., 2016b). By

contrast, it has been shown recently that both substrate binding and physical presence of the SusD-like protein was demonstrated to be crucial for efficient utilization of mixed linkage glucans by *B. ovatus* (Tamura et al., 2019). To date, only two crystallographic structures of two different SusC/D complexes have been solved from *B. thetaiotaomicron* VPI-5482 (Glenwright et al., 2017). One pair (BT2263-64) is found in a four-component complex of unknown function, while the second one (BT1762-63) is part of a levan PUL (Sonnenburg et al., 2010). Glenwright et al. suggested a pedal bin mechanism for nutrient acquisition by SusC/D-like systems, with SusD forming the lid and SusC being the bin (Glenwright et al., 2017). In the absence of the ligand, SusD moves away from SusC in a hinge-like fashion in order to expose the substrate-binding site to the extracellular environment. After substrate binding, the SusD ‘lid’ swings in a closed state leading to a stabilized complex due to ligand interaction with both SusC and SusD. Then, the substrate moves into the SusC ‘bin’ preceding the binding of TonB to the TonB box and subsequent conformational changes for the release of the substrate into the cell. The structure of SusD is not altered in the closed state and the interaction with SusC includes the highly variable region that forms the substrate-binding site of the SusD-like proteins. According to the authors, this pedal bin mechanism would be in agreement with the envelopment of the whole ligand molecule which precludes the import of the undigested glycans with high molecular weights even if the size limit of the substrate that could be transported remains unknown. It is noteworthy that the SusC/D pair doesn’t often link with glycan utilization as some pairs are isolated on the genome or surrounded by proteins of unknown function (Bolam and van den Berg, 2018; Lap  bie et al., 2019). However, as discussed above, isolated SusC/D pair could also be working in synergy with other PULs in order to achieve the complete degradation of a complex polysaccharides such as pectin by *B. thetaiotaomicron* for which one of the three PULs includes only a SusC/D pair (RG-II PUL2) (Ndeh et al., 2017).

The presence of PULs is a unique feature in members of the Gram-negative phylum Bacteroidetes. The SusC/D pair was used as a hallmark for PUL identification and led to the establishment of the Polysaccharide-Utilization Loci DataBase (PULDB, <http://www.cazy.org/PULDB/>) (Terrapon et al., 2018, 2015). This database lists the experimentally characterized PULs (the ‘literature-derived PULs’), based on transcriptomic studies of cultured Bacteroidetes strains cultured on polysaccharides and proteoglycans, and also 35 632 predicted PULs mainly from 1154 Bacteroidetes species (release January 2020). However, TonB-dependent transporters specific for glycosides are also present in

Proteobacteria, such as in the plant pathogen *Xanthomonas campestris* pv. *campestris* (Blanvillain et al., 2007; Déjean et al., 2013).

Interestingly, depending on the strain and the substrate of interest, different mechanisms have been evolved by gut Bacteroidetes, leading to two opposite behaviors, selfish or distributive (Grondin et al., 2017). Co-culturing experiments of *B. thetaiotaomicron* with other *Bacteroides* strains revealed that the extracellular hydrolysis of substrate will benefit only to itself, while for *B. ovatus*, the partial breakdown products will diffuse into the extracellular environment and will benefit to the other species (Cuskin et al., 2015; Rakoff-Nahoum et al., 2014; Rogowski et al., 2015). These PULs thus confer an interesting advantage for environmental niche colonization, and thus, would play a role in the modulation of microbiota composition and the driving of community dynamics depending on the diverse mechanisms by which nutrients are used.

Recently, the SUS paradigm has been extended to members of the other dominant phylum Firmicutes (gpPULs) (Sheridan et al., 2016) finally leading to a general use of the term ‘PULs’ as bacterial loci encoding for at least one glycoside-degrading enzyme, one transport system and one transcriptional regulator .

**Table 1. Summary of the SusD and SusC/D pairs involved in polysaccharide utilization by Bacteroidetes, characterized either by biochemical, or by gene deletion or mutation approaches.**

Except if they were also characterized by one of these techniques, this list does not include the proteins of which the genes have been shown to be upregulated during polysaccharide utilization by transcriptomic analysis (listed in the PULDB).

| Protein name<br>(locus tag) | Substrate                      | Organism   | Characterization method                       | Structure   | Reference                                      |
|-----------------------------|--------------------------------|--|---|---|--|
| <b>SusD proteins</b>        |                                |  |   |   |  |
| BT3701                      | Starch<br>Maltooligosaccharide | <i>B. thetaiotaomicron</i>   | ITC<br>Gene deletion                          | 3CKC apo<br>3CKB maltotriose<br>3CK9 maltoheptaose<br>3CK8 $\beta$ -cyclodextrin<br>3CK7 $\alpha$ -cyclodextrin | (Koropatkin et al., 2008)                      |
| BT1043                      | N-acetyllactosamine            | <i>B. thetaiotaomicron</i>   | ITC   | 3EHM apo<br>3EHN N-acetyllactosamine  | (Koropatkin et al., 2009)                      |
| Uncultured species          | Cellulose                      | Uncultured Svalbard reindeer<br>rumen assigned to<br>Bacteroidetes | AGE   | -   | (Mackenzie et al., 2012)                       |
| BACOVA_02651                | Xyloglucan                     | <i>B. ovatus</i>   | AGE<br>ITC<br>Gene deletion<br>Point mutation | 5E75 apo<br>5E76 xylo-<br>glucooligosaccharide  | (Larsbrink et al., 2014; Tauzin et al., 2016a) |
| BT3984                      | Mannan                         | <i>B. thetaiotaomicron</i>   | ITC<br>Gene deletion                          | -   | (Cuskin et al., 2015)                          |

|                         |  |  |   |   |  |   |
|-------------------------|--|--|---|---|--|---|
| BACOVA_04392            | Birchwood<br>glucuronoxylan<br>Weath arabinoxylan            |  |   |   |  |   |
| BACOVA_03427            | Birchwood<br>glucuronoxylan<br>Corn<br>glucuronoarabinoxylan | <i>B. ovatus</i>                               | ITC   | -   |  | (Rogowski et al., 2015)                               |
| Fjoh_4558               | Chitin   | <i>Flavobacterium johnsoniae</i><br>ATCC 17061 | SDS-PAGE                                      | 5J90 apo<br>5J5U apo                                  |  | (Larsbrink et al., 2016)                              |
| Fjoh_4561               |  |  | ITC<br>Gene deletion                          |   |  |   |
| BACOVA_02095            | Galactomannan<br>Glucomannan                                 | <i>B. ovatus</i>                               | AGE<br>Gene deletion                          | -   |  | (BAGENHOLM et al., 2017)                              |
| BT3311                  | 1,6- $\beta$ -glucan   | <i>B. thetaiotaomicron</i>                     | AGE<br>ITC                                    | -   |  | (Temple et al., 2017)                                 |
| SAMN04488552_1347       | Laminarin  | <i>Gramella</i><br>sp. MAR_2010_102            | AGE<br>Point mutation                         | 6GCZ apo  |  | (Mystkowska et al., 2018)                             |
| BACOVA_02743            | $\beta$ (1,3)/ $\beta$ (1,4)-glucans                         | <i>B. ovatus</i>                               | AGE<br>ITC<br>Gene deletion<br>Point mutation | 6E60 apo<br>6DMF cellohexaose<br>6E61 heptasaccharide |  | (Tamura et al., 2019)                                 |
| BT3331                  | Glycosaminoglycan<br>chondroitin sulfate                     | <i>B. thetaiotaomicron</i>                     | ITC   |   |  | (Ndeh et al., 2020)                                   |
| <b>Complexes SusC/D</b> |  |  |   |   |  |   |
| BT2261-2264             | No   | <i>B. thetaiotaomicron</i>                     | ITC<br>Gene deletion                          | 5FQ6  |  | (Glenwright et al., 2017; Sonnenburg<br>et al., 2010) |
| BT1762-1763             | Levan  | <i>B. thetaiotaomicron</i>                     | ITC<br>Gene deletion                          | 5T3R  |  |   |

### 3.3. ABC transporters of glycosides

The ATP-binding cassette (ABC) is one of the largest super-family of transmembrane spanning proteins. ABC are present in many eukaryotes and prokaryotes, and even in some large viruses (Ford and Beis, 2019). These permeases are, in general, primary active transporters, which may act either as importers or as exporters, depending on the direction of translocation of the cargo molecules (Locher, 2016). They are able to transport various substrates in response to ATP hydrolysis (Mächtel et al., 2019). All ABC transporters have a core architecture constituted of two transmembrane domains (TMDs) and two cytosolic nucleotide-binding domains (NBDs) where ATP binds and is hydrolyzed (Beis, 2015). In prokaryotes, these four domains can be present in a combination of one or two multi-domain polypeptides (Theodoulou and Kerr, 2015; Wilkens, 2015). Glycoside importers also require the cooperation of an additional cognate partner: a substrate-binding protein, also named substrate specific binding component (both abbreviated SBP) (Theodoulou and Kerr, 2015; Wilkens, 2015). However, exceptions do exist in a subgroup of ABC transporters lacking SBPs, involved in the uptake of vitamins and micronutrients and referred as energy coupling factor (ECF) transporters (Erkens et al., 2012). At the beginning of my thesis, no glycoside ABC importer lacking a SBP had been described.

#### 3.3.1. Diversity of sequences and structures

Among the 34 families of ABC importers currently listed in the TCDB, transporters specific to carbohydrates have been clustered in two families: the carbohydrate uptake transporter-1 (CUT1 annotated with TC# 3.A.1.1) and the carbohydrate uptake transporter-2 (CUT2 annotated with TC# 3.A.1.2) (Saier, 2000). Members of the CUT1 family are specific to a variety of oligosaccharides of which the degree of polymerization varies between 2 and 7, as well as polyols, glycerol and glycerol-phosphate. Those of the CUT2 family transport only monosaccharides (Schneider, 2001). But it does not rule out the existence of exceptions, for instance a D-xylose and L-arabinose ABC transporter in *Haloferax volcanii* H26 belonging to the CUT1 family (Johnsen et al., 2019). In addition, members of the peptide/opine/nickel uptake transporter (PepT) family (3.A.1.5) were also found to be involved in glycoside transport (Andersen et al., 2012; Connors et al., 2005; Nanavati et al., 2006). Thus, we can expect that the number of families containing glycoside transporters will increase as new specificities of transport are discovered.

All the characterized glycoside transporters are SBP-dependent. Generally speaking, the CUT1 ABC transporters harness two separate transmembrane proteins interacting with a protein involving two copies of NBDs (Schneider, 2001). For glycoside ABC transporters of the PepT family, either the two TMDs or the two NBDs are in separate proteins.

The selectivity and specificity of carbohydrate ABC transporters are largely determined by the SBPs that are the most diverse components of the ABC systems, whereas the NBDs, and to a lesser extent, the TMDs, are clearly more similar. Indeed, SBPs have relatively low sequence similarity lower than 20%, and phylogenetic analyses based on multiple sequence alignments did not yield a stable tree (Berntsson et al., 2010). Traditionally, SBPs have been grouped into different groups on the basis of their substrate specificity, in which glycoside ABC transporter were clustered into the CUT1 families (Tam and Saier, 1993). In 2010, SBPs of ABC importers were reclassified into six clusters based on their available three-dimensional structures. SBPs binding to glycosides spread in three different clusters (B, C and D) (Berntsson et al., 2010). Recently, a seventh cluster was suggested to extend the structural classification, as the structure of fructooligosaccharide binding protein of ABC shows additional structural elements that define a larger ligand-binding cavity, with respect to the canonical SBPs, and an EF hand-like calcium-binding site (Culurgioni et al., 2017).

### 3.3.2. Functional diversity

Since the term ABC transporters was defined in 1990, a huge amount of proteins were annotated as ABC transporters, as it will be detailed in the next section (Holland, 2019). Meanwhile, only a small set of ABC transporters were experimentally proven to be involved in glycoside utilization by bacteria, by using one or several methods described in section 3.1. of this chapter (**Table 2**). In some cases, only SBPs function and structure were characterized, because these proteins are highly soluble and thus, easier to characterize compared to TMDs. However, substrate specificity determined by binding assays does not always correspond to the real transport specificity. The SBP, MalE, of the maltose ABC from *E. coli* K-12, binds  $\alpha$ - and  $\beta$ -cyclodextrins, while these substrates cannot be up taken by the maltose ABC transporter (Quioco et al., 1997). Moreover, in few cases only, the ABC transporters were deeply characterized, like those targeting maltooligosaccharides and alginate, by combining biochemical, molecular and structural approaches (Mächtel et al., 2019; Maruyama et al., 2019).

**Table 2. List of the bacterial ABC transporters involved in glycoside utilization, characterized either by binding or transport assays, gene deletion and/or transcriptional approaches.**

The glycoside ABC transporters retrieved from literature but which are not listed in the TCDB were classified based on their conserved domains by using a BLAST analysis against the NCBI's Conserved Domain Database (Marchler-Bauer et al., 2017).

| TCDB Family name | TCID       | Protein name (locus tag of SBP) | Substrate                       | Organism                                    | Characterization method   | Reference  |
|------------------|------------|---------------------------------|---------------------------------|---|---|--|
| CUT1             | 3.A.1.1.1  | MalEFGK (b4034)                 | Maltooligosaccharides (DP2-7)   | <i>Escherichia coli</i> K-12                | Gene deletion<br>Binding test<br>Transport assay<br>Crystal structure | (Oldham et al., 2013, 2007; Wandersman et al., 1979)                                   |
|                  | 3.A.1.1.2  | YufKLM (BSU34160)               | Galatooligosaccharides (DP2-4)  | <i>Bacillus subtilis</i> KM0                | Binding test  | (Watzlawick et al., 2016)  |
|                  | 3.A.1.1.9  | GuoEFGK                         | Xylooligosaccharides (DP2-6)    | <i>Geobacillus stearothermophilus</i> T-6   | Binding test  | (Shulami et al., 2007, 1999)   |
|                  | 3.A.1.1.10 | AlgSM1M2Q1Q2                    | Alginate (DP 2-7)               | <i>Sphingomonas</i> sp. strain A1           | Transport assay<br>Gene deletion<br>Crystal structure                 | (Kaneko et al., 2017; Maruyama et al., 2015; Mishima et al., 2003; Momma et al., 2000) |
|                  | 3.A.1.1.11 | TogABMN                         | Oligogalacturonides (DP2-3)     | <i>Erwinia chrysanthemi</i> 3937            | Gene deletion<br>Heterologous expression in <i>E. coli</i>            | (Hugouvieux-Cotte-Pattat et al., 2001; Hugouvieux-Cotte-Pattat and Reverchon, 2001)    |
|                  | 3.A.1.1.17 | ThuEFGK (CN159_00480)           | Trehalose, maltose, sucrose     | <i>Sinorhizobium meliloti</i> Rm1201        | Gene deletion<br>Transport assay                                      | (Jensen et al., 2002)  |
|                  | 3.A.1.1.18 | NacEFG                          | N-Acetylglucosamine, chitobiose | <i>Streptomyces olivaceoviridis</i>         | Transport assay<br>Binding test<br>Gene deletion                      | (Xiao et al., 2002)  |
|                  | 3.A.1.1.19 | PalEFGK                         | Platinose (Isomaltulose)        | <i>Agrobacterium tumefaciens</i> MAFF301001 | Gene deletion   | (De Costa et al., 2003)  |

|            |                          |  |   |  |   |
|------------|--------------------------|--|---|--|---|
| 3.A.1.1.20 | MsmEFGK<br>(D7H66_05730) | Fructooligosaccharides (DP3-5)                   | <i>Lactobacillus acidophilus</i> NCFM       | Transcriptional analysis<br>Gene deletion                    | (Barrangou et al., 2003)                            |
| 3.A.1.1.21 | BxlEFGK<br>(SCO7028)     | Xylobiose, xylotriose                            | <i>Streptomyces thermoviolaceus</i> OPC-520 | Transcriptional analysis<br>Binding test                     | (Tsujiibo et al., 2004)                             |
| 3.A.1.1.22 | MalE1E2FGK<br>(TM_1204)  | Maltose, maltotriose,<br>mannotetraose (MalE1)   | <i>Thermotoga maritima</i> MSB8             | Transcriptional analysis<br>Binding test                     | (Nanavati et al., 2005)                             |
| 3.A.1.1.22 | MalE1E2FGK<br>(TM_1839)  | Maltose, maltotriose, trehalose<br>(MalE2)       | <i>Thermotoga maritima</i> MSB8             | Transcriptional analysis<br>Binding test                     | (Nanavati et al., 2005)                             |
| 3.A.1.1.23 | CebEFG                   | Cellobiose, cellotriose                          | <i>Streptomyces reticuli</i> DSM 40776      | Transcriptional analysis<br>Gene deletion<br>Transport assay | (Schlösser et al., 1999)                            |
| 3.A.1.1.25 | -<br>(TTC1627)           | Trehalose, maltose, sucrose,<br>palatinose       | <i>Thermus thermophilus</i> HB27            | Transport assay<br>Binding test<br>Crystal structure         | (Chandravanshi et al., 2019;<br>Silva et al., 2005) |
| 3.A.1.1.26 | MalEFG<br>(BSU34610)     | Maltodextrin                                     | <i>Bacillus subtilis</i> 168                | Transport assay<br>Binding test<br>Gene deletion             | (Schönert et al., 2006)                             |
| 3.A.1.1.27 | MalXFGK<br>(SMU_1568)    | Maltodextrin (DP2-7)                             | <i>Streptococcus mutans</i> ATCC 700610     | Transport assay<br>Gene deletion                             | (Webb et al., 2008)                                 |
| 3.A.1.1.28 | MsmEFGK<br>(SMU_878)     | Raffinose, stachyose                             | <i>Streptococcus mutans</i> ATCC 700610     | Gene deletion  | (Webb et al., 2008)                                 |
| 3.A.1.1.31 | LpqY-SugA-<br>SugB-SugC  | Trehalose  | <i>Mycobacterium tuberculosis</i> H37Rv     | Transport assay<br>Gene deletion                             | (Kalscheuer et al., 2010)                           |
| 3.A.1.1.32 | GgtABCD<br>(slr0529)     | Sucrose, trehalose                               | <i>Synechocystis</i> sp. strain PCC6803     | Gene deletion<br>Binding test                                | (Mikkat and Hagemann, 2000)                         |
| 3.A.1.1.33 | DasABC                   | Chitobiose                                       | <i>Streptomyces coelicolor</i> A3(2)        | Transcriptional analysis<br>Gene deletion                    | (Saito et al., 2008)                                |
| 3.A.1.1.34 | AraNPQ                   | $\alpha$ -1,5-Arabinooligosaccharides<br>(DP3-4) | <i>Bacillus subtilis</i> IQB496             | Gene deletion  | (Ferreira and De SáNogueira,<br>2010)               |
| 3.A.1.1.41 | TreEGF                   | Trehalose  | <i>Thermotoga maritima</i> MSB8             | Binding test   | (Boucher and Noll, 2011)                            |
| 3.A.1.1.44 | MalEFG<br>(SCO2231)      | Maltose  | <i>Streptomyces coelicolor</i> A3(2)        | Gene deletion  | (Van Wezel et al., 1997)                            |

|            |                                    |   |   |   |                                   |
|------------|------------------------------------|---|---|---|-----------------------------------|
| 3.A.1.1.45 | MusEFGK<br>(Cgl2461)               | Maltose   | <i>Corynebacterium glutamicum</i> ATCC 13032                  | Transcriptional analysis<br>Gene deletion                             | (Henrich et al., 2013)            |
| 3.A.1.1.48 | MalEFGK<br>(BLJ_1686)              | Lacto-N-biose,<br>galacto-N-biose   | <i>Bifidobacterium longum</i> JCM1217                         | Binding test  | (Suzuki et al., 2008)             |
| 3.A.1.1.51 | MalEFGK<br>(Bd1227)                | Maltooligosaccharides (DP2-7)   | <i>Bdellovibrio bacteriovorus</i> ATCC 15356                  | Gene deletion<br>Binding test<br>Transport assay<br>Crystal structure | (Licht et al., 2019)              |
| 3.A.1.1.52 | GlsQPCD<br>(DSM107007_RS<br>23065) | Sucrose, maltose, esculin   | <i>Anabaena</i> sp. PCC 7120                                  | Gene deletion   | (Nieves-Mori ón and Flores, 2018) |
| 3.A.1.1.53 | RafEFGK<br>(BIF_01592)             | $\alpha$ -1,6-Glucosides and<br>galactosides (2-7)                        | <i>Bifidobacterium animalis</i> subsp. <i>lactis</i><br>BI-04 | Transcriptional analysis<br>Binding test<br>Crystal structure         | (Ejby et al., 2016)               |
| literature | XynEFG                             | Xylooligosaccharides (DP2-6)  | <i>Geobacillus stearothermophilus</i> T-6                     | Transcriptional analysis<br>Binding test                              | (Shulami et al., 2007)            |
| literature | CbpA                               | Celotriose  | <i>Clostridium thermocellum</i> DSM 1237                      | Binding test  | (Nataf et al., 2009)              |
| literature | CbpB                               | Cellooligosaccharides (DP2-5)   | <i>Clostridium thermocellum</i> DSM 1237                      | Binding test  | (Nataf et al., 2009)              |
| literature | CbpC                               | Cellooligosaccharides (DP2-5)   | <i>Clostridium thermocellum</i> DSM 1237                      | Binding test  | (Nataf et al., 2009)              |
| literature | CbpD                               | Cellooligosaccharides (DP3-5)   | <i>Clostridium thermocellum</i> DSM 1237                      | Binding test  | (Nataf et al., 2009)              |
| literature | LBP                                | Laminaribiose   | <i>Clostridium thermocellum</i> DSM 1237                      | Binding test  | (Nataf et al., 2009)              |
| literature | GachHFG                            | Acarbose,<br>maltooligosaccharides (DP2-4)                                | <i>Streptomyces glaucescens</i> GLA.O                         | Binding test<br>Crystal structure                                     | (Vahedi-Faridi et al., 2010)      |
| literature | AbnEFG                             | Arabinoooligosaccharides (DP3-8)  | <i>Geobacillus stearothermophilus</i> T-6                     | Binding test  | (Shulami et al., 2011)            |
| literature | XBP1                               | Xyooligosaccharides (DP2-4)   | <i>Caldanaerobius polysaccharolyticus</i><br>ATCC BAA-17      | Binding test<br>Crystal structure                                     | (Han et al., 2012)                |
| literature | BIAXBP<br>(Balac_0514)             | Xylooligosaccharides (DP2-6),<br>arabinoxylan-oligosaccharides<br>(DP3-5) | <i>Bifidobacterium animalis</i> subsp. <i>lactis</i><br>BI-04 | Binding test<br>Crystal structure                                     | (Ejby et al., 2013)               |
| literature | CpMnBP1                            | (Galacto-)<br>Mannoooligosaccharides (DP2-5)                              | <i>Caldanaerobius polysaccharolyticus</i><br>ATCC BAA-17      | Binding test<br>Crystal structure                                     | (Chekan et al., 2014)             |

|            |                              |  |  |  |   |
|------------|------------------------------|--|--|--|---|
| literature | -<br>(EUR_01830)             | Mannooligosaccharides (DP3-7), acarbose                              | <i>Eubacterium rectal</i> DSM 17629                    | Proteomic analysis<br>Binding test   | (Cockburn et al., 2015)                       |
| literature | -<br>(EUR_31480)             | Mannooligosaccharides (DP2-3)  | <i>Eubacterium rectal</i> DSM 17629                    | Proteomic analysis<br>Binding test   | (Cockburn et al., 2015)                       |
| literature | -<br>(EUR_21010)             | Cellooligosaccharides (DP2, 3)                                       | <i>Eubacterium rectal</i> DSM 17629                    | Proteomic analysis<br>Binding test   | (Cockburn et al., 2015)                       |
| literature | ManHIJ                       | (Galacto-) Mannooligosaccharides (DP3-5)                             | <i>Bacillus</i> sp. N16-5                              | Transcriptional analysis<br>Binding test<br>Gene deletion                      | (Song et al., 2016)                           |
| literature | FL-SBP                       | Fucosyllactose   | <i>Bifidobacterium breve</i> BR-A29                    | Gene deletion  | (Matsuki et al., 2016)                        |
| literature | FusABC<br>(SP_1796)          | Fructooligosaccharides (DP3-5)                                       | <i>Streptococcus pneumoniae</i> TIGR4                  | Transcriptional analysis<br>Binding test<br>Gene deletion<br>Crystal structure | (Culurgioni et al., 2017; Linke et al., 2013) |
| literature | SfuABC<br>(SPV_1585)         | Fructooligosaccharides (DP3-5)                                       | <i>Streptococcus pneumoniae</i> D39                    | Binding test<br>Gene deletion<br>Crystal structure                             | (Culurgioni et al., 2017)                     |
| literature | CpSBC4535                    | $\alpha$ -1,5- Arabinooligosaccharides (DP 3-6)                      | <i>Caldanaerobius polysaccharolyticus</i> ATCC BAA-17  | Transcriptional analysis<br>Binding test<br>Gene deletion                      | (Wefers et al., 2017)                         |
| literature | CuaABC<br>(Ccel_2112)        | Cellooligosaccharides (DP2-5)  | <i>Ruminiclostridium cellulolyticum</i> H10 ATCC 35319 | Binding  | (Fosses et al., 2017)                         |
| literature | -<br>(Smon_0123)             | Sulfated GalNAc  | <i>Streptobacillus moniliformis</i> DSM 12112          | Binding; Structure   | (Oiki et al., 2017)                           |
| literature | MdxEFG<br>(HMPREF9496_01133) | Maltooligosacchride (DP4-7)  | <i>Enterococcus faecalis</i> JH2-2                     | Gene deletion  | (Sauvageot et al., 2017)                      |
| literature | RiXBP<br>(ROSINTL182_08199)  | Xylooligosaccharides (DP3-6), arabinoxylotriose, arabinoxylotetraose | <i>Roseburia intestinalis</i> L1-82                    | Transcriptional analysis<br>Binding test                                       | (Leth et al., 2018)                           |
| literature | RiMnBP                       | Mannooligosaccharides (DP3-  | <i>Roseburia intestinalis</i> L1-82                    | Binding test   | (La Rosa et al., 2019)                        |

|      |            |                          |  |  |   |   |
|------|------------|--------------------------|--|--|---|---|
|      |            | (ROSINTL182_05479)       | 5)                                     |  |   |   |
|      | literature | BIMnBP1 (BLAC_00780)     | Mannooligosaccharides (DP3-6)          | <i>Bifidobacterium animalis</i> subsp. <i>lactis</i> ATCC27673 | Transcriptional analysis<br>Binding test<br>Crystal structure | (Ejby et al., 2019)                           |
|      | literature | BIMnBP2 (BLAC_00785)     | Mannooligosaccharides (DP3-6)          | <i>Bifidobacterium animalis</i> subsp. <i>lactis</i> ATCC27673 | Transcriptional analysis<br>Binding test<br>Crystal structure | (Ejby et al., 2019)                           |
|      | literature | Bal6GBP (Balac_0483)     | $\beta$ -(1,3/6)-Galactosides (DP2, 4) | <i>Bifidobacterium animalis</i> subsp. <i>Lactis</i> BI-04     | Binding test<br>Crystal structure                             | (Theilmann et al., 2019)                      |
| PepT | 3.A.1.5.16 | BglpEFGKL (TM_0031)      | Cellobiose, laminaribiose              | <i>Thermotoga maritima</i> MSB8                                | Transcriptional analysis<br>Binding test                      | (Connors et al., 2005; Nanavati et al., 2005) |
|      | 3.A.1.5.32 | F0TFS5-9 (LAC30SC_07285) | Stachyose                              | <i>Lactobacillus acidophilus</i> NCFM                          | Transcriptional analysis<br>Gene deletion                     | (Andersen et al., 2012)                       |

### 3.3.3. Taxonomical diversity

We examined the taxonomic diversity of all the characterized bacterial ABC transporters targeting glycosides from the CUT1 and PepT families (**Table 3**). These 57 glycoside ABC transporters are wide-ranging in Gram-positive and negative bacteria, including seven phyla and 16 orders from different environments, including some which cause human diseases, such as *Streptococcus mutans* and *Enterococcus faecalis* (Sauvageot et al., 2017; Webb et al., 2008). Of the 57 glycosides ABC transporter, more than 90% were assigned to the four phyla Firmicutes, Actinobacteria, Proteobacteria and Thermotogae. However, the actual taxonomic distribution of glycoside ABC transporters is highly biased by the sequencing and biochemical characterization efforts, which are unequal in their coverage of the bacterial diversity. Only by searching the domain (PF01547) of the SBP of maltose ABC transporter in the Pfam 32.0 database (El-Gebali et al., 2019), there were 25,735 sequences across 4,106 species, which highlights the huge diversity of ABC transporters targeting glycosides, and the fact that an extremely few amount of them have been characterized (**Table 3**). Notably, there is no glycoside targeting ABC transporters from the Bacteroidetes phylum in the TCDB, and we did not find any characterized in the literature. However, in the Pfam 32.0 database, 56 Bacteroidetes species harbor 74 sequences containing the Pfam domain of the SBP of maltose ABC transporter. In addition, the PULDB has recently included the annotation of ABC transporters, since they appear, with SusC/D transporters, in around 200 PULs (Terrapon et al., 2018). It is thus likely that some Bacteroidetes ABC transporters play a role in glycoside transport, but it remains to be proved. More generally speaking, the taxonomic diversity of the characterized glycoside-targeting ABC transporters is far from representing the truth, as most of the ABC transporters remain to be characterized.

**Table 3. Phylum and order distribution of the characterized glycoside-targeting ABC transporters in bacteria.**

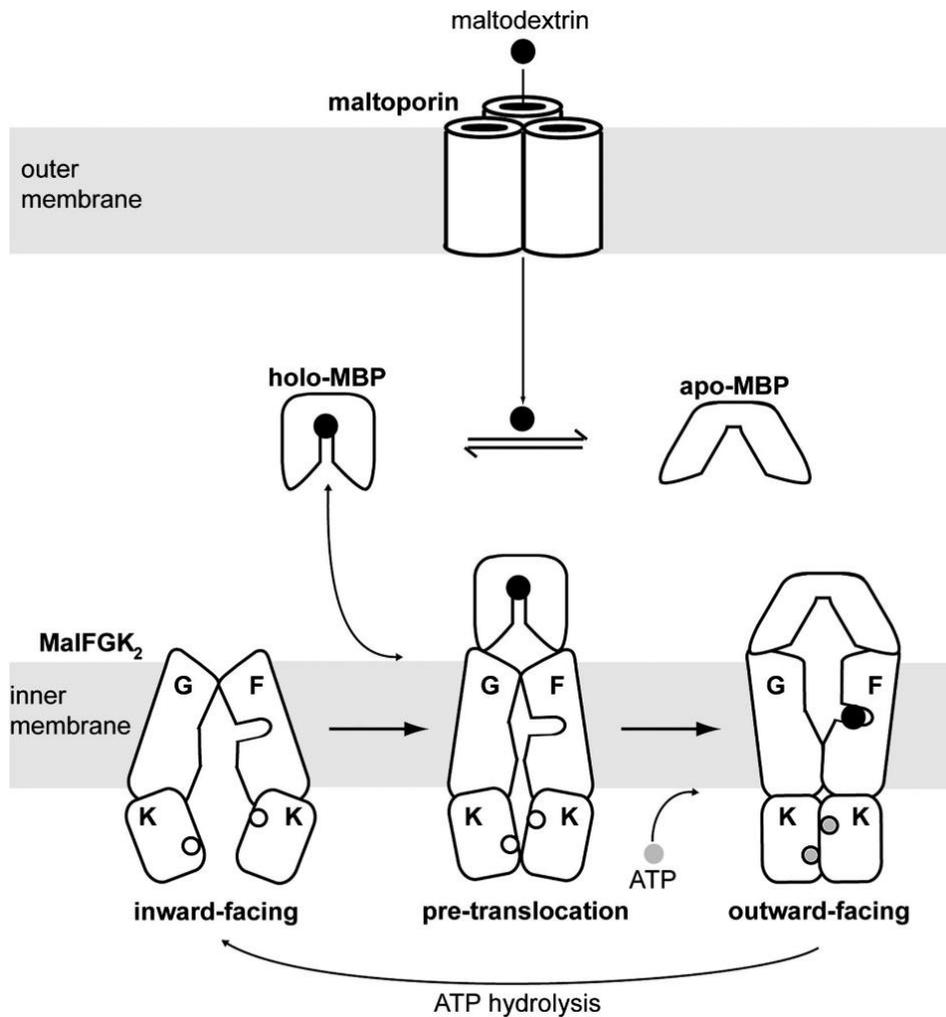
| Phylum              | Order                  | Number of characterized ABC | Percentage of the characterized ABCs |        |
|---------------------|------------------------|-----------------------------|--------------------------------------|--------|
|                     |                        |                             | Order                                | Phylum |
| Actinobacteria      | Actinomycetales        | 4                           | 7.0                                  | 21.1   |
|                     | Corynebacteriales      | 1                           | 1.8                                  |        |
|                     | Bifidobacteriales      | 7                           | 12.3                                 |        |
| Firmicutes          | Bacillales             | 6                           | 10.5                                 | 47.4   |
|                     | Lactobacillales        | 7                           | 12.3                                 |        |
|                     | Clostridiales          | 11                          | 19.3                                 |        |
|                     | Thermoanaerobacterales | 3                           | 5.3                                  |        |
| Proteobacteria      | Enterobacterales       | 3                           | 5.3                                  | 17.5   |
|                     | Rhizobiales            | 5                           | 8.8                                  |        |
|                     | Sphingomonadales       | 1                           | 1.8                                  |        |
|                     | Bdellovibrionales      | 1                           | 1.8                                  |        |
| Cyanobacteria       | Synechococcales        | 1                           | 1.8                                  | 3.6    |
|                     | Nostocales             | 1                           | 1.8                                  |        |
| Thermotogae         | Thermotogales          | 4                           | 7.0                                  | 7.0    |
| Deinococcus-Thermus | Thermales              | 1                           | 1.8                                  | 1.8    |
| Fusobacteria        | Fusobacteriales        | 1                           | 1.8                                  | 1.8    |

### 3.3.4. Mechanism of glycoside import by ABC transporters

Since the sequencing of the genes encoding the maltose ABC transporter in the 1970s (Silhavy et al., 1979), the functional and structural characterization of maltooligosaccharides transporters (especially in *E. coli*) resulted in the establishment of a glycoside ABC transporter paradigm. The import of maltooligosaccharides proceeds by a classical altering access mechanism and the substrate specificity is controlled by the outer membrane maltoporin, the periplasmic protein MalE (also named MBP, for maltose-binding protein, in maltooligosaccharide ABC transporters) and the transmembrane subunits MalFG (**Figure 2**) (Mächtel et al., 2019; Schneider, 2001).

The MalE consists of two globular lobes (the N-terminal and the C-terminal domains) between which a substrate-binding pocket is formed. The shape-complementarity, extensive

hydrogen-bonding, and aromatic stacking interactions together account for the high-affinity binding (in the micromolar range) of MalE to maltooligosaccharides (Quioco et al., 1997). The MalE conformation is predominantly open in the absence of ligand). Once the maltooligosaccharide passes maltoporin, MalE will completely trap the substrate at the interface of the two lobes and then dock to the transmembrane subunits and triggers the ATPase activity, initiating the transport cycle (Hospital et al., 2010; Mächtel et al., 2019). Crystal structures of the transporter complex MBP-MalFGK2 from *E. coli* K-12 bound with a maltooligosaccharide in two different conformational states revealed that i) in the pre-translocation structure, MalG forms two hydrogen bonds with maltooligosaccharide up to four glucosyl units at the reducing end, and ii) in the outward-facing conformation, the transmembrane subunit MalF binds glucosyl units from the nonreducing end (Oldham et al., 2013, 2007). Thus, here the MalFG provides another level of selection of the substrate, in addition to the ones due to LamB and MalE. Thus, the substrate can be unidirectionally imported with the switching of TMDs between inward- and outward-facing conformations, and ATP hydrolysis is stimulated by both MalE, which stabilizes the semi-open configuration of MalFGK2 and open MalE which stabilizes the closed state of MalFGK2 (Mächtel et al., 2019). The binding cavity of the pretranslocation state of the MBP-MalFGK2 complex, would be large enough to accommodate seven glucosyl units, which corresponds to the maximal DP of the oligosaccharides that can be transported (Oldham et al., 2013). The maltooligosaccharide ABC transporter could serve as a smart template to understand how the other glycoside ABC transporters evolve and process, despite their structural and functional diversity.

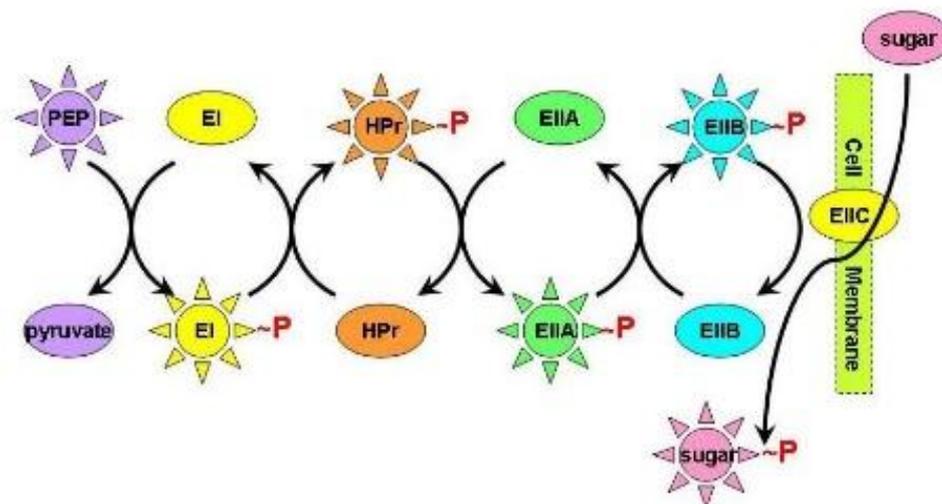


**Figure 2. Transport of maltooligosaccharides by ABC in Gram-negative bacteria.**

This figure shows the three components of the maltose transport system. Maltooligosaccharides diffuse across the outer membrane through maltoporin, bind with MBP in the periplasm, and are transported across the inner membrane by the ABC transporter MalFGK<sub>2</sub>. (from Oldham et al., 2013)

### 3.4. PTS transporters of glycosides

In 1964, the phosphotransferase system (PTS) was firstly described in *Escherichia coli* K235 as a novel system coupling sugar transport to sugar phosphorylation from phosphoenolpyruvate (PEP) as phosphoryl donor, in the laboratory of Saul Roseman (Kundig et al., 1964). Since then, PTSs have been identified in various species of Gram-positive and Gram-negative bacteria as well as in archaea, transporting a wide range of monosaccharides, glycosides and polyols (Saier, 2015). The PTS thus represents a major carbohydrate active transport system in bacteria (Saier, 2000). It is a multicomponent system which always involves enzymes of the plasmic membrane and others of the cytoplasm (Deutscher et al., 2006). The membranous component of the PTS system, enzyme II (EII), is made up of three to four subunits: IIA, IIB, IIC and sometimes IID, which are specific for one substrate or, in a few cases, for a small group of closely related carbohydrates (Erni, 2013). The cytoplasmic components are enzyme I (EI) and histidine protein (HPr), which are not carbohydrate specific. Together, they constitute a phosphorylation cascade allowing the phosphoryl group to be successively transferred from PEP to EI, to the HPr, then to the subunits A and B of EII, and finally to the carbohydrate imported by the subunit EIIC or by the EIICD complex (**Figure 3**) (Deutscher et al., 2014).



**Figure 3. Schematic illustration of the sugar transport and phosphorylation process by PTS.**

This figure shows that the phosphoryl group is transferred to the EI, HPr, EIIA and EIIB serially, which finally phosphorylates the imported sugar by the transmembrane EIIC. (from the website of Proteopedia, [https://proteopedia.org/wiki/index.php/Main\\_Page](https://proteopedia.org/wiki/index.php/Main_Page).)

### 3.4.1. Diversity of sequences and structures

Based on sequence similarities between the C domain of the EII complex, carbohydrate PTSs have been classified into seven families in the TCDB (Saier, 2006): glucose-glucoside (4.A.1, Glc), fructose-mannitol (4.A.2, Fru), lactose-N,N-diacetylchitobiose (4.A.3, Lac), glucitol (4.A.4, Gut), galactitol (4.A.5, Gat), mannose-fructose-sorbose (4.A.6, Man), and L-ascorbate (4.A.7, L-Asc) families. The Glc family was further divided into the glucose (4.A.1.1) and glucoside (4.A.1.2) subfamilies. It is noteworthy that some families include members which are nearly as distant from each other as they are from members from other families (Jeckelmann and Erni, 2019). The name of the PTS families is not always consistent with the function of their members. For example, cellobiose PTS transporters exist either in the Glc family (Francl et al., 2010) or in the Lac family (Kowalczyk et al., 2008). Based on our review of the carbohydrate PTSs in the TCDB and literature (**Table 4**), glycoside PTS transporters are only listed in the Glc, Lac and Man families.

Depending on the family, the EII complexes of the glycoside PTSs are encoded either by one or several encoding genes. The Glc family contains members with either a single multidomain protein including the A, B and C domains, or with two distinct proteins including one for the A domain and another one for the B and C domains. Some of the members of the Glc family lack their own IIA domain, and, instead, use the IIA domain of another system of the same family (Tchieu et al., 2001). By contrast, the members of the Lac family always have multidomain EII complexes classified in two subfamilies: one type (4.A.3.1) consisting of two polypeptide chains (EIIA and EIIBC) and another type (4.A.3.2) including three individual polypeptide chains (EIIA, EIIB and EIIC). The Man family is the only PTS family in which members possess an additional IID domain. The EIIA, B, C and D domains of glycoside PTSs in the Man family can be present either in four individual proteins or three individual proteins with the A and B domains in the same protein. The Glc, Lac and Fru families belong to a phylogenetic PTS-GFL superfamily (Nguyen et al., 2006), because of the structural homogeneity of their IIC domains, with a uniform ten transmembrane helices topology (Cao et al., 2011; McCoy et al., 2016). The members of the Man family are evolutionarily distinct from the GFL superfamily (Chen et al., 2012). In this family, the IIC domains probably consist of six or seven transmembrane helices, and the second integral membrane IID domains probably contains six transmembrane helices, based on the HMMTOP prediction (<http://www.enzim.hu/hmmtop/index.php>). However, the only solved structure of a mannose PTS transporter (ManYZ) from the Man family shows nine transmembrane helices in both the

IIC (ManY) and IID (ManZ) domains (Liu et al., 2019). The structural organization, and thus, classification of the glycoside PTSs have a complex evolutionary history because of the numerous horizontal gene transfer events, duplications, and non-orthologous displacements (Zúñiga et al., 2005).

### **3.4.2. Functional diversity**

Only few dozens of PTSs targeting glycosides were functionally characterized using one or several methods listed above (Table 4). Most of them are specific for  $\alpha$ - or  $\beta$ - linked glucosides, such as maltose, sucrose and cellobiose, or their derivatives such as N,N'-diacetylchitobiose and salicin, as the GFL superfamily is by far the largest (Jeckelmann and Erni, 2019). Recently, transcriptomic analyses revealed that, in strain *Lactobacillus plantarum* WCFS1, the PTS1 (Panwar and Kapoor, 2020; Saulnier et al., 2007) of the Glc family, and the PTS23C (Panwar and Kapoor, 2020) of the Lac family, might be involved in galactomannose/galactomannobiose and mannobiose utilization, respectively, while the specificity of the two PTSs still needs to be further characterized. In addition, in the Man family, there are also few PTSs which are responsible for  $\beta$ -galactosides,  $\beta$ -mannosides, hyaluronic acid and fucosyl- $\alpha$ -1,3-N-acetylglucosamine. Glycoside PTSs are thus mainly specific for disaccharides, although some are able to transport longer oligosaccharides with a DP not higher than four. Other PTS transporters in the list were showed to be involved in the utilization of hyaluronic acid, glucomannooligosaccharides or lichenan oligosaccharides by transcriptional analysis or deletion of one of the PTS encoding genes, but this does not mean that the PTS can transport glycosides with DP longer than four.

**Table 4. List of the bacterial PTS transporters involved in glycoside utilization, characterized either by binding or transport assays, gene deletion and/or transcriptional approaches.**

The glycoside PTS transporters retrieved from literature but which are not listed in the TCDB were classified based on their conserved domains by using a BLAST analysis against the NCBI's Conserved Domain Database (Marchler-Bauer et al., 2017).

| TCDB Family name | TCID       | Protein name (locus tag) | Substrate  | Organism                                | Characterization method   | Literature                               |
|------------------|------------|--------------------------|--|---|---|--|
| Glc              | 4.A.1.1.3  | MalX (b1621)             | Maltose  | <i>Escherichia coli</i> K12             | Heterologous expression in <i>E. coli</i> strain                        | (Reidl and Boos, 1991)                   |
|                  | 4.A.1.1.8  | MalP (BSU08200)          | Maltose  | <i>Bacillus subtilis</i> 168            | Transcriptional analysis<br>Gene deletion                               | (Yamamoto et al., 2001)                  |
|                  | 4.A.1.1.10 | AgIA (FYK33_RS28885)     | Isomaltose, trehalulose, maltulose, leucrose, palatinose | <i>Klebsiella pneumoniae</i> ATCC 23357 | Heterologous express in <i>E. coli</i> strain<br>Transport assay        | (Pikis et al., 2006)                     |
|                  | 4.A.1.1.11 | MalT (SMU_2047)          | Maltose, maltotriose                                     | <i>Streptococcus mutans</i> ATCC 700610 | Gene deletion   | (Webb et al., 2007)                      |
|                  | 4.A.1.1.12 | MalT (SPy_1986)          | Maltose, maltotriose                                     | Group A <i>Streptococcus</i> MGAS5005   | Transcriptional analysis<br>Gene deletion                               | (Shelburne et al., 2008)                 |
|                  | 4.A.1.1.17 | MalT (BCE33L0344)        | Maltose  | <i>Bacillus cereus</i> E33L             | Transport assay<br>Binding test<br>Crystal Structure                    | (McCoy et al., 2016)                     |
|                  | 4.A.1.2.1  | ScrA                     | Sucrose  | <i>Klebsiella pneumoniae</i> KAY2026    | Phosphorylation activity  | (Titgemeyer et al., 1996)                |
|                  | 4.A.1.2.3  | AscF (b2715)             | Salicin  | <i>Escherichia coli</i> K-12            | Gene deletion   | (Desai et al., 2010)                     |
|                  | 4.A.1.2.4  | TreB (b4240)             | Trehalose, maltose, sucrose                              | <i>Escherichia coli</i> K-12            | Heterologous expression in other <i>E. coli</i> strain<br>Gene deletion | (Klein et al., 1995; Steen et al., 2019) |
|                  | 4.A.1.2.5  | BglF                     | Methyl- $\beta$ -glucoside,                              | <i>Corynebacterium</i>                  | Heterologous expression in  | (Kotrba et al., 2003)                    |

|     |            |                       |  |  |   |                          |
|-----|------------|-----------------------|--|--|---|--------------------------|
|     |            | (cgR_2729)            | salicin, arbutin                       | <i>glutamicum</i><br>R                       | <i>C. glutamicum</i>  |                          |
|     | 4.A.1.2.6  | BglP                  | Aesculin, arbutin                      | <i>Streptococcus mutans</i><br>NG8           | Gene deletion<br>Heterologous expression in<br><i>E. coli</i> | (Cote et al., 2000)      |
|     | 4.A.1.2.9  | SacP<br>(BSU38050)    | Sucrose                                | <i>Bacillus subtilis</i> 168                 | Transport assay   | (Fouet et al., 1987)     |
|     | 4.A.1.2.12 | PtsS<br>(Cgl2642)     | Sucrose                                | <i>Corynebacterium glutamicum</i> ATCC 13032 | Gene deletion   | (Moon et al., 2005)      |
|     | 4.A.1.2.14 | PTS15<br>(LGAS_1669)  | Cellobiose                             | <i>Lactobacillus gasseri</i> ATCC 33323      | Transcriptional analysis<br>Gene deletion                     | (Francl et al., 2010)    |
|     | 4.A.1.2.16 | PTS20<br>(LGAS_1778)  | Sucrose                                | <i>Lactobacillus gasseri</i> ATCC 33323      | Transcriptional analysis<br>Gene deletion                     | (Francl et al., 2010)    |
|     | Literature | PTS1<br>(LPST_C0152)  | Fructoligosaccharides (DP2-5)          | <i>Lactobacillus plantarum</i> ST-III        | Transcriptional analysis<br>Gene deletion                     | (Chen et al., 2015)      |
|     | Literature | PTS26<br>(LPST_C2650) | Fructoligosaccharides (DP2-5)          | <i>Lactobacillus plantarum</i> ST-III        | Transcriptional analysis<br>Gene deletion                     | (Chen et al., 2015)      |
|     | Literature | scrT<br>(SP_1722)     | Sucrose                                | <i>Streptococcus pneumoniae</i> TIGR4        | Gene deletion   | (Iyer and Camilli, 2007) |
| Lac | 4.A.3.1.2  | PTS 6CB<br>(LGAS_343) | Lactose                                | <i>Lactobacillusgasseri</i> ATCC33323        | Transcriptional analysis<br>Gene deletion                     | (Francl et al., 2012)    |
|     | 4.A.3.1.3  | PTS 9BC<br>(LGAS_501) | Lactose                                | <i>Lactobacillusgasseri</i> ATCC33323        | Transcriptional analysis<br>Gene deletion                     | (Francl et al., 2012)    |
|     | 4.A.3.2.1  | ChbC<br>(b1737)       | N,N'-diacetylchitobiose, cellobiose    | <i>Escherichia coli</i> K-12                 | Gene deletion   | (Kachroo et al., 2007)   |
|     | 4.A.3.2.2  | LicC<br>(BSU38580)    | Lichenan oligosaccharides (DP unclear) | <i>Bacillus subtilis</i> 168                 | Gene deletion   | (Tobisch et al., 1997)   |

|     |            |                       |  |  |   |                              |
|-----|------------|-----------------------|--|--|---|------------------------------|
|     | 4.A.3.2.5  | ChbC                  | N,N'-diacetylchitobiose                        | <i>Serratia marcescens</i><br>2170               | Gene deletion                                 | (Uchiyama et al., 2003)      |
|     | 4.A.3.2.6  | CelB<br>(VC_1282)     | N,N'-diacetylchitobiose                        | <i>Vibrio cholerae</i> serotype<br>O1 ATCC 39315 | Transcriptional analysis<br>Gene deletion     | (Meiborn et al., 2004)       |
|     | 4.A.3.2.7  | CelC<br>(KP1_4524)    | Cellobiose                                     | <i>Klebsiella pneumoniae</i><br>NTUH-K2044       | Gene deletion                                 | (Wu et al., 2012)            |
|     | 4.A.3.2.8  | CelB<br>(BCE_5325)    | N,N'-diacetylchitobiose                        | <i>Bacillus cereus</i> ATCC<br>10987             | Transport assay<br>Gene deletion<br>Structure | (Cao et al., 2011)           |
|     | 4.A.3.2.9  | Lactose transporter   | Lactose  | <i>Klebsiella pneumoniae</i><br>CG43             | Transport assay                               | (Imai and Hall, 1981)        |
|     | Literature | GenB<br>(OG1RF_10235) | Gentiobiose,<br>amygdalin                      | <i>Enterococcus faecalis</i><br>OG1RF            | Gene deletion<br>Transcriptional analysis     | (Grand et al., 2019)         |
| Man | 4.A.6.1.13 | AlfF<br>(LCAB_28290)  | Fucosyl- $\alpha$ -1,3-N-<br>acetylglucosamine | <i>Lactobacillus casei</i><br>BL23               | Gene deletion                                 | (Rodríguez-Dáz et al., 2012) |
|     | 4.A.6.1.14 | PTS-IIC<br>(spr0294)  | Hyaluronic acid (DP<br>unclear)                | <i>Streptococcus pneumoniae</i> ATCC<br>BAA-255  | Gene deletion                                 | (Marion et al., 2012)        |
|     | 4.A.6.1.18 | GnbC<br>(LCABL_02930) | Galacto N-biose and<br>lacto N-biose           | <i>Lactobacillus casei</i><br>BL23               | Gene deletion<br>Transcriptional analysis     | (Bidart et al., 2014)        |

### 3.4.3. Taxonomical diversity

We examined the taxonomical diversity of all the characterized bacterial PTS transporters targeting glycosides, issued from the TCDB Glc, Lac and Man families and literature (**Table 5**). These 31 glycoside PTS transporters are spread both in Gram-positive and negative bacteria, including three phyla and eleven species. More than half of the characterized glycoside-targeting PTS transporters are from species of the Firmicutes phylum. The PTSs targeting glucosides are distributed in three phyla (Firmicutes, Proteobacteria and Actinobacteria), while PTS targeting mannosides and galactosides all belong to species of the Firmicutes phylum. The PTS content correlates with the potential of the species for facultative anaerobic and anaerobic growth (Jeckelmann and Erni, 2019), which correlates with the fact that the Firmicute phylum is one of the dominant phylum in the mammalian gut microbiota. Thirty percent of the glycoside PTS from Firmicutes belong to the *Lactobacillus* genus, of which the species can utilize diverse glycosides found in the human gut, allowing them to survive in this ecosystem. This trait, and the fact that many *Lactobacillus* species exert benefits to general health and wellness to the host, make of them interesting probiotics (Francl et al., 2010). Nevertheless, PTSs are also present in some pathogenic species, such as *Klebsiella pneumoniae* NTUH-K2044, which causes nosocomial infections and opportunistic infections (Wu et al., 2012), and *Vibrio cholerae* O1, which causes a dehydrating diarrheal illness (Meiborn et al., 2004). PTSs might thus play a role in the habitat colonization by these strains, and thus, in bacterial infection.

**Table 5. Phylum and order distribution of the characterized glycoside-targeting PTS transporters in bacteria.**

| Phylum         | Order           | Number of characterized PTSs | Percentage of the characterized PTSs |        |
|----------------|-----------------|------------------------------|--------------------------------------|--------|
|                |                 |                              | Order                                | Phylum |
| Firmicutes     | Bacillus        | 5                            | 16.1                                 | 61.3   |
|                | Streptococcus   | 5                            | 16.1                                 |        |
|                | Lactobacillus   | 8                            | 25.9                                 |        |
|                | Enterococcus    | 1                            | 3.2                                  |        |
| Proteobacteria | Escherichia     | 4                            | 12.9                                 | 32.2   |
|                | Klebsiella      | 4                            | 12.9                                 |        |
|                | Serratia        | 1                            | 3.2                                  |        |
|                | Vibrio          | 1                            | 3.2                                  |        |
| Actinobacteria | Corynebacterium | 2                            | 6.5                                  | 6.5    |

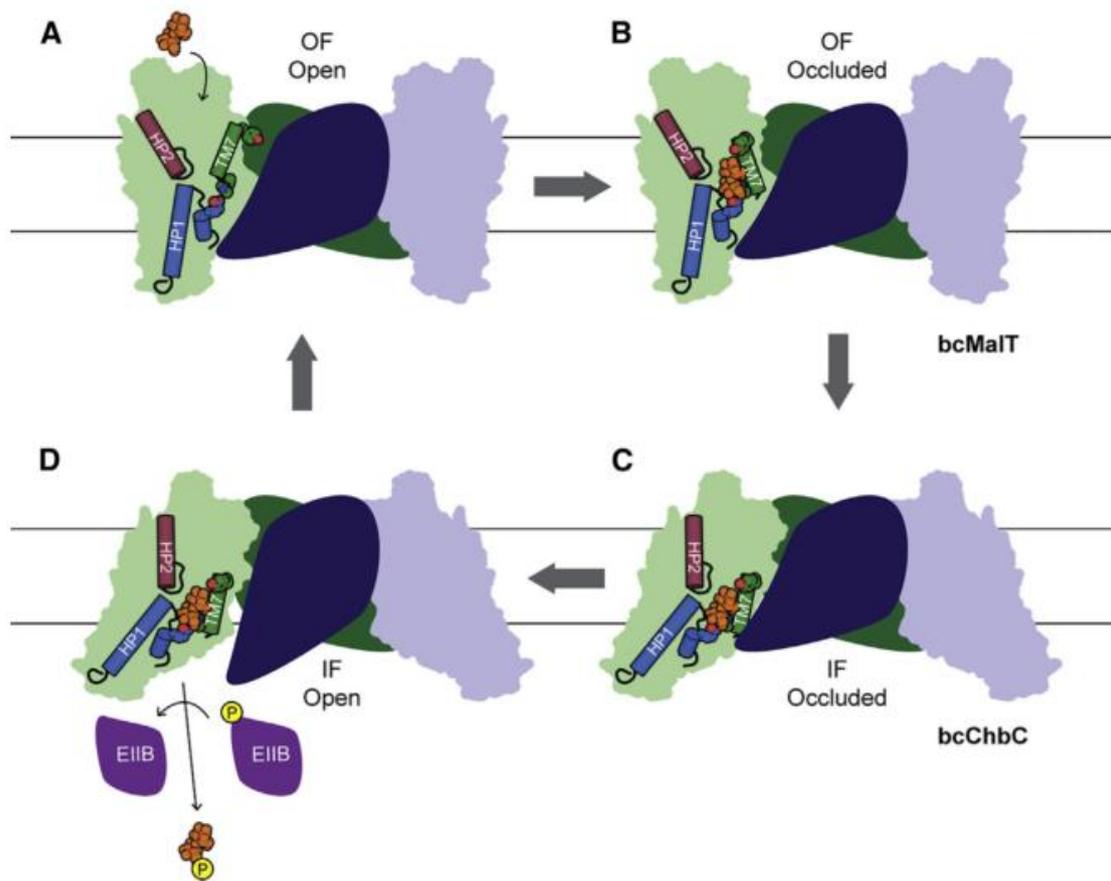
#### 3.4.4. Mechanism of glycoside import by PTS transporters

As described above, the transport of glycosides by PTS is accompanied by cascade of phosphorylations, in which the phosphate moiety is transferred from PEP to the imported glycoside by a set of cytoplasmic phosphocarrier proteins (EI, HPr, EIIA and EIIB, serially). The EI, HPr and IIA are all phosphorylated at a histidine residue. The IIB domains of the GFL superfamily are phosphorylated at a cysteine, while the IIB domains of the mannose family are phosphorylated at a histidine residue. The crystal structures of several EI, HPr, IIA and IIB proteins have been solved, alone or in complex with other PTS, which have been detailed in previous reviews (Deutscher et al., 2006; Erni, 2013; Peterkofsky et al., 2001; Robillard and Broos, 1999).

Little was known about how the transmembrane EIIC domain selectively imports the glycoside across the membrane before a high resolution crystal structure of diacetylchitobiose EIIC protein (bcChbC) from *Bacillus cereus* E33L was solved in the inward-facing conformation (Cao et al., 2011). A few years later, the maltose specific EIIC protein (bcMalT) from the same strain was solved in the inward-facing and outward-facing conformations (McCoy et al., 2016; Ren et al., 2018). Even if bcChbC and bcMalT have different specificities, they belong to the same GFL superfamily and both proteins are homodimers

with 10 transmembrane helices, constituting a scaffold domain (SD) containing the large and mostly hydrophobic protomer-protomer interface and a transport domain (TD) containing sugar binding site and translocation pathway. Thus a so-called elevator mechanism was proposed. Overall, the substrate firstly binds to the transporter (IIC) in the outward-facing open conformation, then the substrate is occluded and translocated by a rigid body motion of TD against the immobile SD across the membrane, where the EIIC isomerizes to the inward-facing open conformation (**Figure 4**) (Jeckelmann and Erni, 2019). The isomerization rate is increased 100–1000 folds by the activity of the phosphorylated IIB domain and rendered unidirectional by the concomitant phosphoryl transfer from phosphorylated-IIB domain to the substrate in the binding cavity and/or by the subsequent release of the phosphorylated substrate. However, it is unknown whether the two protomers isomerize independently or in synchrony, and how the IIB and IIB interact in the phosphate transformation process.

Members of the PTS in Man family contain a unique additional IID transmembrane domain. The only determined structure of EIICD complex is that of the *E. coli* K-12 protein specific for mannose (Liu et al., 2019), which also suggested a possible elevator mechanism. However, the feasibility of this mechanism for the glycoside PTS from the Man family still needs further investigation.

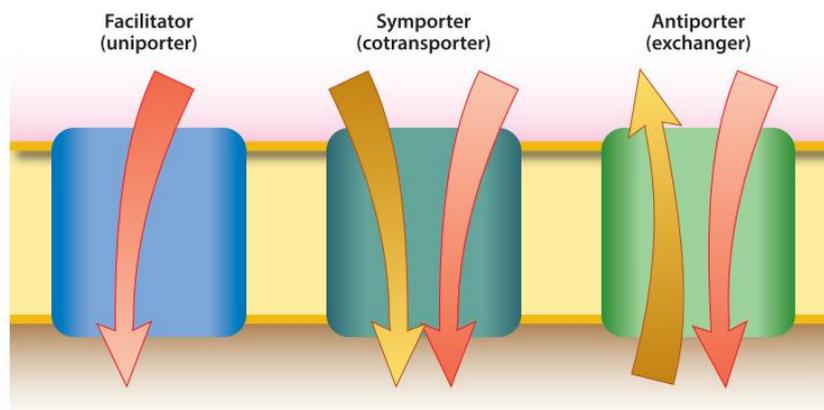


**Figure 4. Carbohydrate transport of through the transmembrane PTS\_EIIC domain.**

This figure shows a model of the glucose superfamily EIIC transport cycle. (A–D) The substrate binds to the outward-facing open structure (A), which is then converted to the outward-facing occluded state (B) by a movement of transport domain. The transport domain moves toward the cytoplasmic side of the membrane to form the inward-facing occluded state (C), and finally the substrate is phosphorylated by the phosphoryl from the phosphorylated-EIIB domain and is released to the cytoplasmic space (D). (from McCoy et al., 2016)

### 3.5. MFS transporters of glycosides

The major facilitator superfamily (MFS) is, with the ABC superfamily, one of the two exceptionally large superfamilies of transporters, including millions of sequenced members found ubiquitously in bacteria, archaea, and eukaryotes (Saier, 2006). The MFS transporters are single-polypeptide secondary carriers only capable of transporting small solutes in response to chemiosmotic ion gradients. MFS transporters represent the largest family of secondary membrane transporters. They can be classified into symporters, antiporters and, in case of a driving substance being absent, facilitators (**Figure 5**) (Yan, 2015). In the TCDB (release of February 2020), the MFS superfamily is subdivided into 104 families, consisting of 89 families classified under the main MFS family and annotated TC# 2.A.1, and 15 other families including distant members of MFSs. Each family contains members specific to a different set of related compounds, including monosaccharides, oligosaccharides, polyols, drugs, neurotransmitters, Krebs cycle metabolites, phosphorylated glycolytic intermediates, amino acids, peptides, organic anions, inorganic anions, etc (Reddy et al., 2012).



**Figure 5. Schematic representation of the three types of MFS transporters.**

This figure shows the difference of the transport process for facilitators, symporters and antiporters. Facilitators catalyze substrate diffusion across the membrane down its concentration gradient. Symporters or antiporters use the energy released from the downhill translocation of one substrate to drive the uphill translocation of another substrate either in the same direction (symporters) or in the opposite direction (antiporters). The color gradients of the arrows indicate the transmembrane electrochemical gradients of the substrates. (from Yan, 2015)

### 3.5.1. Diversity of sequences and structures

Among the 104 established MFS families in the TCDB, the bacterial MFS transporters of glycosides are spread in three of them: the Oligosaccharide: H<sup>+</sup> Symporter (OHS) Family (2.A.1.5), the Fucose: H<sup>+</sup> Symporter (FHS) family (2.A.1.7) and the Glycoside-Pentoside-Hexuronide (GPH): Cation Symporter family (2.A.2) (**Table 6**). The GPH family is one of the distant families of the MFS superfamily, since transporters of the GPH family show only marginal sequence similarities with other members of the other families. Position-Specific Iterated BLAST (PSI-BLAST) analysis suggested that these proteins all arose from a common ancestor (Saier, 2000). The glycoside MFS transporters are generally about 400-550 amino acids in length, except the Gram-positive bacterial lactose permeases (LacS) which are larger because of an additional EII-A like domain at the C-terminal extremity, which is usually found in PTS transporters. Despite their weak sequence similarities, all these proteins share the same topology, with 12 transmembrane helices of which the first six transmembrane helices exhibit significant sequence similarity with the last six (Saier, 2000).

### 3.5.2. Functional diversity

The bacterial MFS transporters which have been characterized to be responsible for the uptake of different glycosides are listed in **Table 6**. The transporters of the OHS family have been reported to be specific for a set of galactosides (such as lactose, raffinose and melibiose), sucrose and maltose. Among the members of the OHS family, the LacY permease involved in the uptake of lactose and other galactosides by *Escherichia coli* K-12 was the first studied member of MFS (Büchel et al., 1980) and nowadays remains one of the best characterized MFS transporters (Kaback and Guan, 2019).

Only one member of the FHS family, the ScrT transporter specific of sucrose uptake from *Shewanella frigidimarina* NCIMB 400, has been functionally characterized as specific for oligosaccharide transport (Rodionov et al., 2010), while the other characterized transporters of this family are involved in monosaccharide or its derivatives transport. By contrast, the members of the GPH family exhibit a broad specificity towards glycosides including galactosides, glucosides, xylosides and fructooligosaccharides.

Some transporters, including members of OHS family as LacY from *Escherichia coli* K-12 (King and Wilson, 1990), CscB from *Escherichia coli* EC3132 (Peng et al., 2009), MelY of *Enterobacter cloacae* ATCC 13047 (Shinnick et al., 2003; Tavoulari and Frillingos, 2008) and only member of GPH family: LacS of *Lactobacillus acidophilus* NCFM (Andersen et al.,

2011), are able to transport different glycosides sharing structural traits. For example, the LacY protein could internalize lactose, melibiose, thio- $\beta$ -methyl galactopyranoside (TMG), 4-nitrophenyl- $\beta$ -D-galactopyranoside and 4-nitrophenyl- $\alpha$ -D-galactopyranoside, because there is one sugar-binding site in LacY that selectively recognizes the galactosyl moiety (Kaback and Guan, 2019).

By listing the substrates of the characterized glycoside MFS transporters, it appeared that they are mainly specific for disaccharides or their derivatives. However, MFS transporters can also import longer glycosides. Using gene deletion and growth dynamic analysis, two transporters were shown to be able to internalize oligosaccharides up to DP 4: the MFS from an uncultured *Bacteroides* strain (Tauzin et al., 2016b) and the FosT from *Escherichia coli* BEN2908 (Schouler et al., 2009), which internalize xylo-tetraose and fructo-tetraose, respectively. Other MFs, such as TogT (Hugouvieux-Cotte-Pattat and Reverchon, 2001) from *Erwinia chrysanthemi* 3937 and LacS (Lba1463) (Andersen et al., 2011) from *Lactobacillus acidophilus* NCFM, have been shown to be responsible for growth on a mixture of oligosaccharides, but the size of the oligosaccharide they were able to transport has not been characterized. It is thus possible that MFS could transport glycosides with DP higher than 4, although further experiments are required to validate this hypothesis.

**Table 6. List of the bacterial MFS transporters involved in glycoside utilization, characterized either by binding or transport assays, gene deletion and/or transcriptional approaches.**

The glycoside MFS transporters retrieved from the literature but which are not listed in the TCDB were classified based on the identified conserved domain using a BLAST analysis against the NCBI's Conserved Domain Database (Marchler-Bauer et al., 2017).

| TCDB Family name | TCID       | Protein name (locus tag) | Substrate  | Organism                               | Characterization method   | Reference  |
|------------------|------------|--------------------------|--|--|---|--|
| OHS              | 2.A.1.5.1  | LacY (b0343)             | Lactose, melibiose, thio- $\beta$ -methyl galactopyranoside (TMG), 4-nitrophenyl- $\beta$ -D-galactopyranoside, 4-nitrophenyl- $\alpha$ -D-galactopyranoside | <i>Escherichia coli</i> K-12           | Transport assay<br>Point mutation<br>Structure                              | (Abramson et al., 2003; Kaback, 2015; King and Wilson, 1990) |
|                  | 2.A.1.5.2  | RafB (RAFB_ECOL)         | Raffinose  | <i>Escherichia coli</i> K-12           | Transport assay<br>Point mutation   | (Camp et al., 2007)  |
|                  | 2.A.1.5.3  | CscB                     | Sucrose, maltose   | <i>Escherichia coli</i> EC3132         | Transport assay<br>Heterologous expression in another <i>E. coli</i> strain | (Peng et al., 2009)  |
|                  | 2.A.1.5.4  | MeIY (ECL_05077)         | Melibiose, lactose   | <i>Enterobacter cloacae</i> ATCC 13047 | Gene deletion<br>Point mutation<br>Transport assay                          | (Shinnick et al., 2003; Tavoulari and Frillingos, 2008)      |
|                  | literature | FosT (Aec46)             | Fructooligosaccharides (DP3-4)   | <i>Escherichia coli</i> BEN2908        | Gene deletion   | (Schouler et al., 2009)                                      |

|     |            |                              |  |  |  |  |
|-----|------------|------------------------------|--|--|--|--|
| FHS | 2.A.1.7.6  | ScrT<br>(Sfri_3989)          | Sucrose  | <i>Shewanella<br/>frigidimarina</i><br>NCIMB 400 | Heterologous<br>expression in <i>E. coli</i>                   | (Rodionov et al.,<br>2010)                           |
| GPH | 2.A.2.1.1  | MelB<br>(EcE24377A_<br>4674) | Melibiose  | <i>Escherichia coli</i><br>E24377A               | Transport assay<br>Binding test<br>Structure                   | (Ethayathulla et al.,<br>2014)                       |
|     | 2.A.2.2.1  | LacS<br>(DID82_<br>03005)    | Lactose  | <i>Streptococcus<br/>thermophilus</i> A147       | Heterologous<br>expressed in <i>E. coli</i><br>Transport assay | (Poolman et al.,<br>1989)                            |
|     | 2.A.2.3.1  | GusB<br>(FAM13_RS0<br>0460)  | Glucuronide  | <i>Escherichia coli</i><br>K-12                  | Transcriptional analysis<br>Transport assay                    | (Liang et al., 2005)                                 |
|     | 2.A.2.3.3  | XylP<br>(XYLP_<br>LACP)      | Isoprimeverose                                       | <i>Lactobacillus<br/>pentosus</i> NZ9000         | Transport assay  | (Heuberger et al.,<br>2001)                          |
|     | 2.A.2.5.1  | TogT                         | Oligogalacturonides<br>(DP>2)                        | <i>Erwinia<br/>chrysanthemii</i> 3937            | Gene deletion  | (Hugouvieux-Cotte-<br>Pattat and<br>Reverchon, 2001) |
|     | 2. A.2.6.2 | MalY<br>(CC_2283)            | Maltose  | <i>Caulobacter<br/>crescentus</i> ATCC<br>19089  | Gene deletion<br>Transport assay                               | (Lohmiller et al.,<br>2008)                          |
|     | literature | F5_MFS<br>(CCG34981.1)       | Xylooligosaccharides<br>(DP2-4)                      | Uncultured<br><i>Bacteroides</i>                 | Gene deletion  | (Tauzin et al.,<br>2016b)                            |
|     | literature | LacS<br>(Lba1463)            | Galactooligosaccharides<br>(DP2-6), lactose, lactiol | <i>Lactobacillus<br/>acidophilus</i> NCFM        | Transcriptional analysis<br>Gene deletion                      | (Andersen et al.,<br>2011)                           |

### 3.5.3. Taxonomical diversity

The 14 characterized bacterial MFS transporters of glycosides are spread in three phyla (Proteobacteria, Firmicutes and Bacteroidetes) (Table 7). Most of them (71%) belong to Proteobacteria, of which 57% are from the order Enterobacterales, including the well-known LacY from *Escherichia coli* (Table 7). The identification of LacY might have facilitated the identification of other Enterobacterales MFS, due to the fact that gene transfer or duplication more frequently occur intra-phyla than inter-phyla (Soucy et al., 2015). However, even if most of the MFS retrieved for now belong to this order, these results could be biased as the analyses are based on the well-known LacY sequences from Proteobacteria but we cannot deny that the vast majority of sequences remain unknown so far.

The glycoside MFS transporters are present in diverse bacterial phyla from different habitats, including the human gut, food and soil microbiota, suggesting the important role of MFSs in bacteria survival and adaptation to their ecological niches. The phytopathogen *Erwinia chrysanthemi* 3937 (Hugouvieux-Cotte-Pattat and Reverchon, 2001), uses a MFS to transport pectin oligomers, a component of plant cell wall, thus causing soft rot of plants. The fructooligosaccharide MFS transporter of the pathogenic *Escherichia coli* BEN2908 strain is important for colonization of the chicken intestine, indicating a potential risk related to high level of fructooligosaccharide prebiotic usage in food and feed (Schouler et al., 2009). Nevertheless, glycoside MFSs are also found in strains that are beneficial to human health, such as in the probiotic strains *Streptococcus thermophilus* A147 (Poolman et al., 1989), *Lactobacillus acidophilus* NCFM (Andersen et al., 2011) and *Lactobacillus pentosus* NZ9000 (Heuberger et al., 2001).

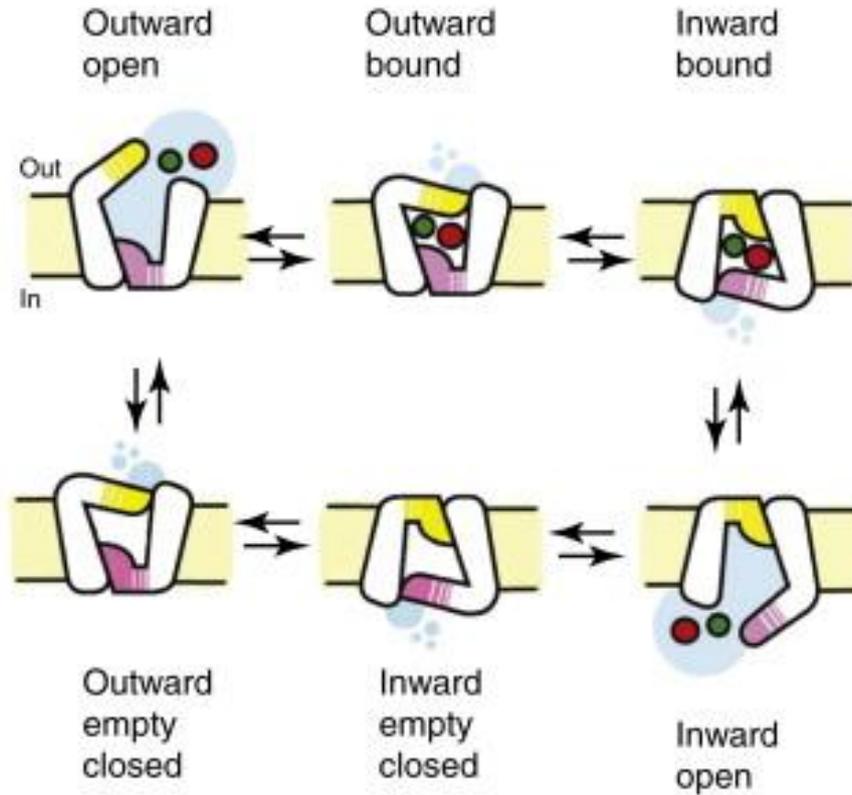
**Table 7. Phylum and order distribution of characterized glycoside-targeting MFS transporters in bacteria.**

| Phylum         | Order            | Number of characterized MFS | Percentage of the characterized MFSs |        |
|----------------|------------------|-----------------------------|--------------------------------------|--------|
|                |                  |                             | Order                                | Phylum |
| Proteobacteria | Enterobacterales | 8                           | 57.2                                 |        |
|                | Alteromonadales  | 1                           | 7.1                                  | 71.4   |
|                | Caulobacterales  | 1                           | 7.1                                  |        |
| Firmicutes     | Lactobacillales  | 3                           | 21.5                                 | 21.5   |
| Bacteroidetes  | Bacteroidales    | 1                           | 7.1                                  | 7.1    |

#### 3.5.4. Mechanism of glycoside import by MFS transporters

Before the resolution of any crystal structure, mutation of some MFS transporters, such as the LacY (King and Wilson, 1990), RafB (Camp et al., 2007) and MelB (Granell et al., 2010) proteins, allowed the identification of amino acid residues that play an important role in the recognition and transport of substrates across the membrane.

The first glycoside MFS structure solved was that of the lactose permease (LacY) from *E. coli* K-12, in an inward-facing conformation with a bound lactose homolog,  $\beta$ -D-galactopyranosyl-1-thio- $\beta$ -D-galactopyranoside (Abramson et al., 2003), and in an inward-opening conformation (Guan et al., 2007). More recently, the structure of a second glycoside MFS transporter, the melibiose specific MelB protein from *Salmonella typhimurium* LT2, has been solved in an outward partially occluded and an outward inactive state (Ethayathulla et al., 2014). These structures allowed the description of the general features of the MFS transporters (**Figure 6**). A canonical MFS fold comprises two domains, the N and C domains, with each consists of six consecutive transmembrane helices. A single substrate binding cavity is located in the center of the membrane and enclosed by the N and C domains. Moreover, an alternating-access mechanism has been proposed to explain the translocation process (Kaback and Guan, 2019; Zhang et al., 2015). Carbohydrate binding is the first step and is the driving force for the alternating access mechanism. The transporters adopt two main open conformations for substrate loading and release in the outward- and inward-facing open conformations, respectively, and alternate between these two conformations through occluded states to translocate substrates across the membrane.



**Figure 6. Alternating access mechanism for an ion/carbohydrate symport.**

This figure shows the transport cycle. It starts with an empty outward-facing transporter which binds the substrate and coupled ion (red and green spheres, respectively) from the extracellular solution in a reaction coupled to the closure of the gate (yellow). The transporter undergoes a conformational transition into an inward-facing state in which the extracellular gate can no longer open while opening of the intracellular gate (pink) is enabled. The release of the substrate and ion into the cytoplasm yields an empty inward-facing state, and isomerization back into the outward-facing state completes the cycle. (from Boudker and Verdon, 2010)

### 3.6. Conclusion and thesis objectives

In this review chapter, we highlighted the functional, taxonomic, structural and mechanistic diversity of glycoside transporters, which have been developed by bacteria to meet the huge structural diversity of glycosides widespread in Nature.

The constant improvements of (meta)genome sequencing technologies, genetic tools and omics technology have not only enabled the prediction of a large number of glycoside transporters, but also allowed the identification and characterization of some of them. However, these characterization works are far from having covered the functional diversity of these proteins. Indeed, the involvement of less than 200 transporters in the utilization of glycosides has been experimentally validated, either by genetic, biochemical, or, more rarely, by structural approaches. In the vast majority of cases, the biochemical characterization of the transporter specificity is performed by targeting the substrate binding proteins, which are much easier to study than the transmembrane proteins through which transit the oligosaccharides between the extracellular medium and the cytoplasm or periplasm. However, the binding specificity might not be identical to the specificity of uptake, whose characterization is based on the use of labeled substrates. Furthermore, among all the characterized transporters, only one is issued from an uncultured bacterium, while uncultured microorganisms dominate in all natural ecosystems.

The development of methods to accelerate the identification of bacterial transporters, and to refine the characterization of their uptake specificity, by discriminating it from their binding one, is therefore a major challenge. The use of recombinant hosts, for which genetic tools are available to modify their metabolism and make them co-produce several recombinant proteins (the different proteins constituting certain transporters, but also the CAZymes encoded in the same loci), appears to be very promising to simplify these systems, avoiding the functional redundancy of native strains. The use of recombinant hosts is also mandatory to study transporters from uncultured bacteria. Recently, it has been shown in the Discomics group that *Escherichia coli* can be an interesting tool to express the genes of PULs cloned in fosmids, and to produce transporters from uncultured *Bacteroides*. Nevertheless, the efficiency, advantages and limits of this system for the identification and characterization of transporters from both Gram-negative and positive bacteria remains to be evaluated. Besides, only a few dozen transporters involved in the utilization of dietary fibers by mammal gut bacteria have been characterized, but without profiling their transport specificity towards

oligosaccharides of defined structures. This lack of functional data limits the understanding of the structure-function relationships of transporters, in general. It is also a major barrier to the prediction of their precise substrate targets, and thus, to the understanding of how bacteria colonize their habitat.

It is to face these challenges that I started this PhD work, nearly three years ago. The main objectives were the following:

- to evaluate the interest of *E. coli* as a recombinant host for the screening of the specificity of oligosaccharide transporters from uncultured gut bacteria
- to develop methods for the detection and quantification of the efficiency of binding and uptake of unlabeled oligosaccharides of defined structures
- to study, at the molecular scale, mechanisms of dietary fiber utilization by prominent human gut bacteria
- to advance in the understanding of the relationships between the sequence, structure and specificity of glycoside transporters, and to revisit their classification.

The scientific questions addressed during this work are presented at the beginning of each of the four 'Results' chapters.

# **Bibliography**



- Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R., Iwata, S., 2003. Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301, 610–615. <https://doi.org/10.1126/science.1088196>
- Alteri, C.J., Smith, S.N., Mobley, H.L.T., 2009. Fitness of *Escherichia coli* during urinary tract infection requires gluconeogenesis and the TCA cycle. *PLoS Pathog.* 5. <https://doi.org/10.1371/journal.ppat.1000448>
- Andersen, J.M., Barrangou, R., Hachem, M.A., Lahtinen, S., Goh, Y.J., Svensson, B., Klaenhammer, T.R., 2011. Transcriptional and functional analysis of galactooligosaccharide uptake by *lacS* in *Lactobacillus acidophilus*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 17785–17790. <https://doi.org/10.1073/pnas.1114152108>
- Andersen, J.M., Barrangou, R., Hachem, M.A., Lahtinen, S.J., Goh, Y.J., Svensson, B., Klaenhammer, T.R., 2012. Transcriptional Analysis of Prebiotic Uptake and Catabolism by *Lactobacillus acidophilus* NCFM. *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0044409>
- Arroyo, J., Farkaš, V., Sanz, A.B., Cabib, E., 2016. Strengthening the fungal cell wall through chitin–glucan cross-links: effects on morphogenesis and cell integrity. *Cell. Microbiol.* 18, 1239–1250. <https://doi.org/10.1111/cmi.12615>
- Ayechu-Muruzabal, V., van Stigt, A.H., Mank, M., Willemsen, L.E.M., Stahl, B., Garssen, J., van't Land, B., 2018. Diversity of human milk oligosaccharides and effects on early life immune development. *Front. Pediatr.* 6, 1–9. <https://doi.org/10.3389/fped.2018.00239>
- BAGENHOLM, V., Reddy, S.K., Bouraoui, H., Morrill, J., 2017. Galactomannan catabolism conferred by a polysaccharide utilisation locus of *Bacteroides ovatus* enzyme synergy and crystal structure of a  $\beta$ -mannanase. *J. Biol. Chem.* 292, 229–243.
- Barbi, F., Prudent, E., Vallon, L., Buée, M., Dubost, A., Legout, A., Marmeisse, R., Fraissinet-Tachet, L., Luis, P., 2016. Tree species select diverse soil fungal communities expressing different sets of lignocellulolytic enzyme-encoding genes. *Soil Biol. Biochem.* 100, 149–159. <https://doi.org/10.1016/j.soilbio.2016.06.008>
- Barrangou, R., Altermann, E., Hutkins, R., Cano, R., Klaenhammer, T.R., 2003. Functional and comparative genomic analyses of an operon involved in fructooligosaccharide

- utilization by *Lactobacillus acidophilus*. Proc. Natl. Acad. Sci. U. S. A 100, 8957–8962.  
<https://doi.org/10.1073/pnas.1332765100>
- Beier, S., Bertilsson, S., 2013. Bacterial chitin degradation-mechanisms and ecophysiological strategies. Front. Microbiol. 4, 1–12. <https://doi.org/10.3389/fmicb.2013.00149>
- Beis, K., 2015. Structural basis for the mechanism of ABC transporters. Biochem. Soc. Trans. 43, 889–893. <https://doi.org/10.1042/bst20150047>
- Bell, A., Brunt, J., Crost, E., Vaux, L., Nepravishta, R., Owen, C.D., Latousakis, D., Xiao, A., Li, W., Chen, X., Walsh, M.A., Claesen, J., Angulo, J., Thomas, G.H., Juge, N., 2019. Elucidation of a sialic acid metabolism pathway in mucus-foraging *Ruminococcus gnavus* unravels mechanisms of bacterial adaptation to the gut. Nat. Microbiol. 4, 2393–2404. <https://doi.org/10.1038/s41564-019-0590-7>
- Berntsson, R.P.A., Smits, S.H.J., Schmitt, L., Slotboom, D.J., Poolman, B., 2010. A structural classification of substrate-binding proteins. FEBS Lett. 584, 2606–2617.  
<https://doi.org/10.1016/j.febslet.2010.04.043>
- Bidart, G.N., Rodríguez-Díaz, J., Monedero, V., Yebra, M.J., 2014. A unique gene cluster for the utilization of the mucosal and human milk-associated glycans galacto-N-biose and lacto-N-biose in *Lactobacillus casei*. Mol. Microbiol. 93, 521–538.  
<https://doi.org/10.1111/mmi.12678>
- Bjelić, S., Jelesarov, I., 2008. A survey of the year 2007 literature on applications of isothermal titration calorimetry. J. Mol. Recognit. 21, 289–311.  
<https://doi.org/10.1002/jmr.909>
- Bjursell, M.K., Martens, E.C., Gordon, J.I., 2006. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. J. Biol. Chem. 281, 36269–36279.  
<https://doi.org/10.1074/jbc.M606509200>
- Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé N., Vasse, J., Lauber, E., Arlat, M., 2007. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS One 2, e224.  
<https://doi.org/10.1371/journal.pone.0000224>

- Boehm, G., Stahl, B., 2007. Oligosaccharides from Milk. *J. Nutr.* 137, 847S-849S.  
<https://doi.org/10.1093/jn/137.3.847s>
- Bolam, D.N., van den Berg, B., 2018. TonB-dependent transport by the gut microbiota: novel aspects of an old problem. *Curr. Opin. Struct. Biol.* 51, 35–43.  
<https://doi.org/10.1016/j.sbi.2018.03.001>
- Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J., 2004. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769–781.  
<https://doi.org/10.1042/BJ20040892>
- Boucher, N., Noll, K.M., 2011. Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of *Thermotoga maritima* MSB8 screened by differential scanning fluorimetry. *Appl. Environ. Microbiol.* 77, 6395–6399.  
<https://doi.org/10.1128/AEM.05418-11>
- Boudker, O., Verdon, G., 2010. Structural perspectives on secondary active transporters. *Trends Pharmacol. Sci.* 31, 418–426. <https://doi.org/10.1016/j.tips.2010.06.004>
- Boulanger, A., Déjean, G., Lautier, M., Glories, M., Zischek, C., Arlat, M., Lauber, E., 2010. Identification and regulation of the N-acetylglucosamine utilization pathway of the plant pathogenic bacterium *Xanthomonas campestris* pv. *campestris*. *J. Bacteriol.* 192, 1487–1497. <https://doi.org/10.1128/JB.01418-09>
- Brockhausen, I., Stanley, P., 2017. O-GalNAc Glycans, in: *Essentials of Glycobiology* [Internet]. 3rd Edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Büchel, D.E., Gronenborn, B., Müller-Hill, B., 1980. Sequence of the lactose permease gene. *Nature* 283, 541–545. <https://doi.org/10.1038/283541a0>
- Cameron, E.A., Kwiatkowski, K.J., Lee, B., Hamaker, B.R., Koropatkin, N.M., Martens, C., 2014. Multifunctional nutrient-binding proteins adapt human symbiotic enhanced sensing and catalysis. *MBio* 5, e01441-14. <https://doi.org/10.1128/mBio.01441-14>.
- Camp, B.M. Van, Crow, R.R., Peng, Y., Varela, M.F., 2007. Amino acids that confer transport of raffinose and maltose sugars in the raffinose permease (RafB) of *Escherichia*

- coli as implicated by spontaneous mutations at Val-35, Ser-138, Ser-139, Gly-389 and Ile-391. *Bone* 23, 1–7. <https://doi.org/10.1038/jid.2014.371>
- Cao, Y., Jin, X., Levin, E.J., Huang, H., Zong, Y., Quick, M., Javitch, J.A., Rajashankar, K.R., 2011. Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature* 473, 50–54. <https://doi.org/10.1038/nature09939>
- Chandravanshi, M., Sharma, A., Gupta, P. Das, Mandal, S.K., Kanaujia, S.P., 2019. Identification and characterization of ABC transporters for carbohydrate uptake in *Thermus thermophilus* HB8. *Gene* 696, 135–148. <https://doi.org/10.1016/j.gene.2019.02.035>
- Chekan, J.R., Kwon, I.H., Agarwal, V., Dodd, D., Revindran, V., Mackie, R.I., Cann, I., Nair, S.K., 2014. Structural and biochemical basis for mannan utilization by *Caldanaerobius polysaccharolyticus* strain ATCC BAA-17. *J. Biol. Chem.* 289, 34965–34977. <https://doi.org/10.1074/jbc.M114.579904>
- Chen, C., Zhao, G., Chen, W., Guo, B., 2015. Metabolism of fructooligosaccharides in *Lactobacillus plantarum* ST-III via differential gene transcription and alteration of cell membrane fluidity. *Appl. Environ. Microbiol.* 81, 7697–7707. <https://doi.org/10.1128/AEM.02426-15>
- Chen, J.S., Reddy, V., Chen, J.H., Shlykov, M.A., Zheng, W.H., Cho, J., Yen, M.R., Saier, M.H., 2012. Phylogenetic characterization of transport protein superfamilies: Superiority of SuperfamilyTree programs over those based on multiple alignments. *J. Mol. Microbiol. Biotechnol.* 21, 83–96. <https://doi.org/10.1159/000334611>
- Cho, K.H., Salyers, A.A., 2001. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* 183, 7224–7230. <https://doi.org/10.1128/JB.183.24.7224-7230.2001>
- Cockburn, D.W., Koropatkin, N.M., 2016. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *J. Mol. Biol.* 428, 3230–3252. <https://doi.org/10.1016/j.jmb.2016.06.021>
- Cockburn, D.W., Orlovsky, N.I., Foley, M.H., Kwiatkowski, K.J., Bahr, C.M., Maynard, M., Demeler, B., Koropatkin, N.M., 2015. Molecular details of a starch utilization pathway

- in the human gut symbiont *Eubacterium rectale*. *Mol. Microbiol.* 95, 209–230.  
<https://doi.org/10.1111/mmi.12859>
- Colombo, P.M., Rascio, N., Cinelli, F., 1983. *Posidonia oceanica* (L.) Delile: a structural study of the photosynthetic apparatus. *Mar. Ecol.* 4, 133–145.  
<https://doi.org/10.1111/j.1439-0485.1983.tb00292.x>
- Connors, S.B., Montero, C.I., Comfort, D.A., Shockley, K.R., Johnson, M.R., Chhabra, S.R., Kelly, R.M., 2005. Prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima* through the use of carbohydrate-specific transcriptional response. *Abstr. Pap. Am. Chem. Soc.* 229, U240–U240.  
<https://doi.org/10.1128/JB.187.21.7267>
- Cote, C.K., Cvitkovitch, D., Bleiweis, A.S., Honeyman, A.L., 2000. A novel  $\beta$ -glucoside-specific PTS locus from *Streptococcus mutans* that is not inhibited by glucose. *Microbiology* 146, 1555–1563. <https://doi.org/10.1099/00221287-146-7-1555>
- Coutinho, P.M., Deleury, E., Davies, G.J., Henrissat, B., 2003. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* 328, 307–317.  
[https://doi.org/10.1016/S0022-2836\(03\)00307-3](https://doi.org/10.1016/S0022-2836(03)00307-3)
- Culurgioni, S., Harris, G., Singh, A.K., King, S.J., Walsh, M.A., 2017. Structural basis for regulation and specificity of fructooligosaccharide import in *Streptococcus pneumoniae*. *Structure* 25, 79–93. <https://doi.org/10.1016/j.str.2016.11.008>
- Cummings, J.H., Stephen, A.M., 2007. Carbohydrate terminology and classification. *Eur. J. Clin. Nutr.* 61, S5–S18. <https://doi.org/10.1038/sj.ejcn.1602936>
- Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E.A., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., Rogowski, A., Hamilton, B.S., Chen, R., Tolbert, T.J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munoz-Munoz, J.L., Day, A., Peña, M.J., McLean, R., Suits, M.D., Boraston, A.B., Atherly, T., Ziemer, C.J., Williams, S.J., Davies, G.J., Abbott, W.D., Martens, E.C., Gilbert, H.J., 2015. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* 517, 165–169. <https://doi.org/10.1038/nature13995>

- D'Elia, J.N., Salyers, A.A., 1996. Contribution of a neopullulanase, a pullulanase, and an  $\alpha$ -glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. *J. Bacteriol.* 178, 7173–7179. <https://doi.org/10.1128/jb.178.24.7173-7179.1996>
- Dagkesamanskaya, A., Langer, K., Tauzin, A.S., Rouzeau, C., Lestrade, D., Potocki-Veronese, G., Boitard, L., Bibette, J., Baudry, J., Pompon, D., Anton-Leberre, V., 2018. Use of photoswitchable fluorescent proteins for droplet-based microfluidic screening. *J. Microbiol. Methods* 147, 59–65. <https://doi.org/10.1016/j.mimet.2018.03.001>
- Davies, G., Henrissat, B., 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859. [https://doi.org/10.1016/S0969-2126\(01\)00220-9](https://doi.org/10.1016/S0969-2126(01)00220-9)
- De Costa, D.M., Suzuki, K., Yoshida, K., 2003. Structural and functional analysis of a putative gene cluster for palatinose transport on the linear chromosome of *Agrobacterium tumefaciens* MAFF301001. *J. Bacteriol.* 185, 2369–2373. <https://doi.org/10.1128/JB.185.7.2369-2373.2003>
- Déjean, G., Blanvillain-Baufumé S., Boulanger, A., Darrasse, A., De Bernonville, T.D., Girard, A.L., Carrère, S., Jamet, S., Zischek, C., Lautier, M., Solé M., Büttner, D., Jacques, M.A., Lauber, E., Arlat, M., 2013. The xylan utilization system of the plant pathogen *Xanthomonas campestris* pv *campestris* controls epiphytic life and reveals common features with oligotrophic bacteria and animal gut symbionts. *New Phytol.* 198, 899–915. <https://doi.org/10.1111/nph.12187>
- Déjean, G., Tamura, K., Cabrera, A., Jain, N., Pudlo, N.A., Pereira, G., Viborg, A.H., Van Petegem, F., Martens, E.C., Brumer, H., 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, 1–21. <https://doi.org/10.1128/mBio.00095-20>
- Déjean, G., Tauzin, A.S., Bennett, S.W., Creagh, A.L., Brumer, H., 2019. Adaptation of syntenic xyloglucan utilization loci of human gut *Bacteroidetes* to polysaccharide side chain diversity, *Applied and Environmental Microbiology*. <https://doi.org/10.1128/AEM.01491-19>

- Deng, B., Sullivan, M.A., Chen, C., Li, J., Powell, P.O., Hu, Z., Gilbert, R.G., 2016. Molecular structure of human-liver glycogen. *PLoS One* 11, 1–10. <https://doi.org/10.1371/journal.pone.0150540>
- Desai, S.K., Nandimath, K., Mahadevan, S., 2010. Diverse pathways for salicin utilization in *Shigella sonnei* and *Escherichia coli* carrying an impaired *bgl* operon. *Arch. Microbiol.* 192, 821–833. <https://doi.org/10.1007/s00203-010-0610-8>
- Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Chambon, C., Yeoman, C.J., Berg Miller, M.E., Fields, C.J., Martens, E., Terrapon, N., Henrissat, B., White, B.A., Mosoni, P., 2016. Xylan degradation by the human gut *Bacteroides xylanisolvens* XB1AT involves two distinct gene clusters that are linked at the transcriptional level. *BMC Genomics* 17, 1–14. <https://doi.org/10.1186/s12864-016-2680-8>
- Deutscher, J., Ak  F.M.D., Derkaoui, M., Z r  A.C., Cao, T.N., Bouraoui, H., Kentache, T., Mokhtari, A., Milohanic, E., Joyet, P., 2014. The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. *Microbiol. Mol. Biol. Rev.* 78, 231–256. <https://doi.org/10.1128/MMBR.00001-14>
- Deutscher, J., Francke, C., Postma, P.W., 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* 70, 939–1031. <https://doi.org/10.1128/MMBR.00024-06>
- Diallinas, G., 2014. Understanding transporter specificity and the discrete appearance of channel-like gating domains in transporters. *Front. Pharmacol.* 5 AUG, 1–17. <https://doi.org/10.3389/fphar.2014.00207>
- Ejby, M., Fredslund, F., Andersen, J.M., agar, A.V., Henriksen, J.R., Andersen, T.L., Svensson, B., Slotboom, D.J., Hachem, M.A., 2016. An atp binding cassette transporter mediates the uptake of  $\alpha$ -(1,6)-linked dietary oligosaccharides in bifidobacterium and correlates with competitive growth on these substrates. *J. Biol. Chem.* 291, 20220–20231. <https://doi.org/10.1074/jbc.M116.746529>
- Ejby, M., Fredslund, F., Vujcic-Zagar, A., Svensson, B., Slotboom, D.J., Abou Hachem, M., 2013. Structural basis for arabinoxylo-oligosaccharide capture by the probiotic

- Bifidobacterium animalis* subsp. *lactis* BI-04. *Mol. Microbiol.* 90, 1100–1112.  
<https://doi.org/10.1111/mmi.12419>
- Ejby, M., Guskov, A., Pichler, M.J., Zanten, G.C., Schoof, E., Saburi, W., Slotboom, D.J., Abou Hachem, M., 2019. Two binding proteins of the ABC transporter that confers growth of *Bifidobacterium animalis* subsp. *lactis* ATCC27673 on  $\beta$ -mannan possess distinct manno-oligosaccharide-binding profiles. *Mol. Microbiol.* 112, 114–130.  
<https://doi.org/10.1111/mmi.14257>
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. <https://doi.org/10.1093/nar/gky995>
- Erkens, G.B., Majsnerowska, M., ter Beek, J., Slotboom, D.J., 2012. Energy coupling factor-type ABC transporters for vitamin uptake in prokaryotes. *Biochemistry* 51, 4390–4396.  
<https://doi.org/10.1021/bi300504v>
- Erni, B., 2013. The bacterial phosphoenolpyruvate: Sugar phosphotransferase system (PTS): An interface between energy and signal transduction. *J. Iran. Chem. Soc.* 10, 593–630.  
<https://doi.org/10.1007/s13738-012-0185-1>
- Ethayathulla, A.S., Yousef, M.S., Amin, A., Leblanc, G., Kaback, H.R., Guan, L., 2014. Structure-based mechanism for Na(+)/melibiose symport by MelB. *Nat. Commun.* 5, 3009. <https://doi.org/10.1038/ncomms4009>
- Ferreira, M.J., De Sá Nogueira, I., 2010. A multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis*. *J. Bacteriol.* 192, 5312–5318.  
<https://doi.org/10.1128/JB.00832-10>
- Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., White, B.A., 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* 6, 121–131. <https://doi.org/10.1038/nrmicro1817>
- Foley, M.H., Cockburn, D.W., Koropatkin, N.M., 2016. The *Sus* operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell. Mol. Life Sci.* 73, 2603–2617.  
<https://doi.org/10.1007/s00018-016-2242-x>

- Ford, R.C., Beis, K., 2019. Learning the ABCs one at a time: structure and mechanism of ABC transporters. *Biochem. Soc. Trans.* 47, 23–36. <https://doi.org/10.1042/bst20180147>
- Fosses, A., Borne, R., Franche, N., de Philip, P., Fierobe, H.-P., Mat é M., Perret, S., Liu, N., Denis, Y., 2017. A seven-gene cluster in *Ruminiclostridium cellulolyticum* is essential for signalization, uptake and catabolism of the degradation products of cellulose hydrolysis. *Biotechnol. Biofuels* 10, 1–14. <https://doi.org/10.1186/s13068-017-0933-7>
- Fouet, A., Arnaud, M., Klier, A., Rapoport, G., 1987. *Bacillus subtilis* sucrose-specific enzyme II of the phosphotransferase system: expression in *Escherichia coli* and homology to enzymes II from enteric bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 84, 8773–8777. <https://doi.org/10.1073/pnas.84.24.8773>
- Francel, A.L., Hoeflinger, J.L., Miller, M.J., 2012. Identification of lactose phosphotransferase systems in *Lactobacillus gasseri* ATCC 33323 required for lactose utilization. *Microbiology* 158, 944–952. <https://doi.org/10.1099/mic.0.052928-0>
- Francel, A.L., Thongaram, T., Miller, M.J., 2010. The PTS transporters of *Lactobacillus gasseri* ATCC 33323. *BMC Microbiol.* 10. <https://doi.org/10.1186/1471-2180-10-77>
- Gelfand, M.S., Rodionov, D.A., 2008. Comparative genomics and functional annotation of bacterial transporters. *Phys. Life Rev.* 5, 22–49. <https://doi.org/10.1016/j.plrev.2007.10.003>
- Glenwright, A.J., Pothula, K.R., Bhamidimarri, S.P., Chorev, D.S., Basl é A., Firbank, S.J., Zheng, H., Robinson, C. V., Winterhalter, M., Kleinekath öfer, U., Bolam, D.N., Van Den Berg, B., 2017. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* 541, 407–411. <https://doi.org/10.1038/nature20828>
- Goh, Y.J., Klaenhammer, T.R., 2015. Genetic mechanisms of prebiotic oligosaccharide metabolism in probiotic microbes. *Annu. Rev. Food Sci. Technol.* 6, 137–156. <https://doi.org/10.1146/annurev-food-022814-015706>
- Grand, M., Aubourg, M., Pikis, A., Thompson, J., Deutscher, J., Hartke, A., Sauvageot, N., 2019. Characterization of the gen locus involved in  $\beta$ -1,6-oligosaccharide utilization by *Enterococcus faecalis*. *Mol. Microbiol.* 112, 1744–1756. <https://doi.org/10.1111/mmi.14390>

- Granell, M., León, X., Leblanc, G., Padrós, E., Lórenz-Fonfrá, V.A., 2010. Structural insights into the activation mechanism of melibiose permease by sodium binding. *Proc. Natl. Acad. Sci. U. S. A.* 107, 22078–22083. <https://doi.org/10.1073/pnas.1008649107>
- Grøftehauge, M.K., Hajizadeh, N.R., Swann, M.J., Pohl, E., 2015. Protein-ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr. Sect. D Biol. Crystallogr.* 71, 36–44. <https://doi.org/10.1107/S1399004714016617>
- Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W., Brumer, H., 2017. Polysaccharide utilization loci: fuelling microbial communities. *J. Bacteriol.* 199, e00860-16. <https://doi.org/10.1128/JB.00860-16>
- Guan, L., Mirza, O., Verner, G., Iwata, S., Kaback, H.R., 2007. Structural determination of wild-type lactose permease. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15294–15298. <https://doi.org/10.1073/pnas.0707688104>
- Han, Y., Agarwal, V., Dodds, D., Kims, J., Bae, B., Mackie, R.I., Nair, S.K., Cann, I.K.O., 2012. Biochemical and structural insights into xylan utilization by the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *J. Biol. Chem.* 287, 34946–34960. <https://doi.org/10.1074/jbc.M112.391532>
- Hemsworth, G.R., Dejean, G., Davies, G.J., Brumer, H., 2016. Learning from microbial strategies for polysaccharide degradation. *Biochem. Soc. Trans.* 44, 94–108. <https://doi.org/10.1042/BST20150180>
- Henrich, A., Kuhlmann, N., Eck, A.W., Krämer, R., Seibold, G.M., 2013. Maltose Uptake by the Novel ABC Transport System MusEFGK2I Causes Increased Expression of ptsG in *Corynebacterium glutamicum*. *J. Bacteriol.* 195, 2573–2584. <https://doi.org/10.1128/JB.01629-12>
- Henrissat, B., Davies, G., 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644. <https://doi.org/10.1109/MCSoc.2016.13>

- Heuberger, E.H.M.L., Smits, E., Poolman, B., 2001. Xyloside Transport by XylP, a Member of the Galactoside-Pentoside-Hexuronide Family. *J. Biol. Chem.* 276, 34465–34472. <https://doi.org/10.1074/jbc.M105460200>
- Hobbs, J.K., Pluvinaige, B., Boraston, A.B., 2018. Glycan-metabolizing enzymes in microbe–host interactions: the *Streptococcus pneumoniae* paradigm. *FEBS Lett.* 592, 3865–3897. <https://doi.org/10.1002/1873-3468.13045>
- Hospital, T.H., Fidalgo, R., Surgery, R., Care, P., 2010. The maltose ABC transporter in the 21st century – towards a structural-dynamic perspective on its mode of action, *Annals of Neurology*.
- Hugouvieux-Cotte-Pattat, N., Blot, N., Reverchon, S., 2001. Identification of TogMNAB, an ABC transporter which mediates the uptake of pectic oligomers in *Erwinia chrysanthemi* 3937. *Mol. Microbiol.* 41, 1113–1123. <https://doi.org/10.1046/j.1365-2958.2001.02564.x>
- Hugouvieux-Cotte-Pattat, N., Reverchon, S., 2001. Two transporters, TogT and TogMNAB, are responsible for oligogalacturonide uptake in *Erwinia chrysanthemi* 3937. *Mol. Microbiol.* 41, 1125–1132. <https://doi.org/10.1046/j.1365-2958.2001.02565.x>
- Imai, K., Hall, B.G., 1981. Properties of the lactose transport system in *Klebsiella* sp. strain CT-1. *J. Bacteriol.* 145, 1459–1462.
- Iyer, R., Camilli, A., 2007. Sucrose metabolism contributes to *in vivo* fitness of *Streptococcus pneumoniae*. *Mol. Microbiol.* 66, 1–13. <https://doi.org/10.1111/j.1365-2958.2007.05878.x>
- Jayani, R.S., Saxena, S., Gupta, R., 2005. Microbial pectinolytic enzymes: A review. *Process Biochem.* 40, 2931–2944. <https://doi.org/10.1016/j.procbio.2005.03.026>
- Jeckelmann, J.M., Erni, B., 2019. Carbohydrate Transport by Group Translocation: The Bacterial Phosphoenolpyruvate: Sugar Phosphotransferase System, *Subcellular Biochemistry*. Springer International Publishing. [https://doi.org/10.1007/978-3-030-18768-2\\_8](https://doi.org/10.1007/978-3-030-18768-2_8)
- Jecklin, M.C., Schauer, S., Dumelin, C.E., Zenobi, R., 2009. Label-free determination of protein-ligand binding constants using mass spectrometry and validation using surface

- plasmon resonance and isothermal titration calorimetry. *J. Mol. Recognit.* 22, 319–329. <https://doi.org/10.1002/jmr.951>
- Jensen, J.B., Peters, N.K., Bhuvaneswari, T. V., 2002. Redundancy in periplasmic binding protein-dependent transport systems for trehalose, sucrose, and maltose in *Sinorhizobium meliloti*. *J. Bacteriol.* 184, 2978–2986. <https://doi.org/10.1128/JB.184.11.2978-2986.2002>
- Joglekar, P., Sonnenburg, E.D., Higginbottom, S.K., Earle, K.A., Morland, C., Shapiro-Ward, S., Bolam, D.N., Sonnenburg, J.L., 2018. Genetic variation of the SusC/SusD homologs from a polysaccharide utilization locus underlies divergent fructan specificities and functional adaptation in *Bacteroides thetaiotaomicron* strains. *mSphere* 3, 1–11. <https://doi.org/10.1128/mSphereDirect.00185-18>
- Johnsen, U., Ortjohann, M., Sutter, J.-M., Geweke, S., Schönheit, P., 2019. Uptake of D-xylose and L-arabinose in *Haloferax volcanii* involves an ABC transporter of the CUT1 subfamily. *FEMS Microbiol. Lett.* 366, 1–19. <https://doi.org/10.1093/femsle/fnz089>
- Kaback, H.R., 2015. A chemiosmotic mechanism of symport. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1259–1264. <https://doi.org/10.1073/pnas.1419325112>
- Kaback, H.R., Guan, L., 2019. It takes two to tango: The dance of the permease. *J. Gen. Physiol.* 151, 878–886. <https://doi.org/10.1085/jgp.201912377>
- Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M., Teeling, H., Amann, R.I., Schweder, T., 2014. Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes “*Gramella forsetii*” KT0803. *ISME J.* 8, 1492–1502. <https://doi.org/10.1038/ismej.2014.4>
- Kachroo, A.H., Kancherla, A.K., Singh, N.S., Varshney, U., Mahadevan, S., 2007. Mutations that alter the regulation of the chb operon of *Escherichia coli* allow utilization of cellobiose. *Mol. Microbiol.* 66, 1382–1395. <https://doi.org/10.1111/j.1365-2958.2007.05999.x>
- Kalscheuer, R., Weinrick, B., Veeraraghavan, U., Besra, G.S., Jacobs, W.R., 2010. Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of

- Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. 107, 21761–21766.  
<https://doi.org/10.1073/pnas.1014642108>
- Kamerling, J.P., Gerwig, G.J., 2007. Strategies for the Structural Analysis of Carbohydrates, in: *Comprehensive Glycoscience: From Chemistry to Systems Biology*. Elsevier, pp. 1–68. <https://doi.org/10.1016/B978-044451967-2/00032-5>
- Kaneko, A., Uenishi, K., Baba, S., Mizuno, N., Maruyama, Y., Murata, K., Mikami, B., Hashimoto, W., Kumasaka, T., 2017. A solute-binding protein in the closed conformation induces ATP hydrolysis in a bacterial ATP-binding cassette transporter involved in the import of alginate. *J. Biol. Chem.* 292, 15681–15690.  
<https://doi.org/10.1074/jbc.m117.793992>
- Kaplan, H., Hutkins, R.W., 2003. Metabolism of fructooligosaccharides by *Lactobacillus paracasei* 1195. *Appl. Environ. Microbiol.* 69, 2217–2222.  
<https://doi.org/10.1128/AEM.69.4.2217-2222.2003>
- Kappelman, L., Krüger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F., Becher, D., Shapiro, N., Schweder, T., Amann, R.L., Teeling, H., 2019. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J.* 13, 76–91.  
<https://doi.org/10.1038/s41396-018-0242-6>
- King, S.C., Wilson, T.H., 1990. Identification of valine 177 as a mutation altering specificity for transport of sugars by the *Escherichia coli* lactose carrier. Enhanced specificity for sucrose and maltose. *J. Biol. Chem.* 265, 9638–9644.
- Klein, W., Horlacher, R., Boos, W., 1995. Molecular analysis of treB encoding the *Escherichia coli* enzyme II specific for trehalose. *J. Bacteriol.* 177, 4043–4052.  
<https://doi.org/10.1128/jb.177.14.4043-4052.1995>
- Koropatkin, N., Martens, E.C., Gordon, J.I., Smith, T.J., 2009. The structure Of a SusD homolog, BT1043, involved in mucin O- glycan utilization in a prominent human gut symbiont. *Biochemistry* 23, 1–7. <https://doi.org/10.1038/jid.2014.371>

- Koropatkin, N.M., Cameron, E.A., Martens, E.C., 2012. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* 10, 323–335.  
<https://doi.org/10.1038/nrmicro2746>
- Koropatkin, N.M., Martens, E.C., Gordon, J.I., Smith, T.J., 2008. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* 16, 1105–1115. <https://doi.org/10.1016/j.str.2008.03.017>
- Kotrba, P., Inui, M., Yukawa, H., 2003. A single V317A or V317M substitution in Enzyme II of a newly identified  $\beta$ -glucoside Phosphotransferase and utilization system of *Corynebacterium glutamicum* R extends its specificity towards cellobiose. *Microbiology* 149, 1569–1580. <https://doi.org/10.1099/mic.0.26053-0>
- Kowalczyk, M., Cocaign-Bousquet, M., Loubiere, P., Bardowski, J., 2008. Identification and functional characterisation of cellobiose and lactose transport systems in *Lactococcus lactis* IL1403. *Arch. Microbiol.* 189, 187–196. <https://doi.org/10.1007/s00203-007-0308-8>
- Kundig, W., Ghosh, S., Roseman, S., 1964. Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proc. Natl. Acad. Sci. U. S. A* 52, 1067–1074. <https://doi.org/10.1073/pnas.52.4.1067>
- La Rosa, S.L., Leth, M.L., Michalak, L., Hansen, M.E., Pudlo, N.A., Glowacki, R., Pereira, G., Workman, C.T., Arntzen, M.Ø., Pope, P.B., Martens, E.C., Hachem, M.A., Westereng, B., 2019. The human gut Firmicute *Roseburia intestinalis* is a primary degrader of dietary  $\beta$ -mannans. *Nat. Commun.* 10, 905. <https://doi.org/10.1038/s41467-019-08812-y>
- Ladevèze, S., Laville, E., Despres, J., Mosoni, P., Potocki-Véronèse, G., 2017. Mannoside recognition and degradation by bacteria. *Biol. Rev.* 92, 1969–1990.  
<https://doi.org/10.1111/brv.12316>
- Laine, R.A., 1994. Invited commentary: a calculation of all possible oligosaccharide isomers both branched and linear yields  $1.05 \times 10^{12}$  structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis. *Glycobiology* 4, 759–767. <https://doi.org/10.1093/glycob/4.6.759>

- Lap bie, P., Lombard, V., Drula, E., Terrapon, N., Henrissat, B., 2019. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat. Commun.* 10, 2043. <https://doi.org/10.1038/s41467-019-10068-5>
- Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., Klinger, S., Pudlo, N.A., Urs, K., Koropatkin, N.M., Creagh, A.L., Haynes, C.A., Kelly, A.G., Cederholm, S.N., Davies, G.J., Martens, E.C., Brumer, H., 2014. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* 506, 498–502. <https://doi.org/10.1038/nature12907>
- Larsbrink, J., Zhu, Y., Kharade, S.S., Kwiatkowski, K.J., Eijsink, V.G.H., Koropatkin, N.M., McBride, M.J., Pope, P.B., 2016. A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnol. Biofuels* 9, 1–16. <https://doi.org/10.1186/s13068-016-0674-z>
- Larsen, R.A., Thomas, M.G., Postle, K., 1999. Protonmotive force, ExbB and ligand-bound FepA drive conformational changes in TonB. *Mol. Microbiol.* 31, 1809–1824. <https://doi.org/10.1046/j.1365-2958.1999.01317.x>
- Lee, S., Kang, M., Bae, J.H., Sohn, J.H., Sung, B.H., 2019. Bacterial Valorization of Lignin: Strains, Enzymes, Conversion Pathways, Biosensors, and Perspectives. *Front. Bioeng. Biotechnol.* 7, 1–18. <https://doi.org/10.3389/fbioe.2019.00209>
- Lee, S.J., Lee, S.-J., Lee, D.-W., 2013. Design and development of synthetic microbial platform cells for bioenergy. *Front. Microbiol.* 4, 1–13. <https://doi.org/10.3389/fmicb.2013.00092>
- Leth, M.L., Ejby, M., Workman, C., Ewald, D.A., Pedersen, S.S., Sternberg, C., Bahl, M.I., Licht, T.R., Aachmann, F.L., Westereng, B., Hachem, M.A., 2018. Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* 3, 570–580. <https://doi.org/10.1038/s41564-018-0132-8>
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., Henrissat, B., 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6, 1. <https://doi.org/10.1186/1754-6834-6-41>

- Li, Z., Chai, W., 2019. Mucin O-glycan microarrays. *Curr. Opin. Struct. Biol.* 56, 187–197.  
<https://doi.org/10.1016/j.sbi.2019.03.032>
- Lian, J., Li, Y., Hamedirad, M., Zhao, H., 2014. Directed evolution of a cellodextrin transporter for improved biofuel production under anaerobic conditions in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 111, 1521–1531.  
<https://doi.org/10.1002/bit.25214>
- Liang, W.J., Wilson, K.J., Xie, H., Knol, J., Suzuki, S., Rutherford, N.G., Henderson, P.J.F., Jefferson, R.A., 2005. The *gusBC* genes of *Escherichia coli* encode a glucuronide transport system. *J. Bacteriol.* 187, 2377–2385. <https://doi.org/10.1128/JB.187.7.2377-2385.2005>
- Licht, A., Bommer, M., Werther, T., Neumann, K., Hobe, C., Schneider, E., 2019. Structural and functional characterization of a maltose/maltodextrin ABC transporter comprising a single solute binding domain (MalE) fused to the transmembrane subunit MalF. *Res. Microbiol.* 170, 1–12. <https://doi.org/10.1016/j.resmic.2018.08.006>
- Linke, C.M., Woodiga, S.A., Meyers, D.J., Buckwalter, C.M., Salhi, H.E., King, S.J., 2013. The ABC transporter encoded at the pneumococcal fructooligosaccharide utilization locus determines the ability to utilize long- and short-chain fructooligosaccharides. *J. Bacteriol.* 195, 1031–1041. <https://doi.org/10.1128/JB.01560-12>
- Liu, X., Zeng, J., Huang, K., Wang, J., 2019. Structure of the mannose transporter of the bacterial phosphotransferase system. *Cell Res.* 29, 680–682.  
<https://doi.org/10.1038/s41422-019-0194-z>
- Liu, Y., Beyer, A., Aebersold, R., 2016. On the dependency of cellular protein levels on mRNA abundance. *Cell* 165, 535–550. <https://doi.org/10.1016/j.cell.2016.03.014>
- Locher, K.P., 2016. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat. Struct. Mol. Biol.* 23, 487–493. <https://doi.org/10.1038/nsmb.3216>
- Lohmiller, S., Hantke, K., Patzer, S.I., Braun, V., 2008. TonB-dependent maltose transport by *Caulobacter crescentus*. *Microbiology* 154, 1748–1754.  
<https://doi.org/10.1099/mic.0.2008/017350-0>

- Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M., Henrissat, B., 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J.* 432, 437–444. <https://doi.org/10.1042/BJ20101185>
- Lombard, V., Ramulu H.G., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. <https://doi.org/10.1093/nar/gkt1178>
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T., 2017. Transcriptomics technologies. *PLOS Comput. Biol.* 13, e1005457. <https://doi.org/10.1371/journal.pcbi.1005457.t005>
- Luis, A.S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., Baslé A., Cartmell, A., Terrapon, N., Stott, K., Lowe, E.C., McLean, R., Shearer, K., Schückel, J., Venditto, I., Ralet, M.C., Henrissat, B., Martens, E.C., Mosimann, S.C., Abbott, D.W., Gilbert, H.J., 2018. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat. Microbiol.* 3, 210–219. <https://doi.org/10.1038/s41564-017-0079-1>
- Mächtel, R., Narducci, A., Griffith, D.A., Cordes, T., Orelle, C., 2019. An integrated transport mechanism of the maltose ABC importer. *Res. Microbiol.* 170. <https://doi.org/10.1016/j.resmic.2019.09.004>
- Mackenzie, A.K., Naas, A.E., Kracun, S.K., Schückel, J., Fangel, J.U., Agger, J.W., Willats, W.G.T., Eijsink, V.G.H., Pope, P.B., 2015. A polysaccharide utilization locus from an uncultured Bacteroidetes phylotype suggests ecological adaptation and substrate versatility. *Appl. Environ. Microbiol.* 81, 187–195. <https://doi.org/10.1128/AEM.02858-14>
- Mackenzie, A.K., Pope, P.B., Pedersen, H.L., Gupta, R., Morrison, M., Willats, W.G.T., Eijsink, V.G.H., 2012. Two SusD-like proteins encoded within a polysaccharide utilization locus of an uncultured ruminant bacteroidetes phylotype bind strongly to cellulose. *Appl. Environ. Microbiol.* 78, 5935–5937. <https://doi.org/10.1128/AEM.01164-12>
- Majd, H., King, M.S., Palmer, S.M., Smith, A.C., Elbourne, L.D.H., Paulsen, I.T., Sharples, D., Henderson, P.J.F., Kunji, E.R.S., 2018. Screening of candidate substrates and

- coupling ions of transporters by thermostability shift assays. *Elife* 7, 1–17.  
<https://doi.org/10.7554/eLife.38821>
- Marchesi, J.R., Ravel, J., 2015. The vocabulary of microbiome research: a proposal. *Microbiome* 3, 1–3. <https://doi.org/10.1186/s40168-015-0094-5>
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., Bryant, S.H., 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203.  
<https://doi.org/10.1093/nar/gkw1129>
- Marion, C., Stewart, J.M., Tazi, M.F., Burnaugh, A.M., Linke, C.M., Woodiga, S.A., King, S.J., 2012. *Streptococcus pneumoniae* can utilize multiple sources of hyaluronic acid for growth. *Infect. Immun.* 80, 1390–1398. <https://doi.org/10.1128/IAI.05756-11>
- Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., Gordon, J.I., 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* 9. <https://doi.org/10.1371/journal.pbio.1001221>
- Martin, S.A., Russell, J.B., 1987. Transport and phosphorylation of disaccharides by the ruminal bacterium *Streptococcus bovis*. *Appl. Environ. Microbiol.* 53, 2388–2393.
- Maruyama, Y., Hashimoto, W., Murata, K., 2019. Structural studies on bacterial system used in the recognition and uptake of the macromolecule alginate. *Biosci. Biotechnol. Biochem.* 83, 794–802. <https://doi.org/10.1080/09168451.2019.1578642>
- Maruyama, Y., Itoh, T., Kaneko, A., Nishitani, Y., Mikami, B., Hashimoto, W., Murata, K., 2015. Structure of a Bacterial ABC Transporter Involved in the Import of an Acidic Polysaccharide Alginate. *Structure* 23, 1643–1654.  
<https://doi.org/10.1016/j.str.2015.06.021>
- Matsuki, T., Yahagi, K., Mori, H., Matsumoto, H., Hara, T., Tajima, S., Ogawa, E., Kodama, H., Yamamoto, K., Yamada, T., Matsumoto, S., Kurokawa, K., 2016. A key genetic

- factor for fucosyllactose utilization affects infant gut microbiota development. *Nat. Commun.* 7, 1–12. <https://doi.org/10.1038/ncomms11939>
- McCann, M.C., Knox, J.P., 2011. Plant cell wall biology: polysaccharides in architectural and developmental contexts, in: Ulvskov, P. (Ed.), *Annual Plant Reviews: Plant Polysaccharides, Biosynthesis and Bioengineering*. Oxford: Wiley-Blackwell, pp. 343–366. <https://doi.org/10.1002/9781444391015.ch14>
- McCoy, J.G., Ren, Z., Stanevich, V., Lee, J., Mitra, S., Levin, E.J., Poget, S., Quick, M., Im, W., Zhou, M., 2016. The structure of a sugar transporter of the glucose EHC superfamily provides insight into the elevator mechanism of membrane transport. *Physiol. Behav.* 176, 139–148. <https://doi.org/10.1016/j.physbeh.2017.03.040>
- McNulty, N.P., Wu, M., Erickson, A.R., Pan, C., Erickson, B.K., Martens, E.C., Pudlo, N.A., Muegge, B.D., Henrissat, B., Hettich, R.L., Gordon, J.I., 2013. Effects of Diet on Resource Utilization by a Model Human Gut Microbiota Containing *Bacteroides cellulosilyticus* WH2, a Symbiont with an Extensive Glycobiome. *PLoS Biol.* 11. <https://doi.org/10.1371/journal.pbio.1001637>
- Meiborn, K.L., Li, X.B., Nielsen, A.T., Wu, C.Y., Roseman, S., Schoolnik, G.K., 2004. The *Vibrio cholerae* chitin utilization program. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2524–2529. <https://doi.org/10.1073/pnas.0308707101>
- Mikkat, S., Hagemann, M., 2000. Molecular analysis of the ggtBCD gene cluster of *Synechocystis* sp. strain PCC6803 encoding subunits of an ABC transporter for osmoprotective compounds. *Arch. Microbiol.* 174, 273–282. <https://doi.org/10.1007/s002030000201>
- Mishima, Y., Momma, K., Hashimoto, W., Mikami, B., Murata, K., 2003. Crystal structure of AlgQ2, a macromolecule (alginate)-binding protein of *Sphingomonas* sp. A1, complexed with an alginate tetrasaccharide at 1.6-Å resolution. *J. Biol. Chem.* 278, 6552–6559. <https://doi.org/10.1074/jbc.M209932200>
- Momma, K., Okamoto, M., Mishima, Y., Mori, S., Hashimoto, W., Murata, K., 2000. A novel bacterial ATP-binding cassette transporter system that allows uptake of macromolecules. *J. Bacteriol.* 182, 3998–4004. <https://doi.org/10.1128/JB.182.14.3998-4004.2000>

- Moon, M.W., Kim, H.J., Oh, T.K., Shin, C.S., Lee, J.S., Kim, S.J., Lee, J.K., 2005. Analyses of enzyme II gene mutants for sugar transport and heterologous expression of fructokinase gene in *Corynebacterium glutamicum* ATCC 13032. *FEMS Microbiol. Lett.* 244, 259–266. <https://doi.org/10.1016/j.femsle.2005.01.053>
- Mystkowska, A.A., Robb, C., Vidal-Melgosa, S., Vanni, C., Fernandez-Guerra, A., Hühne, M., Hehemann, J.H., 2018. Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan-binding protein of a marine Bacteroidetes. *FEBS J.* 285, 4465–4481. <https://doi.org/10.1111/febs.14674>
- Naas, A.E., Mackenzie, A.K., Mravec, J., Schückel, J., Willats, W.G.T., Eijsink, V.G.H., Pope, P.B., 2014. Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio* 5, 4–9. <https://doi.org/10.1128/mBio.01401-14>
- Nanavati, D.M., Nguyen, T.N., Noll, K.M., 2005. Substrate specificities and expression patterns reflect the evolutionary divergence of maltose ABC transporters in *Thermotoga maritima*. *J. Bacteriol.* 187, 2002–2009. <https://doi.org/10.1128/JB.187.6.2002-2009.2005>
- Nanavati, D.M., Thirangoon, K., Noll, K.M., 2006. Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. *Appl. Environ. Microbiol.* 72, 1336–1345. <https://doi.org/10.1128/AEM.72.2.1336-1345.2006>
- Nataf, Y., Yaron, S., Bayer, E.A., Stahl, F., Scheper, T.-H., Shoham, Y., Sonenshein, A.L., Lamed, R., 2009. Cellodextrin and Laminaribiose ABC Transporters in *Clostridium thermocellum*. *J. Bacteriol.* 191, 203–209. <https://doi.org/10.1128/jb.01190-08>
- Ndeh, D., Baslé A., Strahl, H., Yates, E.A., McClurg, U.L., Henrissat, B., Terrapon, N., Cartmell, A., 2020. Metabolism of multiple glycosaminoglycans by *Bacteroides thetaiotaomicron* is orchestrated by a versatile core genetic locus. *Nat. Commun.* 11, 1–12. <https://doi.org/10.1038/s41467-020-14509-4>
- Ndeh, D., Rogowski, A., Cartmell, A., Luis, A.S., Baslé A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R.A., Zhu, Y., O'Neill, M.A., Urbanowicz, B.R., York, W.S., Davies, G.J., Abbott, D.W., Ralet, M.C., Martens, E.C., Henrissat, B., Gilbert, H.J., 2017. Complex pectin

- metabolism by gut bacteria reveals novel catalytic functions. *Nature* 544, 65–70.  
<https://doi.org/10.1038/nature21725>
- Nguyen, H.H., Park, J., Kang, S., Kim, M., 2015. Surface plasmon resonance: A versatile technique for biosensor applications. *Sensors (Switzerland)* 15, 10481–10510.  
<https://doi.org/10.3390/s150510481>
- Nguyen, T.X., Yen, M.R., Barabote, R.D., Saier, M.H., 2006. Topological predictions for integral membrane permeases of the phosphoenolpyruvate:sugar phosphotransferase system. *J. Mol. Microbiol. Biotechnol.* 11, 345–360. <https://doi.org/10.1159/000095636>
- Niesen, F.H., Berglund, H., Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* 2, 2212–2221.  
<https://doi.org/10.1038/nprot.2007.321>
- Nieves-Mori ón, M., Flores, E., 2018. Multiple ABC glucoside transporters mediate sugar-stimulated growth in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *Environ. Microbiol. Rep.* 10, 40–48. <https://doi.org/10.1111/1758-2229.12603>
- Oiki, S., Kamochi, R., Mikami, B., Murata, K., Hashimoto, W., 2017. Alternative substrate-bound conformation of bacterial solute-binding protein involved in the import of mammalian host glycosaminoglycans. *Sci. Rep.* 7, 1–13. <https://doi.org/10.1038/s41598-017-16801-8>
- Oldham, M.L., Chen, S., Chen, J., 2013. Structural basis for substrate specificity in the *Escherichia coli* maltose transport system. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18132–18137. <https://doi.org/10.1073/pnas.1311407110>
- Oldham, M.L., Khare, D., Quiócho, F.A., Davidson, A.L., Chen, J., 2007. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* 450, 515–521.  
<https://doi.org/10.1038/nature06264>
- Panwar, D., Kapoor, M., 2020. Transcriptional analysis of galactomannooligosaccharides utilization by *Lactobacillus plantarum* WCFS1. *Food Microbiol.* 86, 103336.  
<https://doi.org/10.1016/j.fm.2019.103336>

- Parry, S., Hanisch, F.G., Leir, S.H., Sutton-Smith, M., Morris, H.R., Dell, A., Harris, A., 2006. N-Glycosylation of the MUC1 mucin in epithelial cells and secretions. *Glycobiology* 16, 623–634. <https://doi.org/10.1093/glycob/cwj110>
- Patil, N.P., Le, V., Sligar, A.D., Mei, L., Chavarria, D., Yang, E.Y., Baker, A.B., 2018. Algal Polysaccharides as Therapeutic Agents for Atherosclerosis. *Front. Cardiovasc. Med.* 5, 1–18. <https://doi.org/10.3389/fcvm.2018.00153>
- Peng, Y., Kumar, S., Hernandez, R.L., Jones, S.E., Cadle, K.M., Smith, K.P., Varela, M.F., 2009. Evidence for the transport of maltose by the sucrose permease, CscB, of *Escherichia coli*. *J. Membr. Biol.* 228, 79–88. <https://doi.org/10.1007/s00232-009-9161-9>
- Peterkofsky, A., Wang, G., Garrett, D.S., Lee, B.R., Seaok, Y.J., Clore, G.M., 2001. Three-dimensional structures of protein-protein complexes in the *E. coli* PTS. *J. Mol. Microbiol. Biotechnol.* 3, 347–354.
- Phansopa, C., Roy, S., Rafferty, J.B., Douglas, C.W.I., Pandhal, J., Wright, P.C., Kelly, D.J., Stafford, G.P., 2014. Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling *Bacteroidetes* species. *Biochem. J.* 458, 499–511. <https://doi.org/10.1042/BJ20131415>
- Pikis, A., Hess, S., Arnold, I., Erni, B., Thompson, J., 2006. Genetic requirements for growth of *Escherichia coli* K12 on methyl- $\alpha$ -D-glucopyranoside and the five  $\alpha$ -D-glucosyl-D-fructose isomers of sucrose. *J. Biol. Chem.* 281, 17900–17908. <https://doi.org/10.1074/jbc.M601183200>
- Poolman, B., Royer, T.J., Mainzer, S.E., Schmidt, B.F., 1989. Lactose transport system of *Streptococcus thermophilus*: a hybrid protein with homology to the melibiose carrier and enzyme III of phosphoenolpyruvate-dependent phosphotransferase systems. *J. Bacteriol.* 171, 244–253. <https://doi.org/10.1128/jb.171.1.244-253.1989>
- Pope, P.B., Mackenzie, A.K., Gregor, I., Smith, W., Sundset, M.A., McHardy, A.C., Morrison, M., Eijsink, V.G.H., 2012. Metagenomics of the svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PLoS One* 7, 1–10. <https://doi.org/10.1371/journal.pone.0038571>

- Prestegard, J.H., Liu, J., Widmalm, G., 2017. Oligosaccharides and Polysaccharides, in: Essentials of Carbohydrate Chemistry. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, pp. 2015–2017. <https://doi.org/10.1101/glycobiology.3e.003>
- Prydz, K., 2015. Determinants of glycosaminoglycan (GAG) structure. *Biomolecules* 5, 2003–2022. <https://doi.org/10.3390/biom5032003>
- Quioco, F.A., Spurlino, J.C., Rodseth, L.E., 1997. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* 5, 997–1015. [https://doi.org/10.1016/S0969-2126\(97\)00253-0](https://doi.org/10.1016/S0969-2126(97)00253-0)
- Rakoff-Nahoum, S., Coyne, M.J., Comstock, L.E., 2014. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr. Biol.* 24, 40–49. <https://doi.org/10.1016/j.cub.2013.10.077>
- Ravcheev, D.A., Godzik, A., Osterman, A.L., Rodionov, D.A., 2013. Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: Comparative genomics reconstruction of metabolic and regulatory networks. *BMC Genomics* 14, 1–17. <https://doi.org/10.1186/1471-2164-14-873>
- Reddy, V.S., Shlykov, M. a, Castillo, R., Sun, E.I., Saier, M.H., 2012. The major facilitator superfamily (MFS) revisited. *FEBS J.* 279, 2022–2035. <https://doi.org/10.1111/j.1742-4658.2012.08588.x>.
- Reeves, A.R., D’Elia, J.N., Frias, J., Salyers, A.A., 1996. A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J. Bacteriol.* 178, 823–830. <https://doi.org/10.1128/jb.178.3.823-830.1996>
- Reichenbach, T., Kalyani, D., Gandini, R., Svartstrom, O., Aspeborg, H., Divne, C., 2018. Structural and biochemical characterization of the *Cutibacterium acnes*  $\alpha$ -1,4-mannosidase that targets the N-glycan core of host glycoproteins. *PLoS One* 13, 1–27. <https://doi.org/10.1371/journal.pone.0204703>
- Reidl, J., Boos, W., 1991. The malX malY operon of *Escherichia coli* encodes a novel enzyme II of the phosphotransferase system recognizing glucose and maltose and an

- enzyme abolishing the endogenous induction of the maltose system. *J. Bacteriol.* 173, 4862–4876. <https://doi.org/10.1128/jb.173.15.4862-4876.1991>
- Ren, Z., Lee, J., Moosa, M.M., Nian, Y., Hu, L., Xu, Z., McCoy, J.G., Ferreon, A.C.M., Im, W., Zhou, M., 2018. Structure of an EIIC sugar transporter trapped in an inward-facing conformation. *Proc. Natl. Acad. Sci. U. S. A.* 115, 5962–5967. <https://doi.org/10.1073/pnas.1800647115>
- Renzi, F., Manfredi, P., Mally, M., Moes, S., Jenö, P., Cornelis, G.R., 2011. The N-glycan glycoprotein deglycosylation complex (gpd) from *Capnocytophaga Canimorsus* deglycosylates human IgG. *PLoS Pathog.* 7. <https://doi.org/10.1371/journal.ppat.1002118>
- Ridley, B.L., Neill, M.A.O., Mohnen, D., 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* 57, 929–967. [https://doi.org/https://doi.org/10.1016/S0031-9422\(01\)00113-3](https://doi.org/https://doi.org/10.1016/S0031-9422(01)00113-3)
- Robillard, G.T., Broos, J., 1999. Structure/function studies on the bacterial carbohydrate transporters, enzymes II, of the phosphoenolpyruvate-dependent phosphotransferase system. *Biochim. Biophys. Acta - Rev. Biomembr.* 1422, 73–104. [https://doi.org/10.1016/S0304-4157\(99\)00002-7](https://doi.org/10.1016/S0304-4157(99)00002-7)
- Rodionov, D.A., Yang, C., Li, X., Rodionova, I.A., Wang, Y., Obratzsova, A.Y., Zagnitko, O.P., Overbeek, R., Romine, M.F., Reed, S., Fredrickson, J.K., Nealson, K.H., Osterman, A.L., 2010. Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics* 11. <https://doi.org/10.1186/1471-2164-11-494>
- Rodríguez-Dáz, J., Rubio-del-Campo, A., Yebra, M.J., 2012. *Lactobacillus casei* ferments the N-Acetylglucosamine moiety of fucosyl- $\alpha$ -1,3-N-acetylglucosamine and excretes L-fucose. *Appl. Environ. Microbiol.* 78, 4613–4619. <https://doi.org/10.1128/AEM.00474-12>
- Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Baslé, A., Morland, C., Day, A.M., Zheng, H., Rogers, T.E., Thompson, P., Hawkins, A.R., Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J., Bolam, D.N., 2015. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* 6. <https://doi.org/10.1038/ncomms8481>

- Saier, M.H., 2015. The Bacterial Phosphotransferase System: New frontiers 50 years after its discovery. *J. Mol. Microbiol. Biotechnol.* 25, 73–78. <https://doi.org/10.1159/000381215>
- Saier, M.H., 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res.* 34, D181–D186. <https://doi.org/10.1093/nar/gkj001>
- Saier, M.H., 2000. Families of transmembrane sugar transport proteins. *Mol. Microbiol.* 35, 699–710. <https://doi.org/10.1046/j.1365-2958.2000.01759.x>
- Saito, A., Fujii, T., Shinya, T., Shibuya, N., Ando, A., Miyashita, K., 2008. The *msiK* gene, encoding the ATP-hydrolysing component of N,N'-diacetylchitobiose ABC transporters, is essential for induction of chitinase production in *Streptomyces coelicolor* A3(2). *Microbiology* 154, 3358–3365. <https://doi.org/10.1099/mic.0.2008/019612-0>
- Saulnier, D.M.A., Molenaar, D., De Vos, W.M., Gibson, G.R., Kolida, S., 2007. Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays. *Appl. Environ. Microbiol.* 73, 1753–1765. <https://doi.org/10.1128/AEM.01151-06>
- Sauvageot, N., Mokhtari, A., Joyet, P., Budin-Verneuil, A., Blancato, V.S., Repizo, G.D., Henry, C., Pikis, A., Thompson, J., Magni, C., Hartke, A., Deutscher, J., 2017. *Enterococcus faecalis* uses a phosphotransferase system permease and a host colonization-related ABC transporter for maltodextrin uptake. *J. Bacteriol.* 199, 1–17. <https://doi.org/10.1128/JB.00878-16>
- Schlösser, A., Jantos, J., Hackmann, K., Schrempf, H., 1999. Characterization of the binding protein-dependent cellobiose and cellotriose transport system of the cellulose degrader *Streptomyces reticuli*. *Appl. Environ. Microbiol.* 65, 2636–2643.
- Schneider, E., 2001. ABC transporters catalyzing carbohydrate uptake. *Res. Microbiol.* 152, 303–310. [https://doi.org/10.1016/S0923-2508\(01\)01201-3](https://doi.org/10.1016/S0923-2508(01)01201-3)
- Schönert, S., Seitz, S., Krafft, H., Feuerbaum, E.A., Andernach, I., Witz, G., Dahl, M.K., 2006. Maltose and maltodextrin utilization by *Bacillus subtilis*. *J. Bacteriol.* 188, 3911–3922. <https://doi.org/10.1128/JB.00213-06>

- Schouler, C., Taki, A., Chouikha, I., Moulin-Schouleur, M., Gilot, P., 2009. A genomic island of an extraintestinal pathogenic *Escherichia coli* strain enables the metabolism of fructooligosaccharides, which improves intestinal colonization. *J. Bacteriol.* 91, 388–393. <https://doi.org/10.1128/JB.01052-08>
- Schwalm, N.D., Groisman, E.A., 2017. Navigating the gut buffet: control of polysaccharide utilization in *Bacteroides* spp. *Trends Microbiol.* 25, 1005–1015. <https://doi.org/10.1016/j.tim.2017.06.009>
- Selvendran, R.R., 1984. The plant cell wall as a source of dietary fiber: chemistry and structure. *Am. J. Clin. Nutr.* 39, 320–337. <https://doi.org/10.1093/ajcn/39.2.320>
- Shelburne, S.A., Keith, D.B., Davenport, M.T., Horstmann, N., Brennan, R.G., Musser, J.M., 2008. Molecular characterization of group A Streptococcus maltodextrin catabolism and its role in pharyngitis. *Mol. Microbiol.* 69, 436–452. <https://doi.org/10.1111/j.1365-2958.2008.06290.x>
- Sheridan, P.O., Martin, J.C., Lawley, T.D., Browne, H.P., Harris, H.M.B., Bernalier-Donadille, A., Duncan, S.H., O’Toole, P.W., Scott, K.P., Flint, H.J., 2016. Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic *Firmicutes*. *Microb. Genomics* 2, e000043. <https://doi.org/10.1099/mgen.0.000043>
- Shinnick, S.G., Perez, S.A., Varela, M.F., 2003. Altered substrate selection of the melibiose transporter (MelY) of *Enterobacter cloacae* involving point mutations in Leu-88, Leu-91, and Ala-182 that confer enhanced maltose transport. *J. Bacteriol.* 185, 3672–3677. <https://doi.org/10.1128/JB.185.12.3672-3677.2003>
- Shipman, J.A., Cho, K.H., Siegel, H.A., Salyers, A.A., 1999. Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* 181, 7206–7211. <https://doi.org/10.1128/jb.181.23.7206-7211.1999>
- Shulami, S., Gat, O., Sonenshein, A.L., Shoham, Y., 1999. The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6. *J. Bacteriol.* 181, 3695–3704.

- Shulami, S., Raz-Pasteur, A., Tabachnikov, O., Gilead-Gropper, S., Shner, I., Shoham, Y., 2011. The L-arabinan utilization system of *Geobacillus stearothermophilus*. *J. Bacteriol.* 193, 2838–2850. <https://doi.org/10.1128/JB.00222-11>
- Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A.L., Shoham, Y., 2007. A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. *Appl. Environ. Microbiol.* 73, 874–884. <https://doi.org/10.1128/AEM.02367-06>
- Silhavy, T.J., Brickman, E., Bassford, P.J., Casadaban, M.J., Shuman, H.A., Schwartz, V., Guarente, L., Schwartz, M., Beckwith, J.R., 1979. Structure of the *malB* region in *Escherichia coli* K12. *Mol. Gen. Genet.* 174, 261–267. <https://doi.org/10.1007/bf00267798>
- Silva, Z., Sampaio, M., Henne, A., Bo, A., Gutzat, R., Boos, W., Costa, M.S., Santos, H., 2005. The High-Affinity Maltose / Trehalose ABC Transporter in the Extremely Thermophilic Bacterium *Thermus thermophilus* HB27 Also Recognizes Sucrose and Palatinose. *J. Bacteriol.* 187, 1210–1218. <https://doi.org/10.1128/JB.187.4.1210>
- Smith, K.A., Salyers, A.A., 1991. Characterization of a neopullulanase and an  $\alpha$ -glucosidase from *Bacteroides thetaiotaomicron* 95-1. *J. Bacteriol.* 173, 2962–2968. <https://doi.org/10.1128/jb.173.9.2962-2968.1991>
- Song, Y., Li, J., Meng, S., Yin, L., Xue, Y., Ma, Y., 2016. A novel manno-oligosaccharide binding protein identified in alkaliphilic *Bacillus* sp. N16-5 is involved in mannan utilization. *PLoS One* 11, 1–12. <https://doi.org/10.1371/journal.pone.0150059>
- Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N., Sonnenburg, J.L., 2010. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* 141, 1241–1252. <https://doi.org/10.1016/j.cell.2010.05.005>
- Soucy, S.M., Huang, J., Gogarten, J.P., 2015. Horizontal gene transfer: Building the web of life. *Nat. Rev. Genet.* 16, 472–482. <https://doi.org/10.1038/nrg3962>

- Steen, A.D., Crits-Christoph, A., Carini, P., DeAngelis, K.M., Fierer, N., Lloyd, K.G., Cameron Thrash, J., 2019. High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J.* 13, 3126–3130. <https://doi.org/10.1038/s41396-019-0484-y>
- Suzuki, R., Wada, J., Katayama, T., Fushinobu, S., Wakagi, T., Shoun, H., Sugimoto, H., Tanaka, A., Kumagai, H., Ashida, H., Kitaoka, M., Yamamoto, K., 2008. Structural and thermodynamic analyses of solute-binding protein from *Bifidobacterium longum* specific for core 1 disaccharide and lacto-N-biose. *J. Biol. Chem.* 283, 13165–13173. <https://doi.org/10.1074/jbc.M709777200>
- Tam, R., Saier, M.H., 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* 57, 320–346.
- Tamura, K., Foley, M.H., Gardill, B.R., Dejean, G., Schnizlein, M., Bahr, C.M.E., Louise Creagh, A., van Petegem, F., Koropatkin, N.M., Brumer, H., 2019. Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*. *Cell. Mol. Life Sci.* <https://doi.org/10.1007/s00018-019-03115-3>
- Tamura, K., Hemsworth, G.R., Déjean, G., Rogers, T.E., Pudlo, N.A., Urs, K., Jain, N., Davies, G.J., Martens, E.C., Brumer, H., 2017. Molecular mechanism by which prominent human gut bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides. *Cell Rep.* 21, 2030. <https://doi.org/10.1016/j.celrep.2017.11.013>
- Tan, L., Showalter, A.M., Egelund, J., Hernandez-Sanchez, A., Doblin, M.S., Bacic, A., 2012. Arabinogalactan-proteins and the research challenges for these enigmatic plant cell surface proteoglycans. *Front. Plant Sci.* 3, 1–10. <https://doi.org/10.3389/fpls.2012.00140>
- Tancula, E., Feldhaus, M.J., Bedzyk, L.A., Salyers, A.A., 1992. Location and characterization of genes involved in binding of starch to the surface of *Bacteroides thetaiotaomicron*. *J. Bacteriol.* 174, 5609–5616. <https://doi.org/10.1128/jb.174.17.5609-5616.1992>
- Tao, L., Sutcliffe, I.C., Russell, R.R.B., Ferretti, J.J., 1993. Cloning and expression of the multiple sugar metabolism (msm) operon of *Streptococcus mutans* in heterologous streptococcal hosts. *Infect. Immun.* 61, 1121–1125.

- Tauzin, A.S., Kwiatkowski, K.J., Orlovsky, N.I., Smith, C.J., Creagh, A.L., Haynes, C.A., Wawrzak, Z., Brumer, H., Koropatkin, N.M., 2016a. Molecular dissection of xyloglucan recognition in a prominent human gut symbiont. *MBio* 7, e02134-15.  
<https://doi.org/10.1128/mBio.02134-15>
- Tauzin, A.S., Laville, E., Xiao, Y., Nouaille, S., Le Bourgeois, P., Heux, S., Portais, J.C., Monsan, P., Martens, E.C., Potocki-Veronese, G., Bordes, F., 2016b. Functional characterization of a gene locus from an uncultured gut *Bacteroides* conferring xylo-oligosaccharides utilization to *Escherichia coli*. *Mol. Microbiol.* 102, 579–592.  
<https://doi.org/10.1111/mmi.13480>
- Tavoulari, S., Frillingos, S., 2008. Substrate Selectivity of the Melibiose Permease (MelY) from *Enterobacter cloacae*. *J. Mol. Biol.* 376, 681–693.  
<https://doi.org/10.1016/j.jmb.2007.12.015>
- Tchieu, J.H., Norris, V., Edwards, J.S., Saier, M.H., 2001. The complete phosphotransferase system in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 3, 329–46.
- Temple, M.J., Cuskin, F., Baslé A., Hickey, N., Speciale, G., Williams, S.J., Gilbert, H.J., Lowe, E.C., 2017. A Bacteroidetes locus dedicated to fungal 1,6--glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6--glucanase. *J. Biol. Chem.* 292, 10639–10650. <https://doi.org/10.1074/jbc.M117.787606>
- Terrapon, N., Lombard, V., Drula, É., Lap  bie, P., Al-Masaudi, S., Gilbert, H.J., Henrissat, B., 2018. PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* 46, D677–D683. <https://doi.org/10.1093/nar/gkx1022>
- Terrapon, N., Lombard, V., Gilbert, H.J., Henrissat, B., 2015. Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* 31, 647–655.  
<https://doi.org/10.1093/bioinformatics/btu716>
- Theilmann, M.C., Fredslund, F., Svensson, B., Lo Leggio, L., Abou Hachem, M., 2019. Substrate preference of an ABC importer corresponds to selective growth on  $\beta$ -(1,6)-galactosides in *Bifidobacterium animalis* subsp. *lactis*. *J. Biol. Chem.* 294, 11701–11711.  
<https://doi.org/10.1074/jbc.ra119.008843>

- Theodoulou, F.L., Kerr, I.D., 2015. ABC transporter research: going strong 40 years on. *Biochem. Soc. Trans.* 43, 1033–1040. <https://doi.org/10.1042/BST20150139>
- Titgemeyer, F., Jahreis, K., Ebner, R., Lengeler, J.W., 1996. Molecular analysis of the *scrA* and *scrB* genes from *Klebsiella pneumoniae* and plasmid pUR400, which encode the sucrose transport protein Enzyme IIScr of the phosphotransferase system and a sucrose-6-phosphate invertase. *Mol. Gen. Genet.* 250, 197–206. <https://doi.org/10.1007/BF02174179>
- Tobisch, S., Glaser, P., Krüger, S., Hecker, M., 1997. Identification and characterization of a new  $\beta$ -glucoside utilization system in *Bacillus subtilis*. *J. Bacteriol.* 179, 496–506. <https://doi.org/10.1128/jb.179.2.496-506.1997>
- Tomme, P., Boraston, A., Kormos, J.M., Warren, R.A.J., Kilburn, D.G., 2000. Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules. *Enzyme Microb. Technol.* 27, 453–458. [https://doi.org/10.1016/S0141-0229\(00\)00246-5](https://doi.org/10.1016/S0141-0229(00)00246-5)
- Tsujibo, H., Kosaka, M., Ikenishi, S., Sato, T., Miyamoto, K., Inamori, Y., 2004. Molecular Characterization of a High-Affinity Xylobiose Transporter of *Streptomyces thermoviolaceus* OPC-520 and Its Transcriptional Regulation. *J. Bacteriol.* 186, 1029–1037. <https://doi.org/10.1128/JB.186.4.1029-1037.2004>
- Uchiyama, T., Kaneko, R., Yamaguchi, J., Inoue, A., Yanagida, T., Nikaidou, N., Regue, M., Watanabe, T., 2003. Uptake of N,N'-diacetylchitobiose [(GlcNAc)<sub>2</sub>] via the phosphotransferase system is essential for chitinase production by *Serratia marcescens* 2170. *J. Bacteriol.* 185, 1776–1782. <https://doi.org/10.1128/JB.185.6.1776-1782.2003>
- Vahedi-Faridi, A., Licht, A., Bulut, H., Scheffel, F., Keller, S., Wehmeier, U.F., Saenger, W., Schneider, E., 2010. Crystal Structures of the Solute Receptor GacH of *Streptomyces glaucescens* in Complex with Acarbose and an Acarbose Homolog: Comparison with the Acarbose-Loaded Maltose-Binding Protein of *Salmonella typhimurium*. *J. Mol. Biol.* 397, 709–723. <https://doi.org/10.1016/j.jmb.2010.01.054>
- van Bueren, A.L., Saraf, A., Martens, E.C., Dijkhuizen, L., 2015. Differential metabolism of exopolysaccharides from probiotic lactobacilli by the human gut symbiont *Bacteroides*

- thetaitaomicron. *Appl. Environ. Microbiol.* 81, 3973–3983.  
<https://doi.org/10.1128/AEM.00149-15>
- Van Wezel, G.P., White, J., Bibb, M.J., Postma, P.W., 1997. The malEFG gene cluster of *Streptomyces coelicolor* A3(2): Characterization, disruption and transcriptional analysis. *Mol. Gen. Genet.* 254, 604–608. <https://doi.org/10.1007/s004380050458>
- Viegas, A., Manso, J., Nobrega, F.L., Cabrita, E.J., 2011. Saturation-transfer difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding. *J. Chem. Educ.* 88, 990–994. <https://doi.org/10.1021/ed101169t>
- Vuignier, K., Schappler, J., Veuthey, J.L., Carrupt, P.A., Martel, S., 2010. Drug-protein binding: A critical review of analytical tools. *Anal. Bioanal. Chem.* 398, 53–66.  
<https://doi.org/10.1007/s00216-010-3737-1>
- Wandersman, C., Schwartz, M., Ferenci, T., 1979. *Escherichia coli* mutants impaired in maltodextrin transport. *J. Bacteriol.* 140, 1–13.
- Watzlawick, H., Heravi, K.M., Altenbuchner, J., 2016. Role of the ganSPQAB operon in degradation of galactan by *Bacillus subtilis*. *J. Bacteriol.* 198, 2887–2896.  
<https://doi.org/10.1128/JB.00468-16>
- Webb, A.J., Homer, K.A., Hosie, A.H.F., 2008. Two closely related ABC transporters in *Streptococcus mutans* are involved in disaccharide and/or oligosaccharide uptake. *J. Bacteriol.* 190, 168–178. <https://doi.org/10.1128/JB.01509-07>
- Webb, A.J., Homer, K.A., Hosie, A.H.F., 2007. A phosphoenolpyruvate-dependent phosphotransferase system is the principal maltose transporter in *Streptococcus mutans*. *J. Bacteriol.* 189, 3322–3327. <https://doi.org/10.1128/JB.01633-06>
- Wefers, D., Dong, J., Abdel-Hamid, A.M., Paul, H.M., Pereira, G. V., Han, Y., Dodd, D., Baskaran, R., Mayer, B., Mackie, R.I., Cann, I., 2017. Enzymatic mechanism for arabinan degradation and transport in the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *Appl. Environ. Microbiol.* 83. <https://doi.org/10.1128/AEM.00794-17>

- Wei, X., Guo, Y., Shao, C., Sun, Z., Zhurina, D., Liu, D., Liu, W., Zou, D., Jiang, Z., Wang, X., Zhao, J., Shang, W., Li, X., Liao, X., Huang, L., Riedel, C.U., Yuan, J., 2012. Fructose uptake in *Bifidobacterium longum* NCC2705 is mediated by an ATP-binding cassette transporter. *J. Biol. Chem.* 287, 357–367. <https://doi.org/10.1074/jbc.M111.266213>
- Wieczorek, A.S., Schmidt, O., Chatzinotas, A., Von Bergen, M., Gorissen, A., Kolb, S., 2019. Ecological functions of agricultural soil bacteria and microeukaryotes in chitin degradation: A case study. *Front. Microbiol.* 10. <https://doi.org/10.3389/fmicb.2019.01293>
- Wilkins, S., 2015. Structure and mechanism of ABC transporters. *F1000Prime Rep.* 7, 426–431. <https://doi.org/10.1016/j.sbi.2004.06.005>
- Wilson, D.B., 2008. Three microbial strategies for plant cell wall degradation. *Ann. N. Y. Acad. Sci.* 1125, 289–297. <https://doi.org/10.1196/annals.1419.026>
- Wu, M.C., Chen, Y.C., Lin, T.L., Hsieh, P.F., Wang, J.T., 2012. Cellobiose-specific phosphotransferase system of *Klebsiella pneumoniae* and its importance in biofilm formation and virulence. *Infect. Immun.* 80, 2464–2472. <https://doi.org/10.1128/IAI.06247-11>
- Xiao, X., Wang, F., Satio, A., Majka, J., Schläpfer, A., Schrempf, H., 2002. The novel *Streptomyces olivaceoviridis* ABC transporter Ngc mediates uptake of N-acetylglucosamine and N, N'-diacetylchitobiose. *Mol. Genet. Genomics* 267, 429–439. <https://doi.org/10.1007/s00438-002-0640-2>
- Xu, S.Y., Huang, X., Cheong, K.L., 2017. Recent advances in marine algae polysaccharides: Isolation, structure, and activities. *Mar. Drugs* 15, 1–16. <https://doi.org/10.3390/md15120388>
- Yamamoto, H., Serizawa, M., Thompson, J., Sekiguchi, J., 2001. Regulation of the *glv* operon in *Bacillus subtilis*: YfiA (GlvR) is a positive regulator of the operon that is repressed through CcpA and cre. *J. Bacteriol.* 183, 5110–5121. <https://doi.org/10.1128/JB.183.17.5110-5121.2001>

- Yan, N., 2015. Structural Biology of the Major Facilitator Superfamily Transporters. *Annu. Rev. Biophys.* 44, 257–283. <https://doi.org/10.1146/annurev-biophys-060414-033901>
- Zhang, X.C., Zhao, Y., Heng, J., Jiang, D., 2015. Energy coupling mechanisms of MFS transporters. *Protein Sci.* 24, 1560–1579. <https://doi.org/10.1002/pro.2759>
- Zhu, Y., McBride, M.J., 2017. The unusual cellulose utilization system of the aerobic soil bacterium *Cytophaga hutchinsonii*. *Appl. Microbiol. Biotechnol.* 101, 7113–7127. <https://doi.org/10.1007/s00253-017-8467-2>
- Zúñiga, M., Comas, I., Linaje, R., Monedero, V., Yebra, M.J., Esteban, C.D., Deutscher, J., Pérez-Martínez, G., González-Candelas, F., 2005. Horizontal gene transfer in the molecular evolution of mannose PTS transporters. *Mol. Biol. Evol.* 22, 1673–1685. <https://doi.org/10.1093/molbev/msi163>



# Result



## First article

Glycosides represent the main energy sources for numerous microbiota in the gut, terrestrial and marine environments. According to the current paradigm, utilization of the structural complex and diverse glycosides in bacteria is usually orchestrated by polysaccharide utilization loci (PULs) encoding both carbohydrate-active enzymes (CAZymes) and transporters. As described in the previous chapter, polysaccharides are partially degraded to oligosaccharides by cell surface associated CAZymes, and then oligosaccharides are imported into the bacterial cell to be metabolized. Transporters thus play a critical role in glycoside utilization and a better understanding of them represents high potential in health management, to control the microbiota functioning, and industrial applications, to create chassis strains for synthetic biology capable of growing on non-refined plant biomass. However, the experimental investigation of the function of glycoside transporter faces more challenges than for CAZymes.

The activity-based functional metagenomic approach developed in our team, focusing mainly on glycoside degradation in mammalian guts, previously led to the identification of hundreds of *Escherichia coli* clones producing glycoside-hydrolase activities. In these studies, the type of vector used to construct the metagenomic libraries (fosmids, allowing the cloning of large DNA fragments, sizing from 30 to 40 kbp) was deliberately chosen to capture entire bacterial operon-like gene clusters encoding for the synergistic CAZymes which are necessary to deconstruct complex glycoside structures (Tasse et al., 2010). In these metagenomic PULs, the CAZyme-encoding genes are very often clustered with transporter encoding ones. In 2016, Tazuin *et al.* demonstrated that SusCD and MFS transporters can be produced in functional forms from a Bacteroidetes PUL cloned in fosmid and expressed in *E. coli* (Tazuin et al., 2016). But at the beginning of this thesis work, the following questions remained unanswered:

- What is the probability of producing functional transporters by expressing PULs' genes in *E. coli*?

- What are the families of glycoside transporters that can be functionally produced in *E. coli*?
- Could transporters from Gram-positive bacteria be functionally produced in the Gram-negative bacterium *E. coli* from gpPULs cloned in fosmids?
- What is the potential of the activity-based metagenomic approach to discover new transporter functions?

In order to address these questions, we developed a medium-throughput positive selection method to screen the ability of fosmid *E. coli* clones harboring metagenomic PULs to use

oligosaccharides as sole carbon source. The results are presented in this chapter, which constitutes the first draft of a future publication.

# Identification of functional glycoside transporters from uncultured bacteria

Zhi Wang<sup>a</sup>, Alexandra S. Tauzin<sup>a</sup> and Gabrielle Potocki-Veronese<sup>a</sup>

<sup>a</sup>TBI, CNRS, INRA, INSAT, Université de Toulouse, Toulouse, France

Address correspondence to Gabrielle Potocki-Veronese, veronese@insa-toulouse.fr

## Abstract

Transport is a crucial step in the metabolism of glycosides by bacteria, which is itself a key of microbiota functioning and equilibrium. However, the function of most transport proteins is unknown or only predicted, limiting our understanding of how bacteria utilize glycosides. Here, we present a robust activity-based screening method to identify functional glycoside transporters from microbiomes. The method is based on the co-expression in *Escherichia coli* of genes encoding transporters and carbohydrate-active enzymes (CAZymes) from metagenomic polysaccharide utilization loci (PULs) cloned in fosmids. To establish the proof of concept of the methodology, from four different metagenomic libraries issued of the mammalian microbiota, we selected 45 *E. coli* clones of which the metagenomic sequence contains at least one putative glycoside transporter and one functional CAZyme, identified by screening various glycoside-hydrolase activities. Growth tests were performed on glycosides which correspond to the validated or predicted enzymatic activity of the CAZymes identified in each PUL. This led to the identification of 13 clones that are able to utilize oligosaccharides as sole carbon sources, thanks to the production of transporters from PTS, ABC, MFS, and SusCD families. Seven of the 13 hit clones contain one clearly identified transporter, providing direct experimental evidence that these transporters are functional. In the six cases where two transporters are present in the sequence of a clone, the function of the transporters can be predicted from the flanking CAZymes or from similarity with previously characterized transporters, which facilitates further functional characterization. The results expand the vision on how glycosides are selectively metabolized by bacteria and open a new way to screen for glycoside-transporter specificity towards oligosaccharides of defined structures.

## Introduction

Carbohydrates, together with lipids, proteins and nucleic acids, are one of the four biological essential macromolecules in living systems. In Nature, repetitive units of monomers are linked to other glycosyl residues or non-glycosyl group by glycosidic bonds to form structurally diverse glycosides. Based on their chemical composition and degree of polymerization (DP), glycosides are classified into i) oligosaccharides, composed only of glycosyl residues with DP between 2 and 10 (Kamerling, 2007) (3 and 10 according to the IUB-IUPAC nomenclature), ii) polysaccharides, composed only of glycosyl residues with DP higher than 10, and iii) glycoconjugates, when the carbohydrate moiety is linked to a non-carbohydrate moiety (Cummings and Stephen, 2007). These structural diverse glycosides, thus, constitute the largest repository of metabolically accessible carbon and present a ubiquitous energy source for bacteria in the gut, terrestrial and marine environments (Grondin et al., 2017; Hemsworth et al., 2016).

Members of the microbiota have evolved diverse metabolic lifestyles and strategies to harvest energy from glycosides (Ndeh and Gilbert, 2018). According to the current paradigm, utilization of glycosides in the Bacteroidetes, a dominant phylum of the mammalian gut microbiota, is usually orchestrated by polysaccharide utilization loci (PULs), encoding carbohydrate-active enzymes (CAZymes) associated with SusCD-like transport systems, and sometimes also transporters of the ATP-binding cassette (ABC) or Major Facilitator Superfamily (MFS) families (Terrapon et al., 2018). In another dominant phylum of the mammalian gut microbiota, the Firmicutes, glycosides are metabolized by systems called gpPULs (Gram-positive PULs), which are analogous to PUL systems, encoding not TonB dependent receptors but a range of other transporters, including ABC, MFS and phosphotransferase systems (PTS) (Sheridan et al., 2016). Polysaccharides, depending on their DPs, are partially degraded by cell surface associated CAZymes to oligomers, which are then imported into the cell periplasm (for Gram-negative bacteria) or cell cytoplasm (for Gram-positive) through transporters. In the cell periplasm or the cytoplasm, oligosaccharides are processed by intracellular CAZymes to monosaccharides that are ultimately assimilated into the central metabolism of the bacteria. Therefore, glycoside-transporters are critical for glycoside utilization by bacteria, and thus linked to their ability to settle in nutritional niches (Grondin et al., 2017).

Recently, studies show that in the highly diffusive marine environment, SusCD-like transporters could facilitate some marine bacteria to successfully compete for labile carbon sources among the bacterial communities without diffusive loss (Kappelman et al., 2019; Unfried et al., 2018). Similarly, ABC transporters are likely to confer *Bifidobacteria* an important advantage to capture oligosaccharides via their extracellular lipid anchored solute-binding proteins (SBPs) in the competitive human gut niche (Theilmann et al., 2019). Furthermore, the occurrence of glycoside specific PTS or MFS transporters in pathogenic bacteria, including *Streptococcus pneumoniae* TIGR4 (Iyer and Camilli, 2007) and *Escherichia coli* BEN2908 (Schouler et al., 2009), was found to be correlated with the colonization and virulence of pathogens in the gut. These data highlight that understanding bacterial glycoside transport is a key step for further health managements, through the rational design of prebiotics and symbiotic mixtures, and the development of therapeutic gut microbes to the microbiota functioning and thus, improve health and prevent disease (Marques et al., 2018). Besides, for biotechnological applications, a better knowledge of bacterial glycoside uptake will allow one to develop microbial cell factories and microbial consortia able to efficiently convert non-refined biomass to high-value products (Hara et al., 2017).

Despite the importance of bacterial glycoside transporters for health and biotechnology, the rate of functional characterization of transporters is much less than that of CAZymes (Lombard et al., 2014). One of the major challenges is to find the correct substrates from the extremely large diversity of oligosaccharides (estimated to reach a few thousands (Lap  bie et al., 2019)) for each transporter. For CAZymes, which are classified in families based on sequence and mechanistic similarities in the CAZy database, their function can be predicted based on homology with the numerous biochemically characterized members of the same family (Lombard et al., 2014). For transporters, the classification of membrane transport proteins based on Transporter Classification (TC) system, which is accessible in the Transporter Classification Database (TCDB), is helpful too, since it incorporates both functional and phylogenetic information (Saier et al., 2016). However, predicting substrate specificity of transporters is extremely difficult due to their low sequence similarity, and to the fact that the vast majority of annotated glycoside-specific transporters lack experimental validation (Elbourne et al., 2017; Genee et al., 2016). Indeed, the experimental investigation of the function of glycoside transporter faces several challenges. The usual method to identify transporter specificity is based on the monitoring of the accumulation of radiolabeled compounds in whole cells, membrane vesicles or proteoliposomes (Blanvillain et al., 2007;

Cao et al., 2011; McCoy et al., 2016). But these assays can easily fail when the heterologous expressed transmembrane transport proteins are not active due to targeting, insertion or folding issues, or when the transporters are insoluble or unstable in purification or reconstitution process (Majd et al., 2018). In addition, some substrates are not available in the radioactive form or are prohibitively expensive, thus preventing large-scale identification trials. The transporter can also be validated in native strains by transcriptional analysis after cultivating the target strains on polysaccharides (Martens et al., 2011), by complementation studies (Joglekar et al., 2018) or by analysis of the phenotype of knock-out mutants (Linke et al., 2013). Nevertheless, the success of these approaches is still limited because the cost of omics studies is prohibitive if one wants to screen the metabolization potential of a large panel of oligosaccharides of defined structures, and also because of the difficulty of genetic manipulation of some species and the fact that functional redundancy often occurs in bacteria (Jensen et al., 2002; Maharjan et al., 2007). In addition, most bacteria in natural ecosystems are uncultured (Steen et al., 2019). Thus there is unmet demand for the development of new methods to rapidly identify and validate glycoside-transporter functions.

Here, we present a functional metagenomic approach for the rapid identification of glycoside-transporters from uncultured bacteria, which do not require radiolabeled substrates. Then, the specificity of the transport systems was screened by growth assays in micro-plates of fosmid *E. coli* clones containing metagenomic PULs. These growth assays allowed us to rapidly identify several types of glycoside-transporters which are functional in *E. coli*, revealing novel functions of uncultured bacteria of various taxonomical origins.

## **Results**

### **Prediction of potential glycoside-transporters from metagenomic clones**

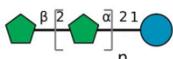
One hundred and two metagenomic clones isolated during previous functional screening campaigns for their ability to hydrolyze plant polysaccharides and oligosaccharides, and subsequently sequenced, have been selected to investigate their ability to transport oligosaccharides. Each of these metagenomic clones comprises within its 30 to 40 kb DNA insert at least one gene encoding a CAZyme responsible for the screened glycoside-hydrolase activity. Among the 102 clones, 26 produce at least one endo-acting glycoside-hydrolase activity. They were issued from a 156,000 clones metagenomic library constructed from the feces of a healthy adult who followed a fiber rich diet (F library), screened for the degradation of plant polysaccharides thanks to chromogenic assays (Tasse et al., 2010). Seventeen other

clones produce at least one glycosidase activity. There were identified from 20,000 clones randomly chosen among the F library, and from another 20,000 clones library constructed from the ileum mucosal microbiota (I library), screened for the ability to breakdown prebiotic oligosaccharides and polysaccharides (Cecchini et al., 2013). And finally, 59 clones were issued from an '*in vivo enriched*' (IVVE, 19,968 clones) and an '*in vitro enriched*' (IVTE, 20,352 clones) cow rumen metagenomic libraries, screened for endo-acting glycoside-hydrolases acting on plant cell wall polysaccharides (Ufarté et al., in preparation).

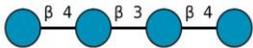
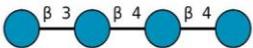
Based on the functional annotation of the metagenomic DNA insert, among the 102 clones mentioned above, we found 45 clones containing one or two predicted glycoside transport systems (**Table 1**) from the SusCD, MFS, ABC, and/or PTS families. The substrates chosen to test the transport ability of these 45 clones correspond to the oligosaccharides used to screen glycosidase activities by using positive selection on agar plates (Cecchini et al., 2013) and to the commercially available oligosaccharides which are components of the polysaccharides used to screen for endo-acting glycoside-hydrolase activities on agar plates (Tasse et al., 2010; Ufarté et al., in preparation). However, the enzymatic activity of the fosmid clones might be due to CAZyme encoding genes which are not part of the same operons (while being present in the same large metagenomic fragment) as the transporter encoding ones, since the boundary of PULs cannot be clearly defined without experimental validation (Moreno-Hagelsieb, 2015). We thus used the list of known activities for each CAZy family, retrieved from the CAZy database (Lombard et al., 2014) (**Table S1**), to enlarge the panel of substrates tested for growth ability, based on all the putative activities of the glycoside hydrolase enzyme (GHs) present in the metagenomic sequences (clones labeled with the asterisk in **Table 1**).

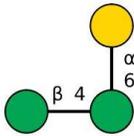
**Table 1. Predicted substrate specificity of glycoside transporters of metagenomic clones, and corresponding transporters and CAZyme families identified in the clone sequences.**

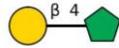
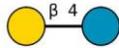
The clone activity corresponding to the GH activities was screened and validated in previous studies (Cecchini et al., 2013; Tasse et al., 2010; Ufarté et al., in preparation). The GH families that could be involved in the hydrolysis of the predicted transporter substrates are mentioned in bold type. For clones with the asterisk, transporter specificity was predicted based on the potential activities of the GHs identified from the metagenomic sequence of the clone. The glycosyl units are represented using the symbol nomenclature for glycans (Varki et al., 2009): glucose in blue circle, galactose in yellow circle, mannose in green circle, fructose in green pentagon and xylose in orange star.

| Predicted substrate for the transporter  | Clone | Clone activity         | Screening substrate | Transport system | Glycoside hydrolase enzyme  | Library |
|--|-------|------------------------|---------------------|------------------|---|---------|
| Fructooligosaccharides<br><br>$(0 \leq n \leq 3)$ | F1    | $\beta$ -fructosidase  | FOS                 | PTS              | <b>GH32</b>   | F       |
|  | F2    | $\beta$ -fructosidase  | FOS                 | MFS              | <b>GH32</b>   |         |
|  | I9    | $\beta$ -fructosidase  | FOS                 | PTS, ABC         | <b>GH32</b>   | I       |
|  | I10   | $\beta$ -fructosidase  | FOS                 | ABC              | <b>GH32</b>   |         |
| Xylooligosaccharides<br><br>$(1 \leq n \leq 7)$ | F3    | $\beta$ -xylosidase    | XOS                 | ABC              | GH2, <b>GH3</b> , GH67  | F       |
|  | I7    | $\beta$ -xylosidase    | XOS                 | MFS              | <b>GH10</b> , <b>GH43</b> , GH115   | I       |
|  | I8    | $\beta$ -xylosidase    | XOS                 | MFS, SusC        | <b>GH10</b> , GH35, <b>GH43</b> , GH67, GH115                                 |         |
|  | 14    | $\beta$ -xylanase      | AZCL-xylan          | SusCD            | <b>GH5</b> , <b>GH10</b> , <b>GH43</b> , GH92                                 | F       |
|  | 15    | 1,4- $\beta$ -xylanase | AZCL-xylan          | SusD             | <b>GH5</b> , <b>GH8</b> , <b>GH10</b> , GH31, <b>GH43</b> , GH95, GH97, GH115 |         |
|  | 10N15 | 1,4- $\beta$ -xylanase | AZCL-xylan          | ABC              | <b>GH10</b> , GH94  | IVVE    |
|  | 16B10 | 1,4- $\beta$ -xylanase | AZCL-xylan          | ABC              | <b>GH5</b> , <b>GH10</b> , GH94   |         |

|  |        |                         |                             |             |                                       |      |
|--|--------|-------------------------|-----------------------------|-------------|---------------------------------------|------|
|  | 25I16  | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | ABC         | <b>GH10</b>                           |      |
|  | 30F10  | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | ABC         | <b>GH10</b>                           |      |
|  | 35C9   | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | ABC         | <b>GH5, GH10</b>                      |      |
|  | 36E1   | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | ABC         | <b>GH10</b>                           |      |
|  | 39F1   | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | ABC         | <b>GH1, GH10</b>                      |      |
|  | 42N18  | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | ABC         | <b>GH10</b>                           |      |
|  | 45P14  | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | ABC         | <b>GH10</b>                           |      |
|  | 50E19  | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | ABC         | <b>GH10</b>                           |      |
|  | 23O13  | 1,4- $\beta$ -xy lanase | AZCL-xy lan                 | SusCD       | <b>GH5, GH11, GH53</b>                | IVTE |
|  | 5*     | $\beta$ -glucanase      | AZCL- $\beta$ -glucan       | SusD, SusCD | <b>GH1, GH3, GH16, GH97</b>           |      |
|  | 9*     | $\beta$ -glucanase      | AZCL- $\beta$ -glucan       | SusD        | <b>GH3, GH7, GH16, GH78</b>           | F    |
|  | 10*    | $\beta$ -glucanase      | AZCL- $\beta$ -glucan       | SusCD       | GH2, <b>GH3, GH5, GH7, GH94, GH97</b> |      |
|  | 11*    | $\beta$ -glucanase      | AZCL- $\beta$ -glucan       | SusCD       | GH2, <b>GH3, GH5, GH7, GH94, GH97</b> |      |
|  | 7D9*   | $\beta$ -mannanase      | AZO-galactomannan           | ABC         | GH4, GH26, GH35, <b>GH43, GH63</b>    | IVVE |
|  | 52E22* | $\beta$ -mannanase      | AZO-galactomannan           | ABC         | <b>GH3, GH26, GH130</b>               |      |
|  | 3F21*  | $\beta$ -glucanase      | AZCL-Barley $\beta$ -glucan | SusCD       | <b>GH3</b>                            |      |
|  | 12E13* | $\beta$ -mannanase      | AZO-galactomannan           | SusCD       | GH26, <b>GH43, GH51</b>               | IVTE |
|  | 40D8*  | $\beta$ -glucanase      | AZCL-Barley $\beta$ -glucan | SusCD       | <b>GH3, GH13, GH16, GH92</b>          |      |
|  | 41J20* | $\beta$ -mannanase      | AZO-galactomannan           | MFS, SusC   | <b>GH3, GH26, GH130</b>               |      |

|   |        |                        |                             |             |  |      |
|---|--------|------------------------|-----------------------------|-------------|--|------|
|   | 41O1*  | $\beta$ -glucanase     | AZCL-Barley $\beta$ -glucan | SusCD       | <b>GH3</b> , GH16, GH78, GH106                         |      |
| Cellobiosyl-cellobiose<br> | 3      | $\beta$ -glucanase     | AZCL- $\beta$ -glucan       | MFS         | GH2, <b>GH5</b>  | F    |
|   | 5      | $\beta$ -glucanase     | AZCL- $\beta$ -glucan       | SusD, SusCD | <b>GH1</b> , <b>GH3</b> , <b>GH16</b> , GH97           |      |
|   | 9      | $\beta$ -glucanase     | AZCL- $\beta$ -glucan       | SusD        | <b>GH3</b> , <b>GH7</b> , <b>GH16</b> , GH78           |      |
|   | 10     | $\beta$ -glucanase     | AZCL- $\beta$ -glucan       | SusCD       | GH2, <b>GH3</b> , <b>GH5</b> , <b>GH7</b> , GH94, GH97 |      |
|   | 11     | $\beta$ -glucanase     | AZCL- $\beta$ -glucan       | SusCD       | GH2, <b>GH3</b> , <b>GH5</b> , <b>GH7</b> , GH94, GH97 |      |
| Glucosyl-cellotriose<br>   | 3F21   | $\beta$ -glucanase     | AZCL-Barley $\beta$ -glucan | SusCD       | <b>GH3</b>   | IVTE |
|   | 10D6   | $\beta$ -glucanase     | AZCL-CM cellulose           | SusCD       | GH2, <b>GH5</b> , <b>GH26</b> , GH94                   |      |
|   | 15K21  | $\beta$ -glucanase     | AZCL-CM cellulose           | SusCD       | <b>GH5</b> , <b>GH26</b> , GH130                       |      |
|   | 38A17  | $\beta$ -glucanase     | AZCL-CM cellulose           | SusCD       | <b>GH5</b> , <b>GH26</b> , GH130                       |      |
|   | 40D8   | $\beta$ -glucanase     | AZCL-Barley $\beta$ -glucan | SusCD       | <b>GH3</b> , GH13, <b>GH16</b> , GH92                  |      |
| Cellotriose<br>          | 41O1   | $\beta$ -glucanase     | AZCL-Barley $\beta$ -glucan | SusCD       | <b>GH3</b> , <b>GH16</b> , GH78, GH106                 | F    |
|   | 44D8   | $\beta$ -glucanase     | AZCL-CM cellulose           | SusCD       | <b>GH5</b> , <b>GH26</b> , GH130                       |      |
|   | 44F17  | $\beta$ -glucanase     | AZCL-CM cellulose           | SusCD       | <b>GH5</b> , <b>GH26</b> , GH130                       |      |
|   | 14*    | 1,4- $\beta$ -xylanase | AZCL-xylan                  | SusCD       | <b>GH5</b> , GH10, GH43, GH92                          |      |
|   | 15*    | 1,4- $\beta$ -xylanase | AZCL-xylan                  | SusD        | <u>GH5</u> , GH8, GH10, GH31, GH43, GH95, GH97, GH115  |      |
| IVVE  | 1C10*  | $\beta$ -mannanase     | AZO-galactomannan           | ABC         | GH4, <u>GH26</u> , GH28, GH39, GH43, GH63              |      |
|   | 7D9*   | $\beta$ -mannanase     | AZO-galactomannan           | ABC         | GH4, <b>GH26</b> , GH35, GH43, GH63                    |      |
|   | 14O19* | $\beta$ -mannanase     | AZO-galactomannan           | ABC         | GH4, <b>GH26</b> , GH28, GH63                          |      |

|  |        |                        |                   |           |  |      |
|--|--------|------------------------|-------------------|-----------|--|------|
|  | 52E22* | $\beta$ -mannanase     | AZO-galactomannan | ABC       | <b>GH3, GH26, GH130</b>                  | IVTE |
|  | 12E13* | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH26, GH43, GH51</b>                  |      |
|  | 20G16* | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH5, GH130</b>                        |      |
|  | 23O13* | 1,4- $\beta$ -xylanase | AZCL-xylan        | SusCD     | <b>GH5, GH11, GH53</b>                   |      |
|  | 28D5*  | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH5, GH26, GH92, GH130</b>            |      |
|  | 41J20* | $\beta$ -mannanase     | AZO-galactomannan | MFS, SusC | <b>GH3, GH26, GH130</b>                  |      |
| Galactosyl-mannobiose<br><br><br><br>Mannotriose<br><br> | 1C10   | $\beta$ -mannanase     | AZO-galactomannan | ABC       | <b>GH4, GH26, GH28, GH39, GH43, GH63</b> | IVVE |
|  | 7D9    | $\beta$ -mannanase     | AZO-galactomannan | ABC       | <b>GH4, GH26, GH43, GH63</b>             |      |
|  | 14O19  | $\beta$ -mannanase     | AZO-galactomannan | ABC       | <b>GH4, GH26, GH28, GH63</b>             |      |
|  | 52E22  | $\beta$ -mannanase     | AZO-galactomannan | ABC       | <b>GH3, GH26, GH130</b>                  | IVTE |
|  | 10D6   | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH2, GH5, GH26, GH94</b>              |      |
|  | 12E13  | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH26, GH43, GH51</b>                  |      |
|  | 15K21  | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH5, GH26, GH130</b>                  |      |
|  | 20G16  | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH5, GH130</b>                        |      |
|  | 28D5   | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH5, GH26, GH92, GH130</b>            |      |
|  | 38A17  | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH5, GH26, GH130</b>                  |      |
|  | 41J20  | $\beta$ -mannanase     | AZO-galactomannan | MFS, SusC | <b>GH3, GH26, GH130</b>                  |      |
|  | 44D8   | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH5, GH26, GH130</b>                  |      |
|  | 44F17  | $\beta$ -mannanase     | AZO-galactomannan | SusCD     | <b>GH5, GH26, GH130</b>                  |      |

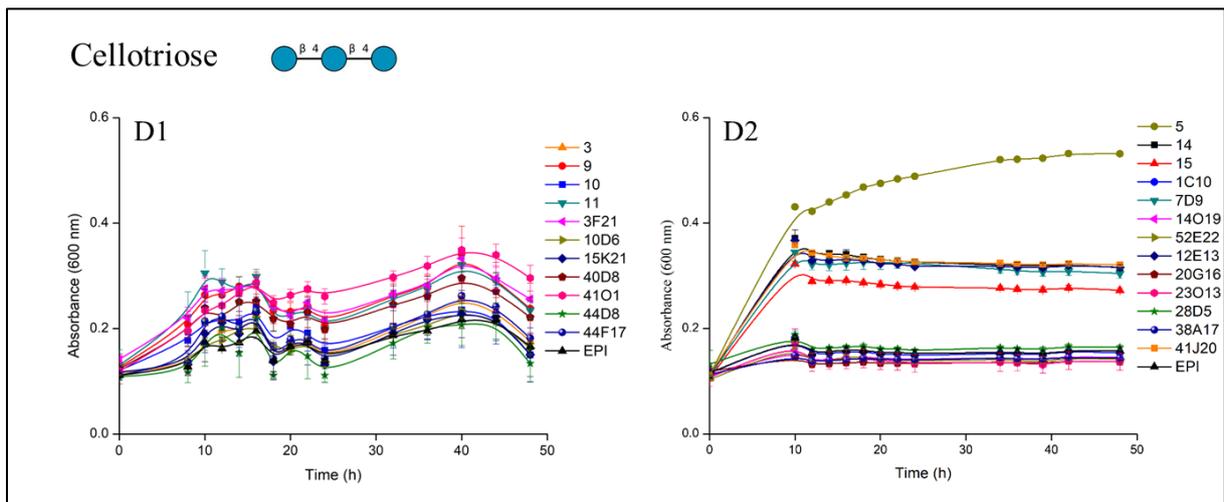
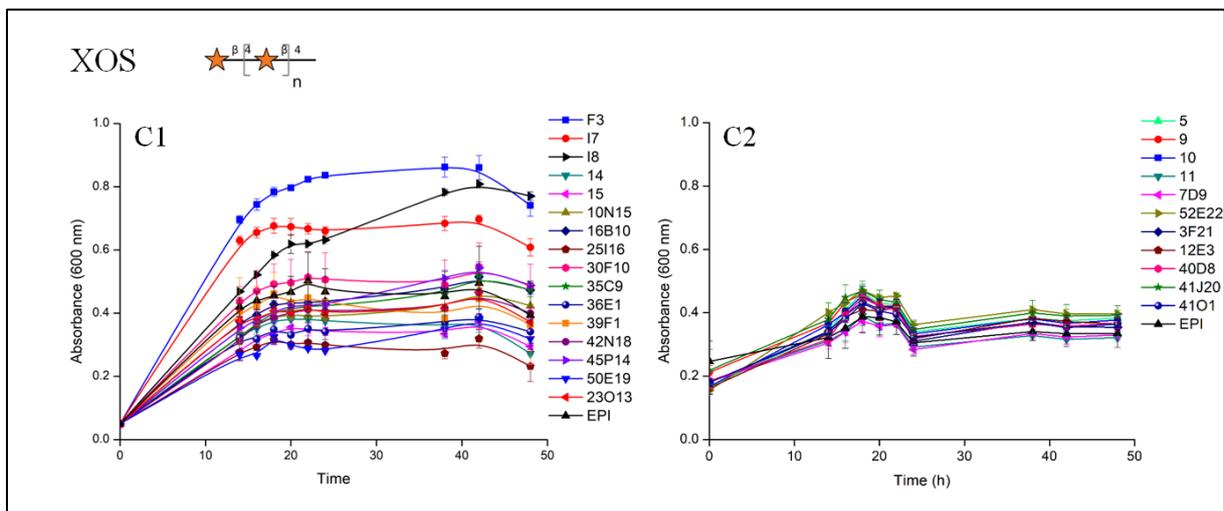
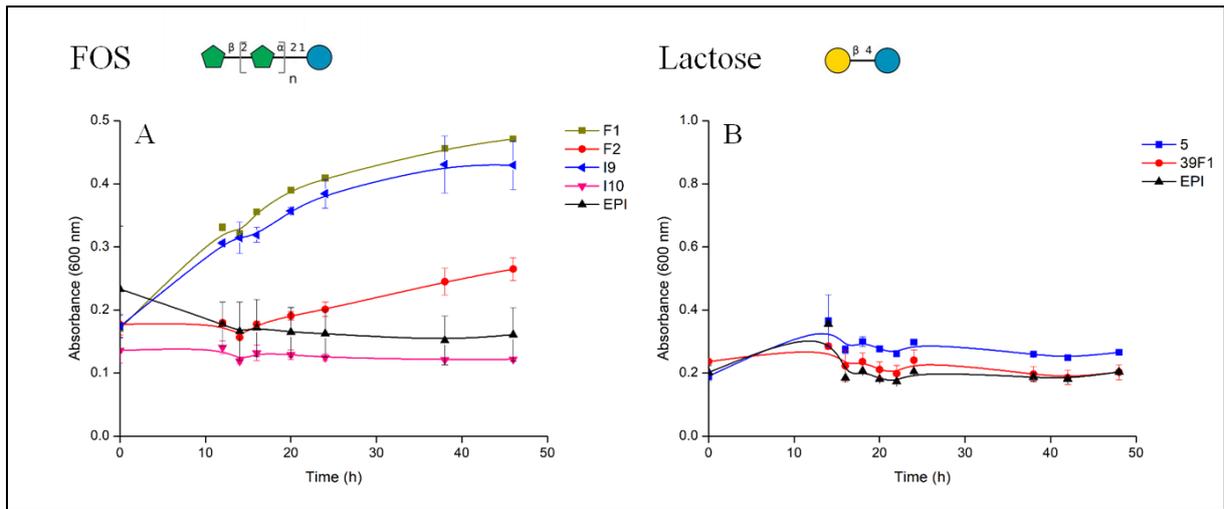
|  |        |                        |                       |              |  |      |
|--|--------|------------------------|-----------------------|--------------|--|------|
|  | 3*     | $\beta$ -glucanase     | AZCL- $\beta$ -glucan | MFS          | <b>GH2, GH5</b>                                      | F    |
|  | 5*     | $\beta$ -glucanase     | AZCL- $\beta$ -glucan | SusD, SusCD  | <b>GH1, GH3, GH16, GH97</b>                          |      |
|  | 10*    | $\beta$ -glucanase     | AZCL- $\beta$ -glucan | SusCD        | <b>GH2, GH3, GH5, GH7, GH94, GH97</b>                |      |
|  | 11*    | $\beta$ -glucanase     | AZCL- $\beta$ -glucan | SusCD        | <b>GH2, GH3, GH5, GH7, GH94, GH97</b>                |      |
|  | 14*    | 1,4- $\beta$ -xylanase | AZCL-xylan            | SusCD        | <b>GH5, GH10, GH43, GH92</b>                         |      |
|  | 15*    | 1,4- $\beta$ -xylanase | AZCL-xylan            | SusD         | <b>GH5, GH8, GH10, GH31, GH43, GH95, GH97, GH115</b> |      |
|  | 16B10* | 1,4- $\beta$ -xylanase | AZCL-xylan            | ABC          | <b>GH5, GH10, GH94</b>                               | IVVE |
|  | 35C9*  | 1,4- $\beta$ -xylanase | AZCL-xylan            | ABC          | <b>GH5, GH10</b>                                     |      |
|  | 39F1*  | 1,4- $\beta$ -xylanase | AZCL-xylan            | ABC          | <b>GH1, GH10</b>                                     |      |
|  | 23O13* | 1,4- $\beta$ -xylanase | AZCL-xylan            | SusCD        | <b>GH5, GH11, GH53</b>                               |      |
| Lactulose<br> | I11    | $\beta$ -galactosidase | Lactulose             | MFS/PTS      | <b>GH2</b>   | I    |
|  | I12    | $\beta$ -galactosidase | Lactulose             | ABC          | <b>GH2, GH42</b>                                     |      |
|  | I13    | $\beta$ -galactosidase | Lactulose             | PTS          | <b>GH2, GH77</b>                                     |      |
|  | I14    | $\beta$ -galactosidase | Lactulose             | ABC, MFS/PTS | <b>GH2, GH13</b>                                     |      |
| Lactose<br> | 5*     | $\beta$ -glucanase     | AZCL- $\beta$ -glucan | SusD, SusCD  | <b>GH1, GH3, GH16, GH97</b>                          | F    |
|  | 39F1*  | 1,4- $\beta$ -xylanase | AZCL-xylan            | ABC          | <b>GH1, GH10</b>                                     | IVVE |

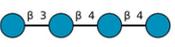
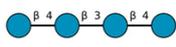
## Screening glycoside metabolizing pathways

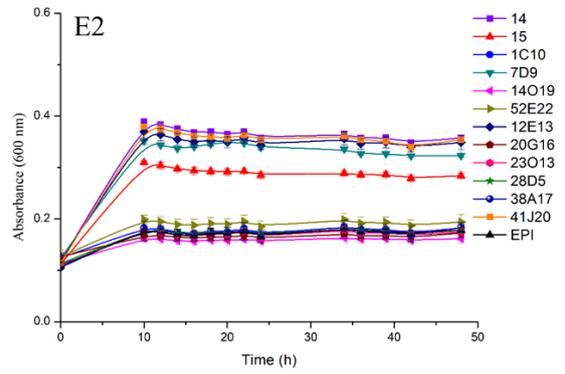
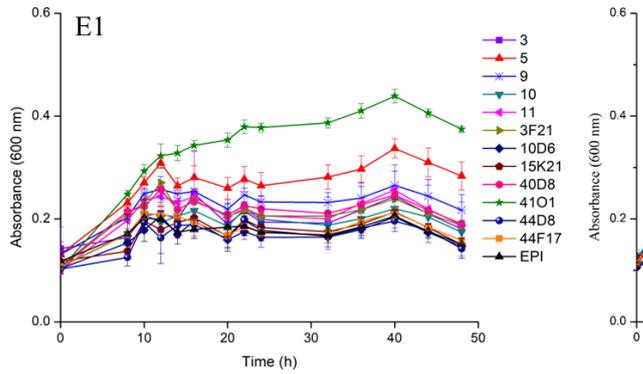
The utilization of glycosides usually needs the cooperation of transporters and CAZymes. Previously, we proved that the metagenomic gene expression in fosmid is not under the control of promoters flanking the cloning site, but rather under the control of sequences randomly scattered on the metagenomic DNA insert, which are recognized as promoters by *E. coli* (Tauzin et al., 2016). Therefore, to assess the efficiency of cloning PULs in fosmid to (randomly) co-express at least two genes (encoding a transporter and a glycosidase), and to validate the specificity prediction of transporters listed in **Table 1**, growth screening of the selected 45 metagenomic clones on the predicted oligosaccharides was performed in micro-plates. Their growth was characterized by monitoring the value of optical density at 600 nm ( $OD_{600\text{ nm}}$ ) in M9 medium supplemented with selected oligosaccharides as sole carbon source within 48 h (**Figure 1**). The *E. coli* host transformed with an empty fosmid (named EPI) was used as a negative control. The increasing value of  $OD_{600\text{ nm}}$  over time compared with that of EPI provides direct evidence of the glycoside utilization by the tested clone.

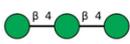
In total, 129 growth assays were carried out for the 45 selected metagenomic clones and allowed the identification of 13 hit clones with a value of  $OD_{600\text{ nm}}$  value higher by 0.2 than that of the negative control EPI. Such an increase is significant, since the  $OD_{600\text{ nm}}$  value of the EPI strain increases by 0.3 after a 48 h growth on 0.5% (m/v) glucose. The mean total hit yield was 77.1% when the primary screening of the GH activity was based on positive selection on oligosaccharides as sole carbon sources, and 12.3% when it was based on chromogenic assays screening on polysaccharides (**Table 2**). Based on sequence analyses, for each of the 13 hit clones, the size of the contigs exhibits a wide range from, 13,000 to 43,000 bp, and four clones (I9, 41O1, 35C9 and 41J20) contain two contigs for one single metagenomic DNA insert (**Table 3**). The clone sequences are mainly assigned to Firmicutes and Bacteroidetes, consisting of the two dominant phyla in the mammalian gut microbiota. The hit clones possess one or two transport systems including ABC, MFS, PTS and/or SusCD, and from one to five glycoside hydrolases, exhibiting various abilities to metabolize FOS, XOS, lactulose, mixed-linkage  $\beta$ -glucans or galactosyl-mannobiose and mannotriose. The clones harboring a GH1 were tested on lactose, but they were not able to grow on this substrate. Twelve of the 13 hit clones could grow on the same oligosaccharides that had been experimentally tested during the initial positive selection on solid medium (nine clones: F1, F3, I7, I8, I9, I11, I12, I13 and I14) or on oligosaccharides derived from the polysaccharide used for chromogenic assays (three clones: 5, 41O1 and 41J20). Only one hit clone (35C9,

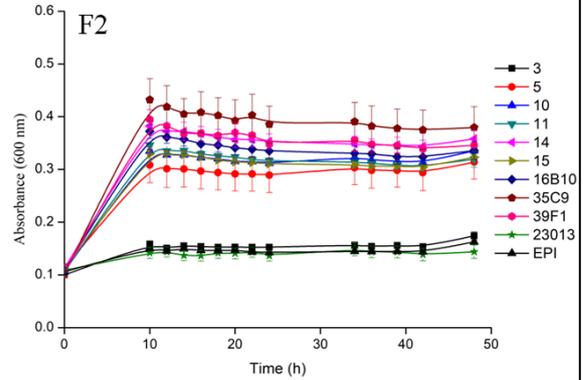
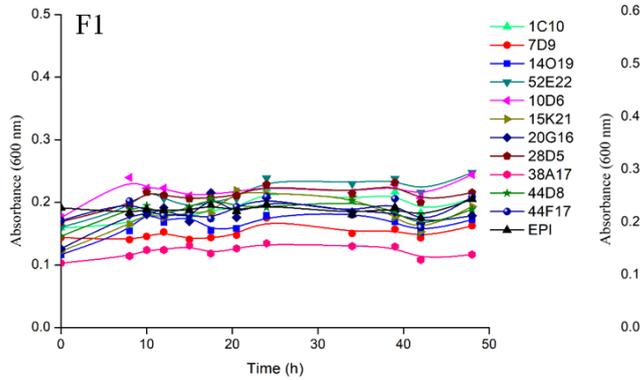
which was able to grow on galactosyl-mannobiose and mannotriose) was obtained from the 55 growth assays involving 22 clones selected because their sequence contains at least one GH enzyme with putative glycosidase activity on the targeted oligosaccharides.

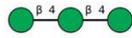


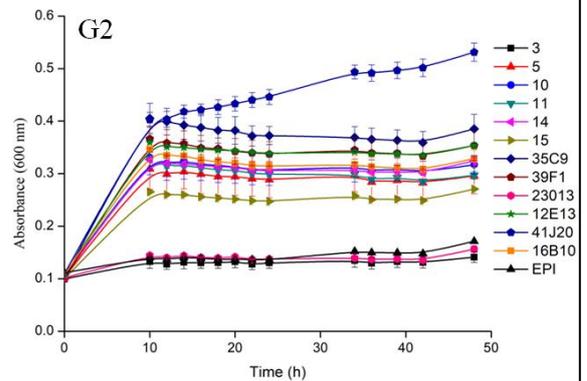
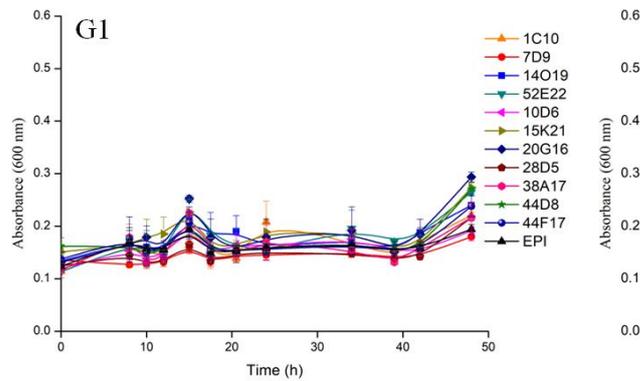
Mixture of cellobiosyl-cellobiose  and glucosyl-celotriose 

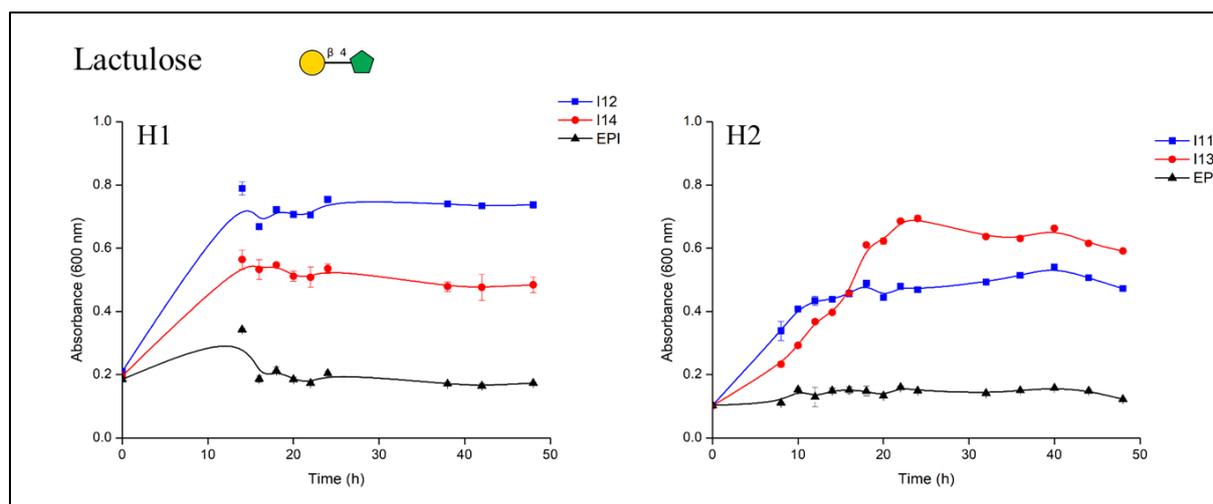


Mannotriose 



Mixture of galactosyl-mannobiose  and mannotriose 





Glycosyl unit legends: ● Glucose; ⬠ Fructose; ★ Arabinose; ● Galactose; ● Mannose

**Figure 1. Growth of metagenomic clones on various oligosaccharides used as sole carbon sources:** (A) FOS, (B) lactulose, (C) XOS, (D) cellotriose, (E) mixture of cellobiosyl-cellobiose and glucosyl-cellotriose, (F) mannotriose, (G) mixture of galactosyl-mannobiose and mannotriose, and (H) lactulose.

**Table 2. Number of hit clones obtained from different libraries and the hit yield in the growth screening.**

| Method for primary screening of GH activity on solid plates | Positive selection on oligosaccharides |      | Chromogenic assays on polysaccharides |      |      |
|---|--|------|---------------------------------------|------|------|
|   | F                                      | I    | F                                     | IVVE | IVTE |
| Library   |  |      |                                       |      |      |
| Number of hit clones/number of tested clones                | 2/3                                    | 7/8  | 1/7                                   | 1/14 | 2/13 |
| Hit yield (%)   | 66.7                                   | 87.5 | 14.3                                  | 7.1  | 15.4 |
| Mean hit yield (%)  | 77.1                                   |      | 12.3                                  |      |      |

**Table 3. Summary of hit clones with the substrate used for the growth screening, their contig size and taxonomical assignation at the level of phylum, order or species, depending on the similarity of their sequence with referenced genomes.**

The fact that the clone 41J20 possesses two non-assembled contigs assigned to different phyla indicates that loci were exchanged by horizontal gene transfers between distantly related gut bacteria, as previously described (Tasse et al., 2010).

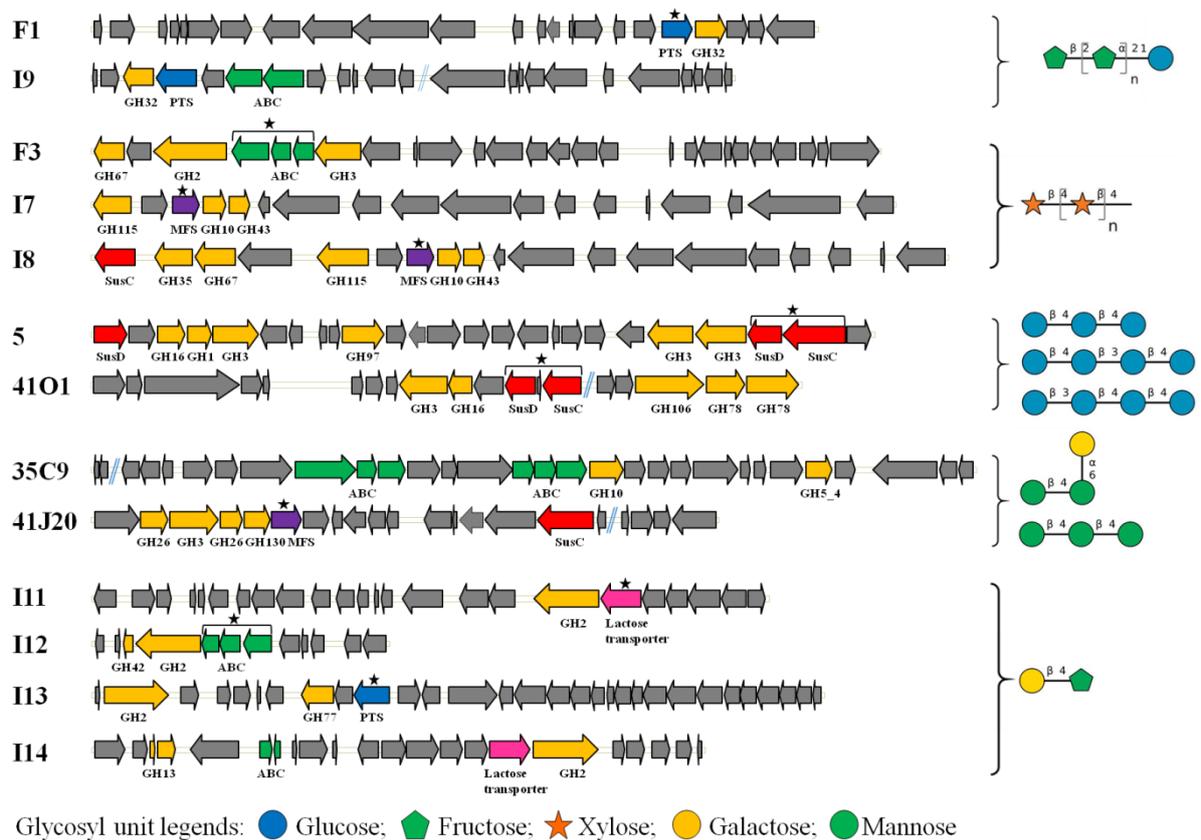
| Substrate   | Hit clone | Contig size (bp) | Taxonomic assignation |  |
|---|-----------|------------------|-----------------------|--|
|   |           |                  | Phylum                | Order/Species                                      |
| FOS   | F1        | 33,125           | Firmicutes            | Clostridiales                                      |
|   | I9        | 14,714           | Firmicutes            | Clostridiales/ <i>Dorea longicatena</i>            |
|   |           | 13,700           | Firmicutes            | Clostridiales/ <i>Clostridium nexile</i>           |
| XOS   | F3        | 35,645           | Firmicutes            | Clostridiales                                      |
|   | I7        | 37,812           | Bacteroidetes         | Bacteroidales                                      |
|   | I8        | 38,518           | Bacteroidetes         | Bacteroidales                                      |
| Mixture of cellobiosyl-cellobiose and glucosyl-celotriose, and celotriose | 5         | 42,497           | Bacteroidetes         | Bacteroidales                                      |
|   | 41O1      | 22,594           | Bacteroidetes         | Bacteroidales                                      |
|   |           | 1,024            | Bacteroidetes         | Bacteroidales                                      |
| Galactosyl-mannobiose and mannotriose                                     | 35C9      | 2,381            | -                     | -  |
|   |           | 41,729           | -                     | -  |
|   | 41J20     | 23,867           | Bacteroidetes         | Bacteroidales/ <i>Bacteroides graminisolvens</i>   |
|   |           | 4,322            | Firmicutes            | Clostridiales                                      |
| Lactulose   | I11       | 30,347           | Firmicutes            | Lactobacillales/ <i>Streptococcus thermophilus</i> |
|   | I12       | 13,317           | Firmicutes            | Clostridiales                                      |
|   | I13       | 32,036           | Firmicutes            | Clostridiales/ <i>Faecalibacterium prausnitzii</i> |
|   | I14       | 30,347           | Firmicutes            | Lactobacillales/ <i>Streptococcus thermophilus</i> |
| Lactose   | -         | -                | -                     | -  |

## Identification of functional glycoside-transporters

In the metagenomic clones, it is likely that the CAZymes could not be secreted or attached at the cell surface of the recombinant host (Tasse et al., 2010), because of the very low secretion potential of *E. coli* (Kleiner-Grote et al., 2018). The growth of hit clones thus requires functional transporters to import oligosaccharides into cell for further degradation by CAZymes. Therefore, the strategy consisting in screening clones encoding both transporter and CAZymes by growth assay provides direct experimental evidence of glycoside transporter functionality.

The sequences of the 13 hit clones presented in **Figure 2**, include 19 transport systems belonging to the ABC, PTS, MFS and/or SusCD families. For seven clones (F1, F3, I7, 41O1, I11, I12 and I13), there is only one putative glycoside transporter in each clone (**Figure 2**), leading to a straight forward identification of the likely specificity of the transporter (given that some genes of unknown function are present in the contigs). Among these seven clones, we identified a PTS transporter for FOS (clone F1), an ABC transport system and a MFS transporter for XOS (clones F3 and I7, respectively), and a SusCD transport system for mixed-linkage  $\beta$ -glucopolysaccharides (clone 41O1). For clones able to grow on lactulose, an ABC transport system and a PTS transporter were identified in clones I12 and I13, respectively, while the clone I11 harbors a putative protein named 'lactose transporter'. This putative protein includes both a PTS\_IIA domain and a MFS domain in the polypeptidic chain. For the other six clones (I8, I9, 5, 35C9, 41J20 and I14), two transport systems were annotated within the same metagenomic insert (**Figure 2**). These two transport systems might be working in synergy, completely independently or share redundant function. It is difficult to determine which of the two detected transporters was responsible for the substrate internalization in the growth assays based only on the organization of genes encoding transporter and CAZymes. It is indeed impossible to define the boundaries of PULs without any experimental evidence of the presence of an operon in the native strain from which are issued the DNA metagenomic fragments, and because PULs could have been truncated during the construction of the metagenomic library. In particular, the clone 5 harbors two SusCD transport systems which could belong to two different PULs based on their distance and genomic environment. One of the PULs lacks the SusC protein, likely due to the truncation during the metagenomic library construction. Thus, the functional transporter in clone 5 can only be the entire SusC/D system. In the native strain, it is also possible that the two PULs work in synergy, as demonstrated for several PULs such as the pectin, xylan and N-glycan

PULs (Cuskin et al., 2015; Luis et al., 2018; Rogowski et al., 2015). In the clone I9, it is impossible at this stage to know which of the PTS or of the ABC transport system is responsible for the transport of FOS. The clone I7 is partially redundant with the clone I8 and share 99% nucleotide identity from the GH115 to the GH43, which covered only one transport system. The clone I8 has indeed two transport systems, a MFS and a SusC, while the clone I7 harbors only the MFS transporter. The absence of SusD in the clone I8, which could be due to the truncation of the PUL during the construction of the metagenomic library, suggests that the SusC cannot import any glycosides (Koropatkin et al., 2008; Tamura et al., 2019; Tauzin et al., 2016). Although further investigations are required to confirm this hypothesis, this suggests that the MFS transporter is more likely responsible for the uptake of XOS than the incomplete SusCD transport system in the clone I8. Similarly, the 41J20 sequence contains a MFS transporter and an incomplete SusCD transport system lacking the SusD, which is also not surrounded by GHs. Additionally, the RPS-BLAST analysis against the NCBI's Conserved Domain Database (CDD) (Marchler-Bauer et al., 2017) indicated that the SusC protein of clone 41J20 might mainly transport inorganic ion, strengthening the prediction that this SusC might not be involved in glycoside transport. The MFS transporter which clusterized with GHs in the 41J20 sequence is thus probably the only transporter involved in the uptake of (galacto-)mannooligosaccharides. In the clone I14, two transport systems are present, a 'lactose transporter' and an ABC transporter. The 'lactose transporter' encoding gene was clustered together with the GH2 one. The DNA segment including both the GH2 and the lactose transporter showed a perfect synteny with the corresponding part of the sequence I11, which contains only one transport system, showed here to be responsible for the use of lactulose. This implies that the 'lactose transporter' might be responsible for import of lactulose by the clone I14 too. Nevertheless, even though the I14 ABC transport system is far from the 'lactose transporter' and the GH2, we cannot exclude, at this stage, whether the ABC was involved in lactulose utilization too. In the clone 35C9, while one ABC transporter is close to a GH10, another ABC transporter is alone without any flanking GHs. A GH5\_4 is located far downstream without any transporter around. Based on the known activities of GH5\_4 and GH10 members, the 35C9 (which is mandatory to allow monosaccharide release, and then, *E. coli* cell growth) towards (galacto)manno-oligosaccharides seems rather due to the GH5\_4 than to an endo-acting  $\beta$ -glycanase activity of the GH10 (Pollet et al., 2010). Thus, we cannot discriminate between both ABC transporters, which one is responsible for the transport of galactosyl-mannobiose and mannotriose by the clone 35C9.



**Figure 2. Gene organization of hit clones that are able to grow on different glycosides.**

Black stars indicate the transporters predicted to be responsible for the clone phenotypes, because they are the only glycoside transporter systems detected in the metagenomic PULs or because the SusD component of the SusC/D system is missing. The two contigs of clones I9, 41O1, 35C9 and 41J20 are separated with blue double lines. Genes are represented by arrows and colored based on their function. Genes encoding GHs in yellow, glycoside transporters in blue for PTS, green for ABC, red for SusCD, purple for MFS and pink for ‘lactose transporter’, and other functions in grey.

To further screen the specificity of the identified transporter, the clone F3 was tested on aldouronic acids, which corresponds to branched XOS with aldouronic residues (**Figure 3a**), since the SBP of a XOS ABC system from *Caldanaerobius polysaccharolyticus* ATCC BAA-17 also was proved to bind aldouronic acids (Han et al., 2012). It can be seen in **Figure 3a** that the clone F3 could grow on aldouronic acids, suggesting the ABC transporter displayed specificity for aldouronic acids, despite the low identity (23.8%) between the SBP in F3 and the SBP from *C. polysaccharolyticus* ATCC BAA-17. Moreover, either the clone F3 or strain



used (Ndeh and Gilbert, 2018). Here, basing our study on the PUL paradigm, we screened the specificity of glycoside-transporters from uncultured bacteria, of which the sequences were previously identified by activity-based functional metagenomics targeting GH activities. We identified ten metagenomic PULs (three being partially or fully redundant) of which the glycoside transporter components are functional in *E. coli*.

In total, 19 transporters of four types (PTS, ABC, MFS and SusCD) were highlighted from only 129 growth assays, showing the efficiency of this strategy to identify glycoside-transporters of various specificities, structures and transport mechanisms. We extended the panel of the biochemically characterized functions for the MFS family, of which the member of clone 41I20 was showed here to transport (galacto-)mannooligosaccharides.

The present method is highly reproducible, quantitative, and can be performed at a medium-throughput, to screen about fifty clones in two days by automatically monitoring the growth with a thermostated micro-plate reader. This assay requires less than 10 mg of oligosaccharides for each clone, tested in triplicates. It does not require the use of chromogenic, fluorogenic or radio-labeled substrates. It thus allows one to easily screen the transporter specificity towards the natural oligosaccharides that bacteria have to uptake in Nature, avoiding missing positive clones and getting false positives with chemically modified substrates, of which the structure differs from that of natural compounds. The screening on oligosaccharides also allows a direct access to the function of transporters, contrarily to transcriptomic studies which are based on the use of the polymeric substrate before any enzymatic degradation to trigger PUL expression. It also requires 10 times less substrate amounts than the growth assays required for transcriptomic analyses (Martens et al., 2011), which necessitate retrieving sufficient cells for RNA extraction and sequencing. And of course, our approach allows the screening of the specificity of transporters from uncultured bacteria.

Nevertheless, there are some limitations in screening the transporter specificity of transporters in recombinant *E. coli* cells transformed with PUL-containing fosmids. First, the yield of hit clones, producing both a functional glycosidase and a functional transporter, is low. As shown in this study, depending on the approach used in the primary screen targeting GH activity, the hit yield varies between 12 and 77%. Of course, the highest hit yield was obtained when the primary screen was a solid-plate assay based on positive selection on oligosaccharides used as sole carbon sources. Even if the screening principles seems identical in solid medium and in liquid medium, it is in fact slightly different. In solid plate screening campaigns, automates

grid several cells of the same clone at a same location on the agar plate. For the GH producing hits, when one of these cells lyses, it releases nearby hydrolytic activity. Monosaccharides are then produced into the immediate environment of the other cells of the clonal population, allowing them to grow in turn, and therefore, to easily visualize cell growth. This principle explains why the hit yield is not 100% after the secondary screening for the transport functions in liquid medium. The screening in liquid medium could be carried out directly in one stage, without primary screening for GHs functions. This would nevertheless require ten times more substrate for each clone (without replicates), and the maximum hit yield would drop to 0.7%. This value corresponds to the maximum mean probability (5.5‰) of isolating a clone producing an active GH from a library of fosmid clone (8‰ for endo-acting GHs (Tasse et al., 2010) and 3‰ for glycosidases (Ufarté et al., in preparation) multiplied by that of isolating a functional transporter from GH producing *E. coli* fosmid clones (12%, as shown in the present study). Also, the probability to produce both a functional glycosidase and a functional transporter from a (meta)genomic PUL cloned into a fosmid is only 1.8%, since one hit was obtained from the 55 assays performed based on sequence analysis. This low hit yield is due to the fact that in fosmids, gene expression is spurious, being due to the random presence of sigma factor sequences along the fosmid DNA insert, which are recognized by *E. coli* to trigger gene expression (Tauzin et al., 2016).

Screening the transporter function by expressing the genes of PULs in *E. coli* fosmid clones is thus a strategy which requires automated screening facilities. It is also a costly approach, but of which the price could be decreased by  $10^6$  by using droplet-microfluidics, as we recently showed (Dagkesamanskaya et al., 2018). Finally, by using this technology, the transporter function can be proved only after truncation of the PUL, in cases where there are several transporter sequences identified in the cloned DNA fragment, and also in order to exclude a potential role of the unknown proteins of which the function cannot be predicted in the detected phenotype.

### **MFS, ABC, PTS and SusCD transporters from various Gram-positive and Gram- negative uncultured bacteria are functional in the *E. coli* membrane**

Except for the clone 35C9, of which the sequence is not sufficiently similar to that of a sequenced genome to predict its taxonomical origin, all the metagenomic DNA sequences of the hit clones were assigned to the two main phyla in mammalian gut microbiomes, Bacteroidetes and Firmicutes. However, more than 50% of the hit sequences couldn't be

annotated to the species level, highlighting the functional diversity that remains to be discovered from gut microbiomes by using metagenomics.

Bacteroidetes transporters from both the MFS and SusCD families were highlighted in the present study. We previously demonstrated that these two types of membrane proteins can be functional when produced in recombinant *E. coli* cells, of which the membrane structure is similar to that of Bacteroidetes, both being Gram negative bacteria (Tauzin et al., 2016). No PTS nor ABC systems were identified in these Bacteroidetes sequences. It is logical for PTS, which have, to our knowledge, never been reported in Bacteroidetes, contrary to MFS and ABC (Tauzin et al., 2016; Zafar and Saier, 2018) which are described in this phylum, although no Bacteroidetes ABC transporter has been shown to be involved in carbohydrate transport until now. A larger screening study could thus, in the future, allow us to identify and characterize Bacteroidetes ABC transporters.

From Firmicutes bacteria, we identified a larger diversity of transporters types, since three PTS, four ABC and two lactose transporters which include one PTS\_IIA domain and one MFS domain were highlighted in the Firmicutes' PULs. MFS, ABC and PTS systems are all used by native *E. coli* strains for nutrient uptake (Saier, 2000). In addition, several PTS involved in oligosaccharide utilization by Gram-positive bacteria were previously successfully expressed in *E. coli* (Lai et al., 1997; Lai and Ingram, 1993), despite the difference in cell membrane organization. The localization in *E. coli* of the different components of the transporters identified here from Gram positive bacteria will have to be studied.

### **Non-canonical ABC transporters are involved in oligosaccharide utilization by bacteria**

Among the identified ABC transport systems, there are two lacking a SBP in clone I9 and I14. Further characterization of this SBP deficient ABC transporter should shed new light on the mechanism of ABC transporters involved in carbohydrate utilization. As for the PTS transport systems, the F1 one lacks the PTS\_EIIA domain (the inner carbohydrate-specific domain through that the phosphate group transits to the PTS\_EIIB domain and, ultimately, to the oligosaccharide internalized through the transmembrane PTS\_EIIC component (Saier, 2015)). Moreover, the 'lactose transporter' identified in clones I11 and I14 for lactulose uptake are original proteins, as the 'lactose transporter' contains two domains (a PTS\_EIIA and a MFS)

belonging to two different types of transporters. The DNA sequences encoding the ‘lactose transporter’ and GH2 of clones I11 and I14 are nearly identical (99% similarity) and they have a perfect synteny (99%) with a part of the genome of *Streptococcus thermophilus* ND07, a potential probiotic strain with anti-inflammatory properties (Ménard et al., 2004). This demonstrates that the chimeric organization of the identified transporters does not correspond to an artefact of read assembly in our metagenomic sequences. The characterization of the role of each component of this I11/I14 original transporter will have to be determined by protein engineering.

## **Conclusion**

The main purpose of this study was to identify functional transporters from uncultured bacteria. The approach we described in this chapter allowed us to identify functional transporters from various families and taxonomical origins, and to highlight functions which had never been biochemically proven for MFS. We demonstrated the genericity of this strategy, which is compatible with all kind of known oligosaccharide transporter families described to date, whatever is the Gram status of the bacteria they originate from. Proving the translocation function of transporters is now possible at medium throughput, while the other recombinant approaches target only the binding ability. We highlighted several gene clusters involved in dietary fiber utilization by gut bacteria, including FOS and lactulose which are widely used as prebiotics (Bindels et al., 2015). The detailed mechanism and specificity of each of these transporters, especially those with the most original structure or of which the genes are the most abundant in gut microbiomes, will have to be deciphered in future studies.

## **Materials and methods**

### **Metagenomic clones**

The metagenomic clones used for present screening were obtained from the F, I, IVVE or IVET metagenomic libraries constructed as previously described (Cecchini et al., 2013; Tasse et al., 2010; Ufart éet al., in preparation). Briefly, the F library was issued from a fecal sample collected from a healthy 30 years old male who followed a vegetarian and fish-eating diet. The person did not eat any functional food such as prebiotics or probiotics, nor did he receive any antibiotics or other drugs during the 6 months before sampling. The I library was issued from a distal ileum sample obtained from a 51 years old male patient undergoing colonoscopy and surgery for lower colon cancer suspicion, after he has been submitted to a cleansing

preparation. The IVVE library was issued of a mix of the ruminal contents of two cows fed for seven weeks with wheat straw enriched food before sampling and library construction. The same sample for IVVE library was enriched *in vitro* in a bioreactor fed only by wheat straw. All the metagenomic DNA fragments were cloned into pCC1FOS fosmid and transformed into the EPI100 *E. coli* cells (Epicentre Technologies).

## **Carbohydrate substrates**

Commercial prebiotic oligosaccharides and polysaccharides used for screening were the following: fructooligosaccharides (Actilight-FOS, Beghin Meiji, France), xylooligosaccharides (Iro Taihe International, China), lactulose (TEVA, France), lactose (Sigma), cellobiosyl-cellobiose (Megazyme), glucosyl-celotriose (Megazyme), celotriose (Megazyme), galactosyl-mannobiose (Megazyme), mannotriose (Megazyme), aldouronic acids (Megazyme) and laminatriose (Megazyme). All of them were prepared at 10 mg/ml in sterilized MilliQH<sub>2</sub>O and then were filtered with a 0.20 mm Minisart RC4 syringe filter.

## **Growth assays**

All the metagenomic clones tested in our study were stored in glycerol stocks stored at -80 °C. They were recovered on Luria-Bertani (LB) agar plates supplement with 12.5 mg/l chloramphenicol. Then single colony inoculates were grown in tubes containing 2 ml LB broth supplemented with 12.5 mg/l chloramphenicol, overnight at 37 °C with orbital shaking at 200 rpm. Cultures in 48-well microplates of 500 µl of M9 medium supplemented with 12.5 mg/l chloramphenicol and 0.5% (w/v) oligosaccharides were then started at an initial OD<sub>600nm</sub> of 0.05 by inoculating the wells from the overnight precultures. The M9 medium contained 17.4 g/l Na<sub>2</sub>HPO<sub>4</sub> ·12H<sub>2</sub>O, 3.03 g/l KH<sub>2</sub>PO<sub>4</sub>, 0.51 g/l NaCl, 2.04 g/l NH<sub>4</sub>Cl, 0.49 g/l MgSO<sub>4</sub>, 4.38 mg/l CaCl<sub>2</sub>, 15 mg/l Na<sub>2</sub>EDTA 2H<sub>2</sub>O, 4.5 mg/l ZnSO<sub>4</sub> ·7H<sub>2</sub>O, 0.3 mg/l CoCl<sub>2</sub> 6H<sub>2</sub>O, 1 mg/l MnCl<sub>2</sub> 4H<sub>2</sub>O, 1 mg/l H<sub>3</sub>BO<sub>3</sub>, 0.4 mg/l Na<sub>2</sub>MoO<sub>4</sub> 2H<sub>2</sub>O, 3 mg/l FeSO<sub>4</sub> ·7H<sub>2</sub>O, 0.3 mg/l CuSO<sub>4</sub> 5H<sub>2</sub>O, 0.1 g/l thiamine and 0.02 g/l leucine. Cell growth was monitored at 37 °C with orbital shaking at 600 rpm. The optical density of the cultures was measured at 600 nm over 48 h, by using a plate reader (Infinite M200pro, 593 TECAN).

## **Bioinformatic analysis**

The functional and taxonomical annotation of the metagenomic sequences were retrieved from previous studies (Cecchini et al., 2013; Tasse et al., 2010; Ufart é et al., in preparation). Functional annotation of the transporter encoding genes was confirmed by analyzing the conserved domains

of the putative proteins by using a BLAST analysis against the NCBI's Conserved Domain Database, with an E-value threshold of 0.01 (Marchler-Bauer et al., 2017). When an incomplete transport system was detected (missing the SusD neighboring gene, or one of the component of canonical PTS and ABC transporters), we checked the presence of the complementary component by analyzing the conserved domains of the two proteins encoded by the genes before or after the main transporter encoding gene. The sequence identity between the metagenomic DNA inserts was analyzed by using NCBI's Basic Local Alignment Search Tool (BLAST) (Boratyn et al., 2013).

## References

- Bindels, L.B., Delzenne, N.M., Cani, P.D., Walter, J., 2015. Towards a more comprehensive concept for prebiotics. *Nat. Rev. Gastroenterol. Hepatol.* 12, 303–310. <https://doi.org/10.1038/nrgastro.2015.47>
- Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé N., Vasse, J., Lauber, E., Arlat, M., 2007. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS One* 2, e224. <https://doi.org/10.1371/journal.pone.0000224>
- Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezuk, Y., Raytselis, Y., Sayers, E.W., Tao, T., Ye, J., Zaretskaya, I., 2013. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* 41, 29–33. <https://doi.org/10.1093/nar/gkt282>
- Cao, Y., Jin, X., Levin, E.J., Huang, H., Zong, Y., Quick, M., Javitch, J.A., Rajashankar, K.R., 2011. Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature* 473, 50–54. <https://doi.org/10.1038/nature09939>
- Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré J., Henrissat, B., Remaud-Simón, M., Monsan, P., Potocki-Véronèse, G., 2013. Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One* 8, e72766. <https://doi.org/10.1371/journal.pone.0072766>
- Cummings, J.H., Stephen, A.M., 2007. Carbohydrate terminology and classification. *Eur. J. Clin. Nutr.* 61, S5–S18. <https://doi.org/10.1038/sj.ejcn.1602936>
- Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E.A., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., Rogowski, A., Hamilton, B.S., Chen, R., Tolbert, T.J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munoz-Munoz, J.L., Day, A., Peñá, M.J., McLean, R., Suits, M.D., Boraston, A.B., Atherly, T., Ziemer, C.J.,

- Williams, S.J., Davies, G.J., Abbott, W.D., Martens, E.C., Gilbert, H.J., 2015. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* 517, 165–169. <https://doi.org/10.1038/nature13995>
- Dagkesamanskaya, A., Langer, K., Tauzin, A.S., Rouzeau, C., Lestrade, D., Potocki-Veronese, G., Boitard, L., Bibette, J., Baudry, J., Pompon, D., Anton-Leberre, V., 2018. Use of photoswitchable fluorescent proteins for droplet-based microfluidic screening. *J. Microbiol. Methods* 147, 59–65. <https://doi.org/10.1016/j.mimet.2018.03.001>
- Elbourne, L.D.H., Tetu, S.G., Hassan, K.A., Paulsen, I.T., 2017. TransportDB 2.0: A database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* 45, D320–D324. <https://doi.org/10.1093/nar/gkw1068>
- Genee, H.J., Bali, A.P., Petersen, S.D., Siedler, S., Bonde, M.T., Gronenberg, L.S., Kristensen, M., Harrison, S.J., Sommer, M.O.A., 2016. Functional mining of transporters using synthetic selections. *Nat. Chem. Biol.* 12, 1015–1022. <https://doi.org/10.1038/nchembio.2189>
- Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W., Brumer, H., 2017. Polysaccharide utilization loci: fuelling microbial communities. *J. Bacteriol.* 199, e00860-16. <https://doi.org/10.1128/JB.00860-16>
- Han, Y., Agarwal, V., Dodds, D., Kims, J., Bae, B., Mackie, R.I., Nair, S.K., Cann, I.K.O., 2012. Biochemical and structural insights into xylan utilization by the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *J. Biol. Chem.* 287, 34946–34960. <https://doi.org/10.1074/jbc.M112.391532>
- Hara, K.Y., Kobayashi, J., Yamada, R., Sasaki, D., Kuriya, Y., Hirono-Hara, Y., Ishii, J., Araki, M., Kondo, A., 2017. Transporter engineering in biomass utilization by yeast. *FEMS Yeast Res.* 17, 1–14. <https://doi.org/10.1093/femsyr/fox061>
- Hemsworth, G.R., Dejean, G., Davies, G.J., Brumer, H., 2016. Learning from microbial strategies for polysaccharide degradation. *Biochem. Soc. Trans.* 44, 94–108. <https://doi.org/10.1042/BST20150180>
- Iyer, R., Camilli, A., 2007. Sucrose metabolism contributes to *in vivo* fitness of *Streptococcus pneumoniae*. *Mol. Microbiol.* 66, 1–13. <https://doi.org/10.1111/j.1365-2958.2007.05878.x>
- Jensen, J.B., Peters, N.K., Bhuvaneshwari, T. V., 2002. Redundancy in periplasmic binding protein-dependent transport systems for trehalose, sucrose, and maltose in *Sinorhizobium meliloti*. *J. Bacteriol.* 184, 2978–2986. <https://doi.org/10.1128/JB.184.11.2978-2986.2002>

- Joglekar, P., Sonnenburg, E.D., Higginbottom, S.K., Earle, K.A., Morland, C., Shapiro-Ward, S., Bolam, D.N., Sonnenburg, J.L., Abbott, W., Koropatkin, N., 2018. Genetic variation of the SusC/SusD homologs from a polysaccharide utilization locus underlies divergent fructan specificities and functional adaptation in *Bacteroides thetaiotaomicron* strains. *mSphere* 3, e00185-18. <https://doi.org/10.1128/mSphereDirect.00185-18>
- Kamerling, J.P., 2007. Basics Concepts and Nomenclature Recommendations in Carbohydrate Chemistry, in: Kamerling, H. (Ed.), *Comprehensive Glycoscience*. Elsevier, pp. 1–38. <https://doi.org/10.1016/B978-044451967-2/00001-5>
- Kappelmann, L., Krüger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F., Becher, D., Shapiro, N., Schweder, T., Amann, R.I., Teeling, H., 2019. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J.* 13, 76–91. <https://doi.org/10.1038/s41396-018-0242-6>
- Kleiner-Grote, G.R.M., Risse, J.M., Friehs, K., 2018. Secretion of recombinant proteins from *E. coli*. *Eng. Life Sci.* 18, 532–550. <https://doi.org/10.1002/elsc.201700200>
- Koropatkin, N.M., Martens, E.C., Gordon, J.I., Smith, T.J., 2008. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* 16, 1105–1115. <https://doi.org/10.1016/j.str.2008.03.017>
- Lai, X., Davis, F. V., Hespell, R.B., Ingram, L.O., 1997. Cloning of cellobiose phosphoenolpyruvate-dependent phosphotransferase genes: functional expression in recombinant. *Appl. Environ. Microbiol.* 63, 355–363. <https://doi.org/10.1113/25146>
- Lai, X., Ingram, L.O., 1993. Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. *J. Bacteriol.* 175, 6441–6450. <https://doi.org/10.1128/jb.175.20.6441-6450.1993>
- Lapébie, P., Lombard, V., Drula, E., Terrapon, N., Henrissat, B., 2019. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat. Commun.* 10, 2043. <https://doi.org/10.1038/s41467-019-10068-5>
- Linke, C.M., Woodiga, S.A., Meyers, D.J., Buckwalter, C.M., Salhi, H.E., King, S.J., 2013. The ABC transporter encoded at the pneumococcal fructooligosaccharide utilization locus determines the ability to utilize long- and short-chain fructooligosaccharides. *J. Bacteriol.* 195, 1031–1041. <https://doi.org/10.1128/JB.01560-12>
- Lombard, V., Ramulu H.G., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495.

<https://doi.org/10.1093/nar/gkt1178>

- Luis, A.S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., Baslé A., Cartmell, A., Terrapon, N., Stott, K., Lowe, E.C., McLean, R., Shearer, K., Schückel, J., Venditto, I., Ralet, M.C., Henrissat, B., Martens, E.C., Mosimann, S.C., Abbott, D.W., Gilbert, H.J., 2018. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. *Nat. Microbiol.* 3, 210–219. <https://doi.org/10.1038/s41564-017-0079-1>
- Maharjan, R.P., Seeto, S., Ferenci, T., 2007. Divergence and redundancy of transport and metabolic rate-yield strategies in a single *Escherichia coli* population. *J. Bacteriol.* 189, 2350–2358. <https://doi.org/10.1128/JB.01414-06>
- Majd, H., King, M.S., Palmer, S.M., Smith, A.C., Elbourne, L.D.H., Paulsen, I.T., Sharples, D., Henderson, P.J.F., Kunji, E.R.S., 2018. Screening of candidate substrates and coupling ions of transporters by thermostability shift assays. *Elife* 7, 1–17. <https://doi.org/10.7554/eLife.38821>
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., Bryant, S.H., 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. <https://doi.org/10.1093/nar/gkw1129>
- Marques, F.Z., Mackay, C.R., Kaye, D.M., 2018. Beyond gut feelings: how the gut microbiota regulates blood pressure. *Nat. Rev. Cardiol.* 15, 20–32. <https://doi.org/10.1038/nrcardio.2017.120>
- Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., Gordon, J.I., 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* 9. <https://doi.org/10.1371/journal.pbio.1001221>
- McCoy, J.G., Ren, Z., Stanevich, V., Lee, J., Mitra, S., Levin, E.J., Poget, S., Quick, M., Im, W., Zhou, M., 2016. The structure of a sugar transporter of the glucose EIIC superfamily provides insight into the elevator mechanism of membrane transport. *Structure* 24, 956–964. <https://doi.org/10.1016/j.str.2016.04.003>
- Ménard, S., Candalh, C., Bambou, J.C., Terpend, K., Cerf-Bensussan, N., Heyman, M., 2004. Lactic acid bacteria secrete metabolites retaining anti-inflammatory properties after intestinal transport. *Gut* 53, 821–828. <https://doi.org/10.1136/gut.2003.026252>

- Moreno-Hagelsieb, G., 2015. The power of operon rearrangements for predicting functional associations. *Comput. Struct. Biotechnol. J.* 13, 402–406. <https://doi.org/10.1016/j.csbj.2015.06.002>
- Ndeh, D., Gilbert, H.J., 2018. Biochemistry of complex glycan depolymerisation by the human gut microbiota. *FEMS Microbiol. Rev.* 42, 146–164. <https://doi.org/10.1093/femsre/fuy002>
- Pollet, A., Delcour, J.A., Courtin, C.M., 2010. Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. *Crit. Rev. Biotechnol.* 30, 176–191. <https://doi.org/10.3109/07388551003645599>
- Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Baslé A., Morland, C., Day, A.M., Zheng, H., Rogers, T.E., Thompson, P., Hawkins, A.R., Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J., Bolam, D.N., 2015. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* 6. <https://doi.org/10.1038/ncomms8481>
- Saier, M.H., 2015. The Bacterial Phosphotransferase System: New frontiers 50 years after its discovery. *J. Mol. Microbiol. Biotechnol.* 25, 73–78. <https://doi.org/10.1159/000381215>
- Saier, M.H., 2000. Families of transmembrane sugar transport proteins. *Mol. Microbiol.* 35, 699–710. <https://doi.org/10.1046/j.1365-2958.2000.01759.x>
- Saier, M.H., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C., Moreno-Hagelsieb, G., 2016. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.* 44, D372–D379. <https://doi.org/10.1093/nar/gkv1103>
- Schouler, C., Taki, A., Chouikha, I., Moulin-Schouleur, M., Gilot, P., 2009. A genomic island of an extraintestinal pathogenic *Escherichia coli* strain enables the metabolism of fructooligosaccharides, which improves intestinal colonization. *J. Bacteriol.* 91, 388–393. <https://doi.org/10.1128/JB.01052-08>
- Sheridan, P.O., Martin, J.C., Lawley, T.D., Browne, H.P., Harris, H.M.B., Bernalier-Donadille, A., Duncan, S.H., O’Toole, P.W., Scott, K.P., Flint, H.J., 2016. Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic *Firmicutes*. *Microb. Genomics* 2, e000043. <https://doi.org/10.1099/mgen.0.000043>
- Steen, A.D., Crits-Christoph, A., Carini, P., DeAngelis, K.M., Fierer, N., Lloyd, K.G., Cameron Thrash, J., 2019. High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J.* 13, 3126–3130. <https://doi.org/10.1038/s41396-019-0484-y>
- Tamura, K., Foley, M.H., Gardill, B.R., Dejean, G., Schnizlein, M., Bahr, C.M.E., Louise

- Creagh, A., van Petegem, F., Koropatkin, N.M., Brumer, H., 2019. Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*. *Cell. Mol. Life Sci.* <https://doi.org/10.1007/s00018-019-03115-3>
- Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., Doré J., Monsan, P., Remaud-Simeon, M., Potocki-Veronese, G., 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* 20, 1605–1612. <https://doi.org/10.1101/gr.108332.110>
- Tauzin, A.S., Laville, E., Xiao, Y., Nouaille, S., Le Bourgeois, P., Heux, S., Portais, J.C., Monsan, P., Martens, E.C., Potocki-Veronese, G., Bordes, F., 2016. Functional characterization of a gene locus from an uncultured gut *Bacteroides* conferring xylo-oligosaccharides utilization to *Escherichia coli*. *Mol. Microbiol.* 102, 579–592. <https://doi.org/10.1111/mmi.13480>
- Terrapon, N., Lombard, V., Drula, É., Lap bie, P., Al-Masaudi, S., Gilbert, H.J., Henrissat, B., 2018. PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* 46, D677–D683. <https://doi.org/10.1093/nar/gkx1022>
- Theilmann, M.C., Fredslund, F., Svensson, B., Lo Leggio, L., Abou Hachem, M., 2019. Substrate preference of an ABC importer corresponds to selective growth on  $\beta$ -(1,6)-galactosides in *Bifidobacterium animalis* subsp. *lactis*. *J. Biol. Chem.* 294, 11701–11711. <https://doi.org/10.1074/jbc.ra119.008843>
- Ufart é L., Laville, E., Cecchini, D., Rizzo, A., Amblard, E., Drula, E., Henrissat, B., Cleret, M., Lazuka, A., Hernandez, G., Morgavi, D., Dumon, C., Robe, P., Klopp, C., Bozonnet, S., n.d. Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown. in preparation.
- Unfried, F., Becker, S., Robb, C.S., Hehemann, J.H., Markert, S., Heiden, S.E., Hinzke, T., Becher, D., Reintjes, G., Krüger, K., Avcı, B., Kappelmann, L., Hahnke, R.L., Fischer, T., Harder, J., Teeling, H., Fuchs, B., Barbeyron, T., Amann, R.I., Schweder, T., 2018. Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. *ISME J.* 12, 2894–2906. <https://doi.org/10.1038/s41396-018-0243-5>
- Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Marth, J.D., Bertozzi, C.R., Hart, G.W., Etzler, M.E., 2009. Symbol nomenclature for glycan representation. *Proteomics* 9, 5398–5399. <https://doi.org/10.1002/pmic.200900708>

- Viborg, A.H., Terrapon, N., Lombard, V., Michel, G., Czjzek, M., Henrissat, B., Brumer, H., 2019. A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *J. Biol. Chem.* 294, 15973–15986. <https://doi.org/10.1074/jbc.RA119.010619>
- Zafar, H., Saier, M.H., 2018. Comparative genomics of transport proteins in seven *Bacteroides* species. *PLoS One* 13, 1–25. <https://doi.org/10.1371/journal.pone.0208151>

## Supplemental material

**Table S1. Known enzymatic activity of glycoside hydrolase (GH) families in the selected clones based on CAZy database and the substrates selected (based on activities in bold) for growth tests.**

| GH families | Known activities listed the CAZy database   | Selected substrates tested for growth screening   |
|-------------|---|---|
| GH1         | <b><math>\beta</math>-glucosidase</b> (EC 3.2.1.21); <b><math>\beta</math>-galactosidase</b> (EC 3.2.1.23); <b><math>\beta</math>-mannosidase</b> (EC 3.2.1.25); $\beta$ -glucuronidase (EC 3.2.1.31); <b><math>\beta</math>-xylosidase</b> (EC 3.2.1.37); $\beta$ -D-fucosidase (EC 3.2.1.38); phlorizin hydrolase (EC 3.2.1.62); exo- $\beta$ -1,4-glucanase (EC 3.2.1.74); 6-phospho- $\beta$ -galactosidase (EC 3.2.1.85); 6-phospho- $\beta$ -glucosidase (EC 3.2.1.86); strictosidine $\beta$ -glucosidase (EC 3.2.1.105); lactase (EC 3.2.1.108); amygdalin $\beta$ -glucosidase (EC 3.2.1.117); prunasin $\beta$ -glucosidase (EC 3.2.1.118); vicianin hydrolase (EC 3.2.1.119); raucaffricine $\beta$ -glucosidase (EC 3.2.1.125); thioglucosidase (EC 3.2.1.147); $\beta$ -primeverosidase (EC 3.2.1.149); isoflavonoid 7-O- $\beta$ -apiosyl- $\beta$ -glucosidase (EC 3.2.1.161); ABA-specific $\beta$ -glucosidase (EC 3.2.1.175); DIMBOA $\beta$ -glucosidase (EC 3.2.1.182); $\beta$ -glycosidase (EC 3.2.1.-); hydroxyisourate hydrolase (EC 3.-.-) | XOS;<br>Galactosyl-mannobiose and mannotriose;<br>Cellobiosyl-cellobiose, glucosyl-cellotriase, and cellotriase;<br>Lactose |
| GH2         | <b><math>\beta</math>-galactosidase</b> (EC 3.2.1.23); <b><math>\beta</math>-mannosidase</b> (EC 3.2.1.25); $\beta$ -glucuronidase (EC 3.2.1.31); $\alpha$ -L-arabinofuranosidase (EC 3.2.1.55); mannosylglycoprotein endo- $\beta$ -mannosidase (EC 3.2.1.152); exo- $\beta$ -glucosaminidase (EC 3.2.1.165); $\alpha$ -L-arabinopyranosidase (EC 3.2.1.-); $\beta$ -galacturonidase (EC 3.2.1.-); $\beta$ -xylosidase (EC 3.2.1.37); $\beta$ -D-galactofuranosidase (EC 3.2.1.146)  | Galactosyl-mannobiose and mannotriose;<br>Lactulose   |

|     |  |  |
|-----|--|--|
| GH3 | <p><b><math>\beta</math>-glucosidase (EC 3.2.1.21); xylan 1,4-<math>\beta</math>-xylosidase (EC 3.2.1.37); <math>\beta</math>-glucosylceramidase (EC 3.2.1.45); <math>\beta</math>-N-acetylhexosaminidase (EC 3.2.1.52); <math>\alpha</math>-L-arabinofuranosidase (EC 3.2.1.55); glucan 1,3-<math>\beta</math>-glucosidase (EC 3.2.1.58); glucan 1,4-<math>\beta</math>-glucosidase (EC 3.2.1.74); isoprimeverose-producing oligoxyloglucan hydrolase (EC 3.2.1.120); coniferin <math>\beta</math>-glucosidase (EC 3.2.1.126); exo-1,3-1,4-glucanase (EC 3.2.1.-); <math>\beta</math>-N-acetylglucosaminide phosphorylases (EC 2.4.1.-); <math>\beta</math>-1,2-glucosidase (EC 3.2.1.-); <math>\beta</math>-1,3-glucosidase (EC 3.2.1.-); xyloglucan-specific exo-<math>\beta</math>-1,4-glucanase / exo-xyloglucanase (EC 3.2.1.155)</b></p>  | <p>Xylosaccharides;<br/>Cellobiosyl-cellobiose,<br/>glucosyl-cellotriase, and<br/>cellotriase (<math>\beta</math>-glucosidase);</p>  |
| GH4 | <p>maltose-6-phosphate glucosidase (EC 3.2.1.122); <b><math>\alpha</math>-glucosidase (EC 3.2.1.20)</b>; <math>\alpha</math>-galactosidase (EC 3.2.1.22); 6-phospho-<math>\beta</math>-glucosidase (EC 3.2.1.86); <math>\alpha</math>-glucuronidase (EC 3.2.1.139); <math>\alpha</math>-galacturonase (EC 3.2.1.67); palatinase (EC 3.2.1.-)</p>   | <p>Galactosyl-mannobiose and<br/>mannotriose</p>   |
| GH5 | <p>endo-<math>\beta</math>-1,4-glucanase / cellulase (EC 3.2.1.4); <b>endo-<math>\beta</math>-1,4-xylanase (EC 3.2.1.8); <math>\beta</math>-glucosidase (EC 3.2.1.21); <math>\beta</math>-mannosidase (EC 3.2.1.25)</b>; <math>\beta</math>-glucosylceramidase (EC 3.2.1.45); glucan <math>\beta</math>-1,3-glucosidase (EC 3.2.1.58); exo-<math>\beta</math>-1,4-glucanase / cellodextrinase (EC 3.2.1.74); glucan endo-1,6-<math>\beta</math>-glucosidase (EC 3.2.1.75); mannan endo-<math>\beta</math>-1,4-mannosidase (EC 3.2.1.78); cellulose <math>\beta</math>-1,4-cellobiosidase (EC 3.2.1.91); steryl <math>\beta</math>-glucosidase (EC 3.2.1.104); endoglycoceramidase (EC 3.2.1.123); chitosanase (EC 3.2.1.132); <math>\beta</math>-primeverosidase (EC 3.2.1.149); xyloglucan-specific endo-<math>\beta</math>-1,4-glucanase (EC 3.2.1.151); endo-<math>\beta</math>-1,6-galactanase (EC 3.2.1.164); <math>\beta</math>-1,3-mannanase (EC 3.2.1.-); arabinoxylan-specific endo-<math>\beta</math>-1,4-xylanase (EC 3.2.1.-); mannan transglycosylase (EC 2.4.1.-); lichenase / endo-<math>\beta</math>-1,3-1,4-glucanase (EC 3.2.1.73); <math>\beta</math>-glycosidase (EC 3.2.1.-); endo-<math>\beta</math>-1,3-glucanase / laminarinase (EC 3.2.1.39); <math>\beta</math>-N-acetylhexosaminidase (EC 3.2.1.52); chitosanase (EC 3.2.1.132); <math>\beta</math>-D-galactofuranosidase (EC 3.2.1.146); <math>\beta</math>-galactosylceramidase (EC</p> | <p>XOS;<br/>Cellobiosyl-cellobiose,<br/>glucosyl-cellotriase, and<br/>cellotriase;<br/>Galactosyl-mannobiose and<br/>mannotriose</p> |

|      |   |   |
|------|---|---|
|      | 3.2.1.46)   |   |
| GH6  | endoglucanase (EC 3.2.1.4); cellobiohydrolase (EC 3.2.1.91); <b>lichenase / endo-<math>\beta</math>-1,3-1,4-glucanase (EC 3.2.1.73)</b>   | Cellobiosyl-cellobiose, glucosyl-cellotriase, and cellotriase         |
| GH7  | endo- $\beta$ -1,4-glucanase (EC 3.2.1.4); reducing end-acting cellobiohydrolase (EC 3.2.1.176); chitosanase (EC 3.2.1.132); <b>endo-<math>\beta</math>-1,3-1,4-glucanase (EC 3.2.1.73)</b>                                       | Cellobiosyl-cellobiose, glucosyl-cellotriase, and cellotriase         |
| GH8  | Family chitosanase (EC 3.2.1.132); cellulase (EC 3.2.1.4); licheninase (EC 3.2.1.73); <b>endo-1,4-<math>\beta</math>-xylanase (EC 3.2.1.8)</b> ; reducing-end-xylose releasing exo-oligoxyylanase (EC 3.2.1.156)                  | XOS   |
| GH10 | <b>endo-1,4-<math>\beta</math>-xylanase (EC 3.2.1.8)</b> ; endo-1,3- $\beta$ -xylanase (EC 3.2.1.32); tomatinase (EC 3.2.1.-); xylan endotransglycosylase (EC 2.4.2.-); <b>endo-<math>\beta</math>-1,4-glucanase (EC 3.2.1.4)</b> | XOS;<br>Cellobiosyl-cellobiose, glucosyl-cellotriase, and cellotriase |
| GH11 | <b>endo-<math>\beta</math>-1,4-xylanase (EC 3.2.1.8)</b> ; endo- $\beta$ -1,3-xylanase (EC 3.2.1.32)  | XOS   |

|      |  |   |
|------|--|---|
| GH16 | xyloglucan:xyloglucosyltransferase (EC 2.4.1.207); keratan-sulfate endo-1,4- $\beta$ -galactosidase (EC 3.2.1.103); endo-1,3- $\beta$ -glucanase / laminarinase (EC 3.2.1.39); <b>endo-1,3(4)-<math>\beta</math>-glucanase (EC 3.2.1.6)</b> ; licheninase (EC 3.2.1.73); $\beta$ -agarase (EC 3.2.1.81); $\kappa$ -carrageenase (EC 3.2.1.83); xyloglucanase (EC 3.2.1.151); endo- $\beta$ -1,3-galactanase (EC 3.2.1.181); [retaining] $\beta$ -porphyranase (EC 3.2.1.178); hyaluronidase (EC 3.2.1.35); endo- $\beta$ -1,4-galactosidase (EC 3.2.1.-); chitin $\beta$ -1,6-glucanosyltransferase (EC 2.4.1.-); $\beta$ -transglycosidase (EC 2.4.1.-); $\beta$ -glycosidase (EC 3.2.1.-); endo- $\beta$ -1,3-galactanase (EC 3.2.1.181); $\beta$ -carrageenase (EC 3.2.1.-) | Cellobiosyl-cellobiose, glucosyl-cellotriase, and cellotriase   |
| GH26 | $\beta$ -mannanase (EC 3.2.1.78); exo- $\beta$ -1,4-mannobiohydrolase (EC 3.2.1.100); $\beta$ -1,3-xylanase (EC 3.2.1.32); <b>lichenase / endo-<math>\beta</math>-1,3-1,4-glucanase (EC 3.2.1.73)</b> ; mannobiose-producing <b>exo-<math>\beta</math>-mannanase (EC 3.2.1.-)</b>  | Cellobiosyl-cellobiose, glucosyl-cellotriase, and cellotriase;<br>Galactosyl-mannobiose and mannotriase |
| GH32 | invertase (EC 3.2.1.26); endo-inulinase (EC 3.2.1.7); $\beta$ -2,6-fructan 6-levanbiohydrolase (EC 3.2.1.64); endo-levanase (EC 3.2.1.65); exo-inulinase (EC 3.2.1.80); <b>fructan <math>\beta</math>-(2,1)-fructosidase/1-exohydrolase (EC 3.2.1.153)</b> ; fructan $\beta$ -(2,6)-fructosidase/6-exohydrolase (EC 3.2.1.154); sucrose:sucrose 1-fructosyltransferase (EC 2.4.1.99); fructan:fructan 1-fructosyltransferase (EC 2.4.1.100); sucrose:fructan 6-fructosyltransferase (EC 2.4.1.10); fructan:fructan 6G-fructosyltransferase (EC 2.4.1.243); levan fructosyltransferase (EC 2.4.1.-); [retaining] sucrose:sucrose 6-fructosyltransferase (6-SST) (EC 2.4.1.-); cycloinulo-oligosaccharide fructanotransferase (EC 2.4.1.-)                                       | FOS   |

|       |  |   |
|-------|--|---|
| GH39  | $\alpha$ -L-iduronidase (EC 3.2.1.76); <b><math>\beta</math>-xylosidase (EC 3.2.1.37)</b>  | XOS   |
| GH42  | <b><math>\beta</math>-galactosidase (EC 3.2.1.23)</b> ; $\alpha$ -L-arabinopyranosidase (EC 3.2.1.-)   | Lactulose   |
| GH43  | <b><math>\beta</math>-xylosidase (EC 3.2.1.37)</b> ; $\alpha$ -L-arabinofuranosidase (EC 3.2.1.55); xylanase (EC 3.2.1.8); $\alpha$ -1,2-L-arabinofuranosidase (EC 3.2.1.-); exo- $\alpha$ -1,5-L-arabinofuranosidase (EC 3.2.1.-); [inverting] exo- $\alpha$ -1,5-L-arabinanase (EC 3.2.1.-); $\beta$ -1,3-xylosidase (EC 3.2.1.-); [inverting] exo- $\alpha$ -1,5-L-arabinanase (EC 3.2.1.-); [inverting] endo- $\alpha$ -1,5-L-arabinanase (EC 3.2.1.99); exo- $\beta$ -1,3-galactanase (EC 3.2.1.145); $\beta$ -D-galactofuranosidase (EC 3.2.1.146) | XOS   |
| GH51  | endoglucanase (EC 3.2.1.4); endo- $\beta$ -1,4-xylanase (EC 3.2.1.8); <b><math>\beta</math>-xylosidase (EC 3.2.1.37)</b> ; $\alpha$ -L-arabinofuranosidase (EC 3.2.1.55); lichenase / endo- $\beta$ -1,3-1,4-glucanase (EC 3.2.1.73)   | XOS   |
| GH94  | <b>cellobiose phosphorylase (EC 2.4.1.20)</b> ; laminaribiose phosphorylase (EC 2.4.1.31); <b>cellodextrin phosphorylase (EC 2.4.1.49)</b> ; chitobiose phosphorylase (EC 2.4.1.-); cyclic $\beta$ -1,2-glucan synthase (EC 2.4.1.-); cellobionic acid phosphorylase (EC 2.4.1.321); $\beta$ -1,2-oligoglucan phosphorylase (EC 2.4.1.-)   | Cellobiosyl-cellobiose, glucosyl-cellotriase, and cellotriase |
| GH130 | $\beta$ -1,4-mannosylglucose phosphorylase (EC 2.4.1.281); <b><math>\beta</math>-1,4-mannooligosaccharide phosphorylase (EC 2.4.1.319)</b> ; $\beta$ -1,4-mannosyl-N-acetyl-glucosamine phosphorylase (EC 2.4.1.320); $\beta$ -1,2-mannobiose phosphorylase (EC 2.4.1.-); $\beta$ -1,2-oligomannan phosphorylase (EC 2.4.1.-); $\beta$ -1,2-mannosidase (EC 3.2.1.-)   | Galactosyl-mannobiose and mannotriase                         |



## Second article

Fructooligosaccharides (FOS) are increasingly being used as prebiotics in human and livestock alimentation, to selectively stimulate the growth of beneficial human gut bacteria, such as *Bifidobacteria* and *Lactobacilli*. However, it has been observed that FOS can also be fermented by commensal and even pathogenic gut bacteria, but the molecular mechanisms that they use to utilize FOS have rarely been investigated. The growth-based screening of metagenomic clones presented in the first chapter allowed us to identify a FOS metabolization pathway in the *Escherichia coli* clone I9. The I9 sequence is taxonomically assigned to the Firmicute species *Dorea longicatena*, and contains both a PTS transporter and an ABC transporter. By reviewing the bibliography in order to know what was known on *Dorea longicatena* and its role in the human gut ecosystem, I realized that this species is not beneficial to human health. At this stage, several questions arose:

- Is the I9 metagenomic locus conserved in several gut bacteria?
- Would it be a biomarker of pathology in the human gut microbiome?
- Could we predict the specificity of the I9 transporters by using the Transporter Classification Database?
- What is the mechanism of FOS utilization by *Dorea longicatena*? In particular, are the I9 ABC and PTS transporters both involved in FOS uptake? And what are the contributions of substrate binding and uptake in the FOS metabolization efficiency?

In the following article, I describe the strategy we used to address these questions, and the results we obtained. This paper has been published in mSphere in January 2020 (doi: 10.1128/mSphere.00771-19). In this study, I performed all the bioinformatic and biochemical analyses, except the mass spectrometry analysis, which was carried out by Pietro Tedesco and Fabien L'écuyer, who also provided the *E. coli* knockout mutants that I tested. I performed the transcriptional analysis with the help of Pascale Lepercq and Myriam Mercade. I would also like to thank Nicolas Terrapon and Matthieu Arlat for the fruitful discussions during my thesis committees, leading to significant improvement of the manuscript.



# Harvesting of prebiotic fructooligosaccharides by nonbeneficial human gut bacteria

Zhi Wang<sup>a</sup>, Alexandra S. Tauzin<sup>a</sup>, Elisabeth Laville<sup>a</sup>, Pietro Tedesco<sup>a</sup>, Fabien Létisse<sup>a</sup>, Nicolas Terrapon<sup>b,c</sup>, Pascale Lepercq<sup>a</sup>, Myriam Mercade<sup>a</sup> and Gabrielle Potocki-Veronese<sup>a#</sup>

<sup>a</sup>TBI, CNRS, INRA, INSAT, Université Toulouse, F-31077 Toulouse, France

<sup>b</sup>AFMB, UMR 7257 CNRS, Aix-Marseille Université F-13288 Marseille, France

<sup>c</sup>INRA, USC 1408 AFMB, F-13288 Marseille, France

Running Head: Prebiotic internalization by gut bacteria

<sup>#</sup>Address correspondence to Gabrielle Potocki-Veronese, veronese@insa-toulouse.fr.

## ABSTRACT

Prebiotic oligosaccharides, such as fructooligosaccharides, are increasingly being used to modulate the composition and activity of the gut microbiota. However, carbohydrate utilization analyses and metagenomic studies recently revealed the ability of deleterious and uncultured human gut bacterial species to metabolize these functional foods. Moreover, because of the difficulties of functionally profiling transmembrane proteins, only a few prebiotic transporters have been biochemically characterized to date, while carbohydrate binding and transport are the first and thus, the crucial steps in their metabolization. Here, we describe the molecular mechanism of a phosphotransferase system, highlighted as a dietary and pathology biomarker in the human gut microbiome. This transporter is encoded by a metagenomic locus that is highly conserved in several human gut Firmicutes, including *Dorea* species. We developed a generic strategy to deeply analyze, *in vitro* and *in cellulo*, the specificity and functionality of recombinant transporters in *Escherichia coli*, combining carbohydrate utilization locus and host genome engineering, quantification of the binding, transport and growth rates with analysis of phosphorylated carbohydrates by mass spectrometry. We demonstrated that the *Dorea* fructooligosaccharide transporter is specific for kestose, whether for binding, transport or phosphorylation. This constitutes the biochemical proof of effective phosphorylation of glycosides with a degree of polymerization of more than 2, extending the known functional diversity of phosphotransferase systems. Based on these new findings, we revisited the classification of these carbohydrate transporters.

## IMPORTANCE

Prebiotics are increasingly used as food supplements, especially in infant formulas, to modify the functioning and composition of the microbiota. However, little is currently known about the mechanisms of prebiotic recognition and transport by gut bacteria, while these steps are crucial in their metabolism. In this study, we established a new strategy to profile the specificity of oligosaccharide transporters, combining microbiomics, genetic locus and strain engineering, and state-of-the art metabolomics. We revisited the transporter classification database and proposed a new way to classify these membrane proteins, based on their structural and mechanistic similarities. Based on these developments, we identified and characterized, at the molecular level, a fructooligosaccharide transporting phosphotransferase system, which constitutes a biomarker of diet and gut pathology. The deciphering of this

prebiotic metabolization mechanism by a nonbeneficial bacterium highlights the controversial use of prebiotics, especially in the context of chronic gut diseases.

## KEYWORDS

*Dorea*, chronic gut diseases, fructooligosaccharides, microbiome, phosphotransferase system

## Introduction

Prebiotics are functional foods, defined as “nondigestible compounds that, through their metabolization by microorganisms in the gut, modulate composition and/or activity of the gut microbiota, thus conferring a beneficial physiological effect on the host” (1–3). Inulin and fructooligosaccharides (FOS), comprising  $\beta(2\rightarrow1)$  fructosyl units linked to sucrose, are universally agreed to be prebiotic fibers. They are naturally present in a variety of plants such as Jerusalem artichoke, wheat, barley, rye, onions, garlic, asparagus and banana (4, 5). They are increasingly used as food supplements, especially in infant formulas (5, 6), with the global inulin and FOS markets estimated to exceed \$5.8 billion by 2024 (7, 8). Numerous studies indicate that inulin and FOS are selectively fermented by human gut bifidobacteria and lactobacilli, inducing beneficial effects for the host, such as facilitating defecation and reducing endotoxaemia, insulin resistance and colon cancer risks (9–12). Notwithstanding this evidence and the commercial interest, some nonbeneficial gut bacterial species (13, 14), like *Dorea longicatena*, a species associated with several pathologies such as irritable bowel syndrome (IBS) (15, 16), have been shown to feed on these compounds. Moreover, abdominal pain in IBS is alleviated by limiting the consumption of fermentable oligo-, di-, mono- saccharides, and polyols (FODMAPs) such as FOS, resulting in a decrease in *Bacteroides*, *Ruminococcaceae*, *Faecalibacterium*, and *Dorea* species in the gut microbiota composition (17). In addition, the abundance of *Dorea* species increases after fermentation of the human fecal microbiota on FOS (18). *Dorea longicatena* is stimulated in the mice fecal microbiota after mice were transplanted with a human microbiota and fed with FOS (19). Furthermore, we previously demonstrated that many prominent uncultured human gut bacteria, of which the benefits or risks for human health are not known, are well equipped to metabolize prebiotics, including FOS (20). Besides, several pathogenic *Escherichia coli* strains, including strain BEN2908 that has the ability to invade human epithelial cells (21),

can metabolize FOS (13, 22). In any event, regardless of their microbial origins, there has been little investigation of the mechanisms of FOS metabolization at the molecular level.

In bacteria, the complete metabolization of glycosides requires at least a set of carbohydrate-active enzymes (CAZymes, <http://www.cazy.org/>) (23) and one or more specific transporters (24, 25). This complex catabolic machinery is encoded by genes clustered within the same genomic loci (26, 27), called polysaccharide utilization loci (PULs) for Bacteroidetes (28). This name is further extended to the Gram-positive bacteria, known as Gram-positive PULs (gpPULs) (29). As with the other glycans, the proposed mechanism of inulin and FOS metabolization requires four steps (30): (i) extracellular hydrolysis of the longest chains, by cell surface-associated inulinases and/or fructosidases, releasing monosaccharides or short oligosaccharides; (ii) internalization of these short compounds by transporters; (iii) final depolymerization of the oligosaccharides by intracellular fructosidases; and (iv) monosaccharide assimilation in the central metabolism. A variety of bacterial inulin-targeting Bacteroidetes PULs (31, 32) and gpPULs (33) have been identified in the past few decades, and many inulinases and fructosidases have been biochemically characterized (34). However, only a few transporters have been proven to internalize FOS: the ATP-binding cassette (ABC) transport systems from the probiotic strain *Lactobacillus acidophilus* NCFM (35) and the oral pathogenic strain *Streptococcus pneumoniae* TIGR4 (36), the major facilitator superfamily (MFS) transporter from the extraintestinal pathogenic *Escherichia coli* strain BEN2908 (22), and two phosphotransferase systems (PTSs) from the probiotic strain *Lactobacillus plantarum* ST-III (37). Nevertheless, the binding and transport specificity of these transporters have still not been determined.

There are, in fact, several technological barriers to the functional characterization of transporters in native bacteria. Despite major advances for some bacterial genera over the last 10 years (38, 39), the difficulties of genetic manipulation, often combined with functional redundancy, render it challenging to characterize transport systems. In the case of uncultured species, it is impossible to inactivate or delete genes. Notably, these species make up more than 70% of the human gut ecosystem (40). In addition, transmembrane proteins cannot be studied as easily as soluble proteins. Consequently, experimental validation of function has taken place for only a few glycoside transporters, while there are probably hundreds of thousands in bacteria, considering that the PUL database lists more than 30,000 Bacteroidetes sequences for the SusC/D transporter family alone, from 1,154 species (28).

Here, we present a highly generic approach to *in vitro* and *in cellulo* characterization of the binding and transport specificities of membrane-bound proteins based on heterologous expression in *E. coli*. Our strategy combines PUL engineering, analysis of growth dynamics, characterization of transporter functionality, and quantification of binding and transport rates. This novel strategy allowed us to deeply characterize a FOS PTS transporter issued from a dominant uncultured *Dorea* strain, previously identified by mining the human ileal microbiome for prebiotic metabolization pathways (20). This *Dorea* FOS utilization system is the first prebiotic metabolization pathway to be characterized from a nonbeneficial human gut bacterium.

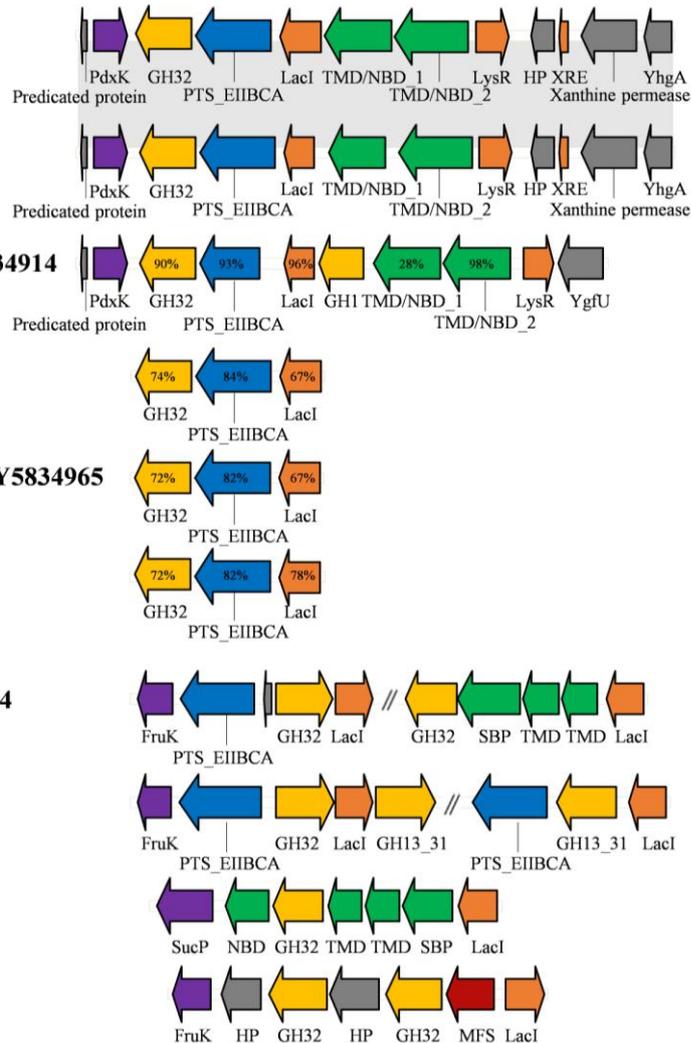
## Results

### Organization of the *Dorea* FOS utilization locus

By means of activity-based screening of the human ileal metagenome, we previously identified a fosmid *E. coli* clone (clone I9) producing intracellular fructosidase activity responsible for the hydrolysis of a mixture of FOS, containing 1-kestose (GF<sub>2</sub>), nystose (GF<sub>3</sub>), and 1<sup>F</sup>-fructofuranosyl-nystose (GF<sub>4</sub>) (20). After sequencing the I9 metagenomic DNA, two contigs were obtained, I9a and I9b, with sizes of 14,714 and 13,700 bp, respectively. The I9a contig encodes a FOS utilization locus, containing genes annotated as encoding a glycoside hydrolase belonging to the family GH32 of the CAZy classification, a PTS component, a transcriptional regulator of the LacI family and two ABC transporter proteins (**Figure 1a**). No peptide signal was identified in the GH32 sequence, indicating a cytoplasmic location of this enzyme in the native organism. The RPS-BLAST analysis against the NCBI's Conserved Domain Database (CDD) indicated that both proteins annotated as ABC transporters are multi-domain proteins, each with a transmembrane domain (TMD) and a cytosolic nucleotide-binding domain (NBD), which together constitutes a complete ABC transporter that nevertheless lacks a solute binding protein (SBP). The same bioinformatic analysis of the PTS transporter sequence of the I9a contig was performed. PTSs are multiple component carbohydrate uptake systems, allowing the phosphoryl group from phosphoenolpyruvate (PEP) to be successively transferred from the Enzyme I (EI) to the Histidine Protein (HPr), then to the domains A and B of the Enzyme II (EII), and finally to the carbohydrate bound to the EII C domain (41). The I9 PTS sequence analysis revealed that it is a single multi-domain protein, composed of the three domains B, C and A of the PTS EII complex. We thus named it PTS\_EIIBCA. The I9 FOS utilization locus lacks genes encoding EI and HPr. As EI and HPr

have been proven to be non-specific and shared by different PTS systems (42, 43), we hypothesized that *E. coli* EI and HPr could complement the I9 PTS\_EIIBCA. Moreover, no transmembrane helix was predicted in the EIIA and EIIB domains, but ten were found in the EIIC domain, suggesting that the EIIC domain is membrane-spanning and could potentially bind extracellular substrates.

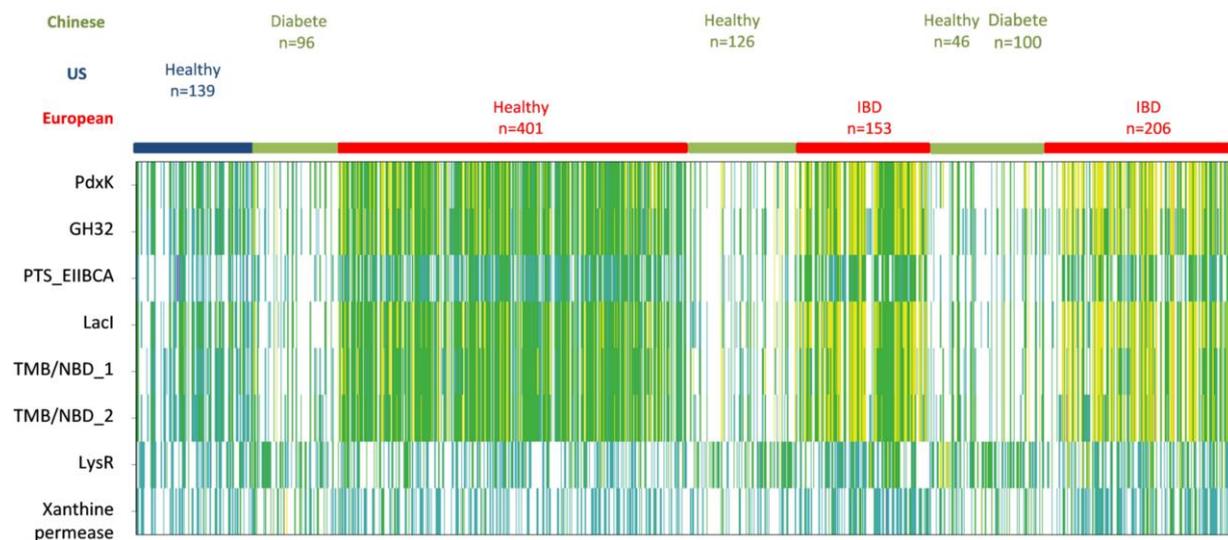
The I9a nucleic acid sequence presents 99% identity on 93% of its length with a part of the *Dorea longicatena* DSM 13814 genome (**Figure 1a**). In addition, the I9a sequence presents a microsynteny, covering from the GH32- to the LacI-encoding genes, with several human gut Firmicutes genomes, such as *Dorea longicatena*, *Ruminococcus torques* and *Coprococcus* sp., which all belong to the Clostridiales order (**Figure 1a**). To date, no FOS utilization locus has been characterized from Clostridiales, even though several species in this order have been proven to metabolize FOS *in vitro* (13, 44). In contrast, several other Firmicutes loci harboring just one of the PTS or ABC transport systems associated with at least one GH32 and a regulator have been proposed to be involved in the utilization of FOS by *Lactobacillus* spp. using transcriptomic analysis, sometimes combined with gene inactivation (**Figure 1b**) (35, 37, 45). In *Streptococcus pneumoniae*, two GH32-containing loci, the *scr*-locus including a PTS transporter, and the *sus*-locus including an ABC transport system were shown to be responsible for the utilization of sucrose and FOS, respectively (**Figure 1b**) (36). Here though, we observed both transport systems localized in the same locus. The presence within the same locus of both transport systems suggests a difference in the specificities of the I9 ABC and PTS.

**a****I9 metagenomic locus***Dorea longicatena* DSM 13814*Dorea longicatena* 2789STDY5834914*Ruminococcus torques* L2-14*Ruminococcus torques* 2789STDY5834965*Coproccoccus* sp. 43\_8**b***Streptococcus pneumoniae* TIGR4*Lactobacillus plantarum* ST-III*Lactobacillus acidophilus* NCFM*Escherichia coli* BEN2908**Figure 1. Representation of FOS and sucrose utilization loci.**

(a) Alignment of the I9 metagenomic FOS utilization locus with similar loci from cultured bacteria. Synteny between the I9 sequence and the genomic locus of *D. longicatena* DSM 13814 is shown by grey bars. Percentages of identity are indicated when lower than 100%. (b) Alignment of the characterized FOS and sucrose utilization loci involving either an ABC, an MFS or a PTS transport system, a GH32 and a related regulator. Genes predicted as being involved in carbohydrate sensing, transport and hydrolysis are color coded: glycoside hydrolases (GH) in yellow, PTS in deep blue, ABC transporter components in green (NBD, nucleotide binding domain; TMD, transmembrane domain; SBP, solute binding protein), MFS in red, LacI regulator and XRE (xenobiotic response element) family transcriptional regulator in orange, and PdxK (pyridoxine kinase), FruK (fructokinase) and SucP (sucrose phosphorylases) in purple. The other genes are indicated in grey: hypothetical protein (HP), Xanthine permease, YhgA (putative transposase YhgAlike) and ygfU (putative purine permease YgfU).

## Abundance of the *Dorea* FOS utilization locus in the human gut microbiome

The prevalence and abundance of the genes within the *Dorea* FOS utilization locus in the human gut microbiome were quantified from the abundance table of the human gut metagenomic gene catalogue (**Figure 2**) built from fecal samples from 1,267 individuals of various geographic origins (America, Europe and China), lifestyles and medical status (46).



**Figure 2. I9 FOS utilization locus abundance and prevalence in the human gut microbiome.**

Gene abundance in the fecal metagenome is represented by a color scale: white, not detected; blue, turquoise, green, yellow, orange and red, increasing abundance with a 10-fold change between colors. Based on its annotation, the xanthine permease encoding gene is not involved in carbohydrate binding, sensing, transport or modification.

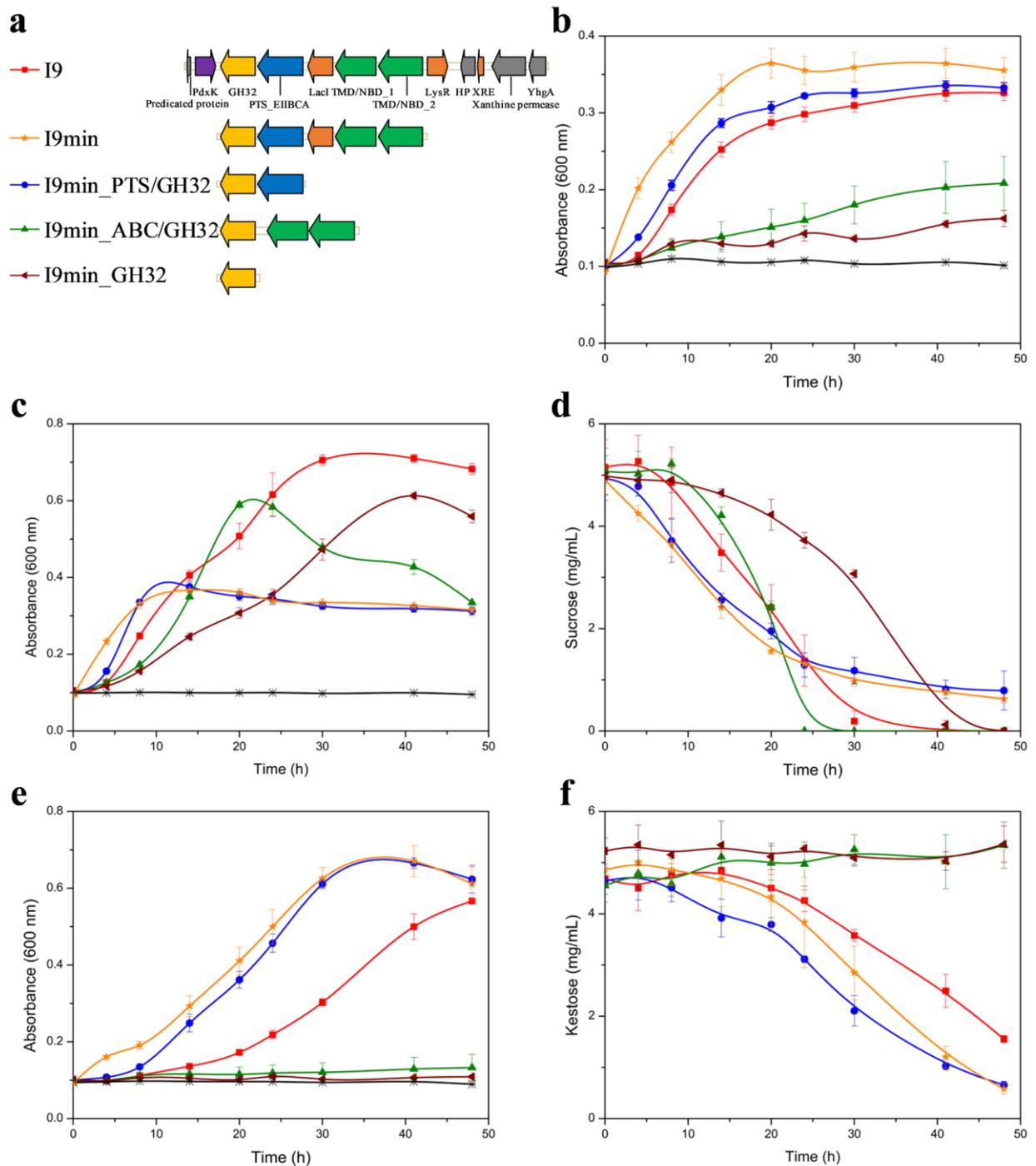
We found that eight genes of the I9 locus have homologs (with more than 97% sequence identity and 100% coverage) in the microbiome. These genes are highly prevalent, being present in at least 50% of the individuals in the cohort. Interestingly, the *pts\_EIIBCA* gene is 4.6 times less abundant than the other genes of the locus, although with a similar prevalence. This result indicates that the structure of the *Dorea* FOS locus has been altered during its evolution, which might be due to (i) the *a posteriori* gene acquisition of the *Dorea pts* gene

from an existing locus or (ii) its loss within some bacterial populations over time. Both events could have been driven by horizontal gene transfer, an essential phenomenon for bacterial genome plasticity and metabolic adaptability (47). Nevertheless, when comparing gene abundance and prevalence in the microbiome of individuals of various origins and medical status, the *pts* gene presents the same relative profile as the other genes of the FOS utilization locus. The six genes are indeed on average 3 and 6.5 times more prevalent and abundant in Westerners (including Americans and Europeans) than in individuals of Chinese origin. This is probably due to the dietary differences between Westerners and Asians. Wheat is a major source of FOS (5), and bread made with wheat is a staple food for Westerners (48, 49), while for Asians, it is rice (50, 51).

In addition, in the European cohort, these genes are similarly prevalent but twice as abundant in the microbiome of patients suffering from inflammatory bowel diseases (IBD; including Crohn's disease and ulcerative colitis) than in that of healthy individuals. This suggests that bacteria possessing this FOS utilization locus may be positively associated with IBD.

### **The *Dorea* phosphotransferase system confers FOS utilization ability to *E. coli***

First, in order to identify which genes expressed in *E. coli* could explain the *Dorea* I9 FOS utilization phenotype, the gene expression levels were quantified by reverse transcription-quantitative PCR (RT-qPCR) for the four genes potentially involved in substrate internalization (*pts\_EIIBCA*, *tmb/nbd\_1*, and *tmb/nbd\_2*) and hydrolysis (*gh32*) (**Figure S1**). Except for *tmb/nbd\_2*, which displayed a relatively low expression, all the other genes were expressed at increased levels compared to that of the endogenous *E. coli* housekeeping gene (*ihfB*). The moderate expression of the ABC transporter-encoding genes suggests that the ability of *E. coli* to internalize FOS might be mostly due to the expression of the *pts\_EIIBCA* gene. In addition, this transcriptomic study highlights the potential of *E. coli* to properly express Firmicutes genes cloned in fosmids. This result is consistent with our previous study performed with a Bacteroidetes PUL, demonstrating that a significant proportion of metagenomic genes could be expressed spuriously in *E. coli* due to the presence within the metagenomic sequence of numerous  $\sigma^{70}$  promoter sequences for *E. coli* (52).



**Figure 3. Functional analysis of the I9 FOS utilization locus.**

(a) Gene organization in the I9 variants. Growth curves of clone I9, its variants and the control clone EPI on FOS mixture (b), on sucrose (c) and kestose (e), and HPAEC-PAD analysis of the culture supernatant for growth on sucrose (d) and kestose (f). Color code: clone I9, red; I9min, yellow; I9min\_PTS/GH32, blue; I9min\_ABC/GH32, green; I9min\_GH32, brown; and EPI; black. The symbols indicate the sampling time points. The data represent the averages of biological triplicates.

Second, in order to establish the contribution of each gene in the *Dorea* FOS locus in the ability of *E. coli* to metabolize FOS, different truncated variants of the I9 metagenomic DNA insert were generated (**Figure 3a**). The *E. coli* host transformed with the empty fosmid pCC1FOS, here named EPI, was used as a negative-control strain.

We showed that only the I9 GH32 enzyme was responsible for FOS hydrolysis, by testing cytoplasmic cell extracts of the I9 variant harboring only the GH32 (I9min\_GH32) on pure sucrose, kestose, nystose, fructosyl-nystose, inulotriose, inulin and levan. Among these carbohydrates, only sucrose, kestose, nystose, fructosyl-nystose, and inulotriose were hydrolyzed (**Figure S2**; see also **Table S2** in supplemental material). Moreover, this enzyme prefers sucrose and FOS up to degree of polymerization 4 (DP4) to longer oligosaccharides, since fructosyl-nystose was not fully hydrolyzed, even present at a lower molar concentration than nystose and kestose (**Table S2**). In contrast, EPI cytoplasmic extracts displayed no activity on the same substrates, validating the fact that the *E. coli* EPI100 host strain was not able to hydrolyze FOS. In addition, no secreted activity was detected in I9min\_GH32 (data not shown), confirming that the GH32 enzyme is produced exclusively inside the *E. coli* cells.

In order to determine the internalization specificity of the two I9 transporters (the PTS and the ABC transport system), the ability of truncated variants to use FOS as sole carbon source was assessed (**Figure 3b**). First, by constructing an I9-reduced variant consisting of the genes encoding from *gh32* to *tmb/nbd\_2* (I9min), we confirmed that only this minimal locus was responsible for the growth on FOS. The I9 and I9min clones in fact displayed similar growth on FOS. The slight difference might result from the high number of recombinant proteins that have to be produced in I9, which slowed down its growth compared to I9min. Interestingly, the clone I9min\_GH32 grew slowly on the FOS mixture. This result indicates that the EPI100 host strain possesses at least one nonspecific transporter able to internalize at least one of the components of the FOS mixture that will be further metabolized by the host after hydrolysis by the intracellular GH32. The I9min\_PTS/GH32 variant, harboring the PTS transporter and the GH32, retained the growth ability of I9 on FOS. By contrast, the growth of I9min\_ABC/GH32 on FOS was strongly impaired, with a growth rate barely higher than that of I9min\_GH32. The modest growth of I9min\_ABC/GH32 might be due to the uptake of sucrose from the FOS mixture but at this stage of the study, we could not exclude that the ABC transporter has a very low ability to transport FOS.

To characterize the transport specificity of each I9 transport system more precisely, for each variant, we monitored the consumption of individual FOS components in the culture medium using HPAEC-PAD (see **Figure S3**). First, as proposed above, the host strain EPI100 possesses its own sucrose transporter, since the amount of sucrose slightly decreased over time for I9min\_GH32. When growing on the FOS mixture, clones I9, I9min and I9min\_PTS/GH32 preferentially metabolized sucrose (GF) and kestose (GF<sub>2</sub>), while nystose (GF<sub>3</sub>) and 1<sup>F</sup>-fructofuranosyl-nystose (GF<sub>4</sub>) were not used. In contrast, I9min\_ABC/GH32 consumed only sucrose. No increase of oligosaccharide concentration was observed in the culture medium during growth, indicating that the I9 ABC and PTS transporters are both importers, contrary to what could be observed with similar analyses for other transport systems, such as those from the major facilitator superfamily (MFS) (52). In order to confirm the specificity of the transporters either for sucrose or kestose, the uptake of these two oligosaccharides was examined by growth test and HPAEC-PAD analyses (**Figure 3c to e and f**). All the variants were able to use sucrose to grow except EPI. However, sucrose appears not be the favorite substrate for the PTS transporter, as the growth rate of I9min\_PTS/GH32 on sucrose is lower than for I9. I9min\_ABC/GH32 grew more rapidly than I9min\_GH32 on sucrose, but these PTS-devoid clones were unable to use kestose. This demonstrates that even if the host strain possesses its own sucrose transporter, the I9 ABC transporter has the ability to transport sucrose. As far as we know, it is the first example of carbohydrate importer of the ABC transporter family which lacks a SBP (53). When SBP lacks, such as in ABC transporters catalyzing micronutrient uptake (54), the TMDs determine the specificity of the transporter through substrate-binding sites. Nevertheless, the mechanism explaining how the I9 ABC transporter could import sucrose without SBP remains to be characterized.

To deeper analyze the specificity of the PTS and ABC transporters to fructans and fructooligosaccharides, the growth of I9 was also tested on inulin (GF<sub>n</sub>, with  $\beta$ -1,2-linked fructosyl units), levan (GF<sub>n</sub>, with  $\beta$ -2,6-linked fructosyl units) and inulotriose (F3, with  $\beta$ -1,2-linked fructosyl units). No growth was observed with any of these carbohydrates (data not shown). However, we showed that the GH32 enzyme was able to hydrolyze inulotriose *in vitro* (see **Figure S2b** and **Table S2**), demonstrating that the metabolization of inulotriose by I9 is blocked at the transport step. From these results, we conclude that (i) the I9 PTS is specific to sucrose and kestose (with a clear preference for kestose) and requires a terminal

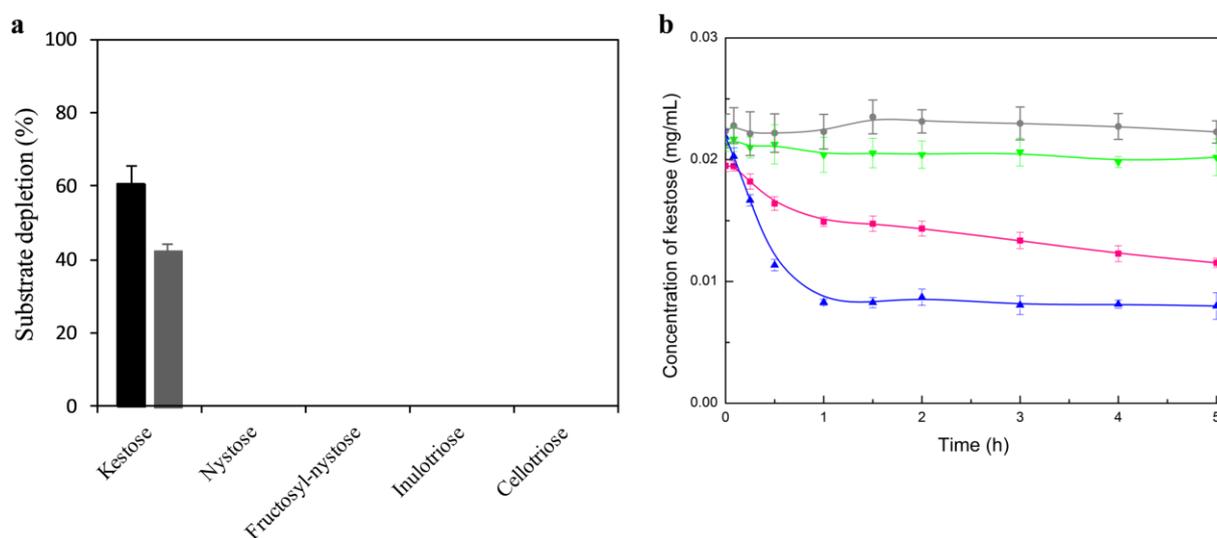
glucosyl moiety to recognize the oligosaccharide to be internalized and (ii) the I9 ABC transporter is able to transport sucrose only.

### **Binding specificity of the *Dorea* PTS transporter**

To determine the transport and binding specificities of the PTS transporter, we had to establish a new methodology. Up until now, the characterization of transporter specificity, performed *in vivo* by growth analysis after knock-out (55), determination of the internalization rate of radiolabeled substrates (56), and/or transcriptional analysis (57), did not allow for separate analysis of the binding, transport and metabolization specificities. In addition, the substrate of transcription activation differs from the substrate that is transported (58). Binding specificity can be determined *in vitro* by characterizing purified binding proteins by isothermal calorimetry, affinity gel electrophoresis (59), thermal shift assay in the presence of substrates to identify their impact on protein stability, pull-down analysis on beads covalently linked to the different sugars (60), or scintillation proximity assay with radiolabeled substrates (61). However, protein purification can be challenging, especially for membrane-bound proteins (like I9 PTS\_EIIBCA). Moreover, in all the previous studies aiming to characterize binding specificity, the binding protein is dissociated from the multiproteic assembly, which could potentially bias the results of specificity determination.

Here, we developed an easy-to-implement method based on the use of crude extracts of the cells harboring the target transporter, and non-labeled natural substrates. The unbound amount of these is quantified by HPAEC coupled to PAD, a highly sensitive detection method. This generic approach can be used to quantitatively analyze the specificity of any transporter. We established the proof of concept by characterizing the *Dorea* PTS produced in *E. coli*. A truncated variant of I9, consisting only of the PTS (named I9min\_PTS), was constructed, in order to avoid any bias due to oligosaccharide hydrolysis in oligosaccharide uptake rate quantification. First, the uptake assays were performed with intact I9min\_PTS cells incubated with various oligosaccharides in order to determine the PTS ability to internalize substrate (**Figure 4a**). Using HPAEC-PAD, the amount of residual substrate was measured in the supernatant at different time points. We observed that among the different gluco- and fructooligosaccharides tested (kestose, nystose, fructosyl-nystose, inulotriose and cellotriase), only kestose decreased in amount after incubation with I9min\_PTS, confirming that the *Dorea* FOS PTS was specific to kestose. However, by incubating intact cells with the substrate, we could not determine the contribution of binding in the uptake process.

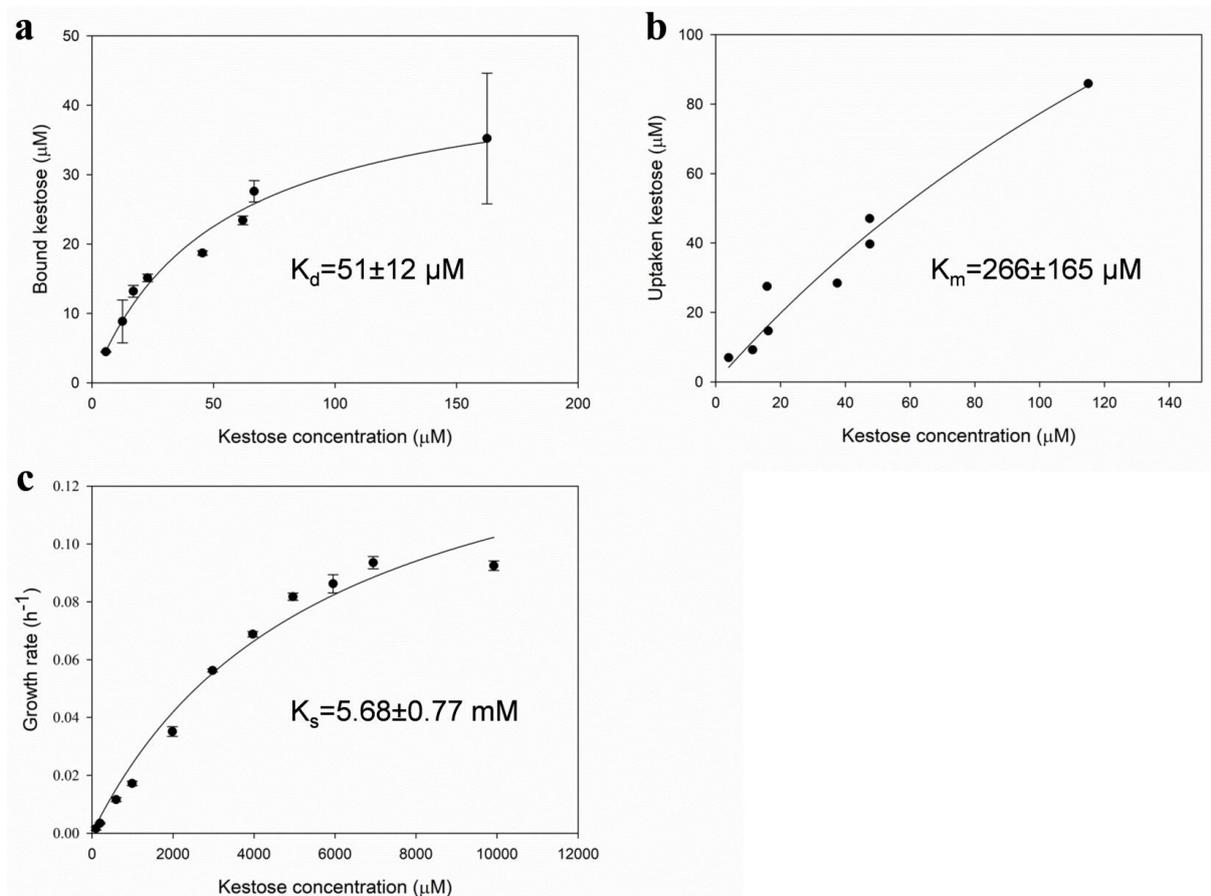
The ability of membrane debris to bind substrate was thus determined by incubating I9min\_PTS lysed cells with oligosaccharides. To focus only on the binding step, cell debris were washed several times before incubation to prevent the presence of *E. coli* cytoplasmic phosphate, HPr and EI in the reaction mixture, as this could lead to substrate phosphorylation and bias in binding rate quantification. As with the uptake assays, only kestose was bound by the I9min\_PTS membranes (**Figure 4a**). The PTS\_EIIBCA was more slowly saturated by kestose with membrane extracts than with intact cells (**Figure 4b**), indicating that kestose internalization shifts the binding equilibrium. Such a low rate of translocation had already been observed with the diacetylchitobiose-specific PTS from *Vibrio furnissii* 7225, where the transport kinetic was characterized by incubating growing *V. furnissii* cells on radiolabeled diacetylchitobiose (62). The low rate of this internalization process might be due to the covalent modification of substrate, in contrast to what happens with ABC transporters which hydrolyze ATP (54) or with secondary transporters like MFS, which use chemiosmotic ion gradient to accelerate uptake (63).



**Figure 4. Oligosaccharide uptake and binding by I9min\_PTS cells.**

(a) Screening of PTS specificity by testing binding (gray bar) and uptake (black bar) of various gluco- and fructooligosaccharides by I9min\_PTS cells. Oligosaccharide concentration was determined by HPAEC-PAD after 1h incubation of the oligosaccharides with I9min\_PTS crude membrane cell extracts (for binding) or intact cells (for uptake). (b) Kestose binding and uptake kinetics by I9min\_PTS crude membrane cell extracts and intact cells, respectively. Kestose concentration in the soluble phase was determined by HPAEC-PAD; EPI cells in

green, EPI membrane extracts in gray, I9min\_PTS cells in blue, and I9min\_PTS membrane extracts in pink. The data represent the means of three biological replicates.



**Figure 5. Saturation experiments on the *Dorea* PTS.**

Binding saturation of PTS\_EIIBCA in I9min\_PTS membranes (a), uptake saturation of I9min\_PTS cells (b) and growth saturation of I9min\_PTS/GH32 (c) with increasing concentrations of kestose. The data represent the means from two biological replicates.

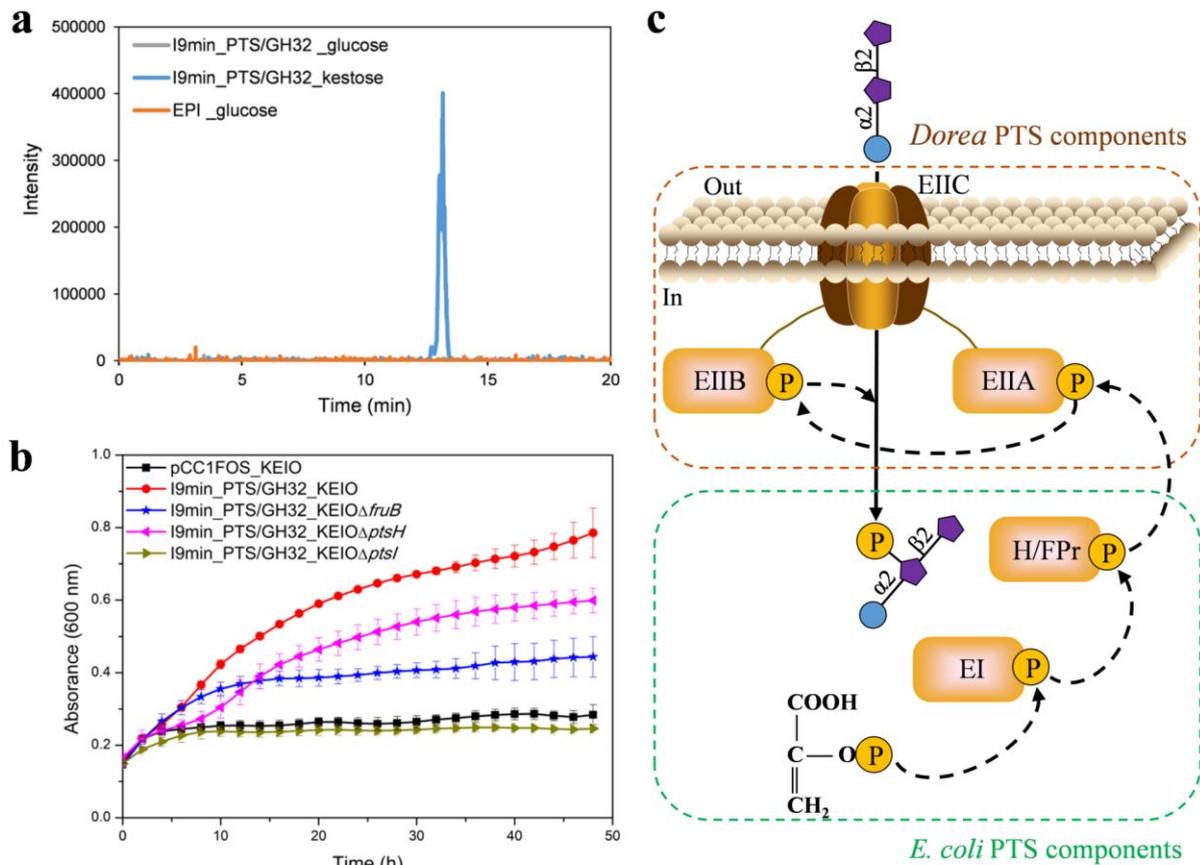
Carbohydrate binding to the transporter is the first and essential step in an active transport system. The ligand-protein binding kinetic is fundamentally related to transport efficiency and microbial chemotaxis in ecological niches (64, 65). Here, we determined reliable kinetic parameters for binding, transport and growth of I9 PTS to kestose (**Figure 5**). Saturation experiments were performed on I9min\_PTS cells and their membranes with increasing concentrations of kestose. The affinity constants of kestose uptake by I9min\_PTS cells ( $K_m$ ) and of kestose binding by PTS\_EIIBCA membranes (dissociation constant [ $K_d$ ]) were  $266 \pm 165$  and  $51 \pm 12 \mu\text{M}$ , respectively. Even if these values are of the same order of magnitude,

the fact that the  $K_m$  value is higher than the  $K_d$  one indicates that uptake is slightly less efficient than binding. It might be due to the complexity of the translocation process, which involves three PTS subunits (EIIC, EIIA and EIIB), against only one (the EIIC part that faces the exterior of the cell) for binding. In addition, to infer the dependence of growth rate on kestose concentration, we estimated the substrate affinity constant ( $K_s$ ) by applying the Monod model (e.g., Michaelis-Menten hyperbolic dependence) to the growth rate of the I9min\_PTS/GH32 cells vs. kestose concentration data (66). The estimated  $K_s$  value, of  $5.68 \pm 0.77$  mM differs by 2 orders of magnitude from the uptake and binding constants. These kinetic binding parameters indicate that the I9min\_PTS/GH32 strain has a strong binding affinity and efficient uptake system for kestose, but a poor capacity to metabolize it. In the human gut, this kind of high-affinity uptake system is crucial in competing with other microbes for limited amounts of kestose. Furthermore, the inconsistency between the values of  $K_s$ ,  $K_d$  and  $K_m$  shows that kinetic data for growth cannot be used to predict transporter binding and internalization ability.

### **Functionality of the *Dorea* PTS transporter in *E. coli***

PTS transporters couple transmembrane transport of substrate to its concomitant phosphorylation. However, the ability of PTS to phosphorylate a trisaccharide has never been shown until now. In order to demonstrate the full functionality of *Dorea* PTS\_EIIBCA in *E. coli*, and to identify the potential transient presence of intracellular phosphorylated kestose, we analyzed the intracellular metabolites of clone I9min\_PTS/GH32 grown on kestose by ion chromatography coupled with electrospray ion ionization high-resolution mass spectrometry (IC-ESI-HRMS) (67), after quenching the metabolism (**Figure 6a**).

As the *E. coli* without I9 PTS\_EIIBCA could not grow on kestose, it could not be used as negative control. We thus compared the results obtained on kestose with those obtained with clones I9min\_PTS/GH32 and of the KEIO collection (68) transformed with the empty fosmid pCC1FOS grown on glucose. For all strains, we searched for an  $m/z$  value of 583.1270, which corresponds to the theoretically expected value in negative mode for the phosphorylated kestose. For the I9min\_PTS/GH32 and EPI strains grown on glucose, the extracted ion chromatogram (XIC) does not show any peak. In contrast, we detected a chromatographically unresolved peak at retention time of 13.17 min for the I9min\_PTS/GH32 strain grown on kestose (**Figure 6a**), indicating the presence of several different forms of phosphorylated kestose.



**Figure 6. Functionality of the *Dorea* PTS in *E. coli*.**

(a) Extracted ion chromatogram (XIC) from negative ionization mode analysis at  $m/z$  583.1270, corresponding to the exact mass of the  $[M-H]^-$  ion of phosphorylated kestose (calculated from the molecular formula  $C_{18}H_{33}O_{19}P$ ). Color code: I9min\_PTS/GH32 clone grown on kestose, blue line; I9min\_PTS/GH32 grown on glucose, gray line; EPI grown on glucose, orange line. (b) Growth curves on FOS of *E. coli* K-12 BW25113 (KEIO) single-knockout mutants for PTS genes (*ptsI*, *ptsH*, and *fruB*) transformed with fosmid I9min\_PTS/GH32. The KEIO wild-type strain transformed with the empty fosmid pCC1FOS was used as the negative control. HPr (histidine protein) encoded by *ptsH* is a cytoplasmic proteic component of the *E. coli* glucose-PTS. It accepts a phosphate group from EI (encoded by *ptsI*), a cytoplasmic protein required for *E. coli* PTS functioning, and transfers it to the carbohydrate-specific EIIA domain. *FruB* encodes the HPr-like component FPr of the *E. coli* fructose-PTS. These data represent the means from three biological replicates. (c) Schematic

representation of 1-kestose transport and phosphorylation by the *Dorea* PTS complemented with the *E. coli* HPr, FPr and EI component. As shown in panel a, kestose can be phosphorylated at different carbon positions, while only one form of phosphorylated kestose is represented here.

Based on the above-described results, the I9 PTS\_EIIBCA encoding the three well known domains of PTS transporters (A, B and C domains) appears to be responsible for kestose internalization and phosphorylation by *E. coli*. As explained above, phosphorylation requires a multiproteic machinery to transfer the phosphoryl group from intracellular PEP to the carbohydrate bound to the EIIC domain of the PTS (41). Because the I9 locus lacks the EI and HPr components of the PTS (while being present elsewhere in the genome of *Dorea longicatena* DSM 13814, with the EI and HPr conserved histidine residues required for phosphate transfer), we hypothesized that the *E. coli* PTS machinery could complement I9 PTS\_EIIBCA to achieve kestose phosphorylation. To investigate this question, we verified the ability of *E. coli* K-12 BW25113 (KEIO collection strains) single-knockout mutants for glucose and fructose-specific PTS genes transformed with fosmid I9min\_PTS/GH32 to grow on a mixture of FOS, from which only kestose is utilized, as previously demonstrated above in the present paper. The *ptsI* (encoding for EI) defective strain was unable to utilize FOS as carbon sources (**Figure 6b**), demonstrating that this *E. coli* enzyme is required to ensure FOS-PTS functionality. Indeed, EI is involved in the first step of phosphorylation from PEP, which explains why deletion of *ptsI* in the KEIO strain completely abolished its growth on glucose and fructose (see **Figure S4**). In contrast, mutants defective for *ptsH* (encoding histidine protein HPr, which transfers the phosphate group from EI to the EIIA component of the glucose PTS) or *fruB* (encoding HPr-like FPr in the fructose PTS) were able to grow on FOS, even though their growth ability was reduced compared to that of the wild-type strain transformed with I9min\_PTS/GH32 (**Figure 6b**). Similarly, deletion of *ptsH* or *fruB* in the KEIO strain resulted in a delayed lag phase when grown on glucose and fructose, respectively (**Figure S4**), indicating that HPr could compensate FPr deficiency, and reversely. We thus conclude that *Dorea* PTS\_EIIBCA requires the *E. coli* EI to transfer phosphate from PEP to *E. coli* HPr (which can also be replaced, less effectively, by FPr), which, in turn, phosphorylates the I9 EIIA domain. Phosphate is then transferred to the I9 EIIB domain, and, ultimately, to the kestose molecules internalized through the I9 EIIC component (**Figure 6c**).

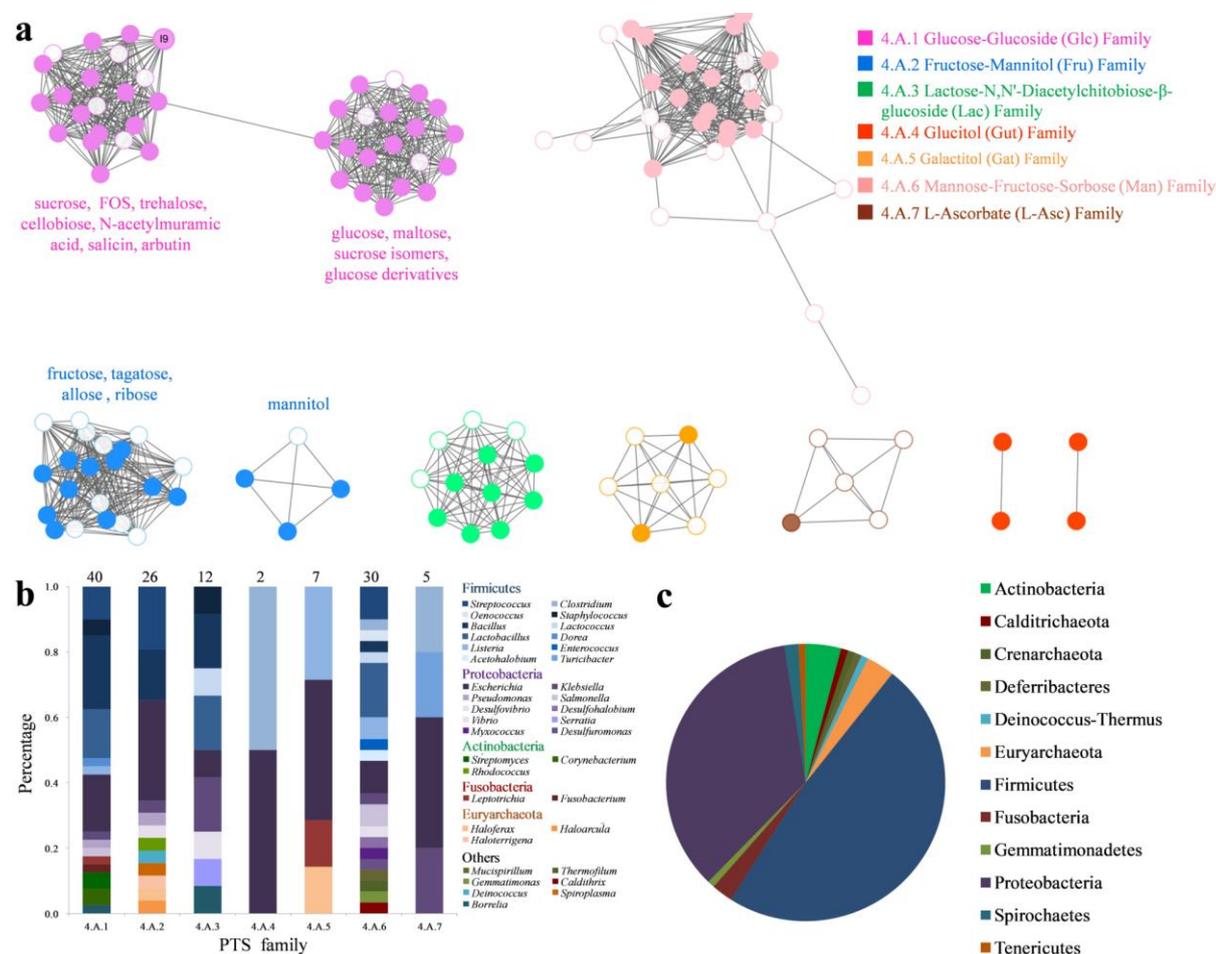
## PTS structural, functional and taxonomical diversity

By analyzing the specificity of *Dorea* PTS\_EIIBCA, we extended the panel of characterized glycoside transporters, which is still limited. Prior to this study, only two PTS (PTS1 and PTS26 from *L. plantarum* ST-III) were known to internalize FOS (37). However, these proteins are not listed in the Transporter Classification Database (TCDB) (69). In the TCDB, based on functional and phylogenetic information about the EIIC domain (as in the first classification of PTS proposed by Nguyen et al. (70)), PTSs have been classified into seven families of phosphotransfer-driven group translocators, consisting of the glucose-glycoside (4.A.1, Glc), fructose-mannitol (4.A.2, Fru), lactose-N,N-diacetylchitobiose (4.A.3, Lac), glucitol (4.A.4, Gut), galactitol (4.A.5, Gat), mannose-fructose-sorbose (4.A.6, Man), and L-ascorbate (4.A.7, L-Asc) families.

To locate the FOS transporters in the diversity of PTS and to investigate the sequence-function relationships of these proteins, we constructed a sequence similarity network (SSN) from the 125 sequences of EIIC component retrieved from TCDB (120 sequences for 117 PTS, since 3 have 2 EIIC domains), the literature (4 sequences) and the present study (**Figure 7a**). In a SSN network, nodes (representing protein sequences) in the same cluster are connected with an edge if they share a percentage of identity higher than a threshold value, providing a more visually tractable view of divergent proteins than alignment-based dendrograms or phylogenetic trees [44]. First, despite the low sequence identity threshold used to link the nodes in the sequence network to form clusters (21%), we observed that the 3 FOS PTS (one from *Dorea* and two others from *L. plantarum* DT-III) do not belong to the cluster of sequences containing the members of the Fru family 4.A.2. Even though FOS contain more fructosyl than glucosyl moieties (one terminal glucosyl moiety in a FOS molecule, regardless of its degree of polymerization), the FOS PTS belong to the Glc family 4.A.1. This could correlate to the fact that the *Dorea* PTS requires the terminal glucosyl moiety of FOS to bind and internalize its substrate.

Second, we observed that the two large Glc and Fru PTS families are subdivided into two clusters each in the SSN. The 4.A.1 Glc family contains two subfamilies in the TCDB, each corresponding to one of the SSN clusters: (i) subfamily 4.A.1.1, named “glucose subfamily” in the TCDB, contains in fact PTS specific to glucose, its derivatives (such as N-acetylglucosamine), and also  $\alpha$ -glucosides like maltose and sucrose isomers; (ii) subfamily 4.A.1.2, named “glucoside subfamily”, contains PTSs targeting  $\alpha$ - and  $\beta$ -glucosides, such as

sucrose (GF), trehalose ( $\alpha$ -1,1-linked GG, which has a close structural similarity with sucrose (71)), FOS (GFn), cellobiose ( $\beta$ -1,4-linked GG) and other  $\beta$ -glucosides (salicilin, arbutin), and N-acetylmuramic acid. If the present SSN clusters correspond to the TCDB 4.A.1 subfamilies, they are not the same as those previously proposed by Nguyen et al., who categorized the proteins now listed in the 4.A.1 family into the Glucose (Glc) and  $\beta$ -Glucoside (Bgl) families (70). Regarding the 4.A.2 Fru family, for which no subfamily is described in the TCDB, one SSN cluster contains the transporters targeting fructose, tagatose, ribose and allose, while the other one, clearly separated, contains mannitol-PTS. In this case, this is consistent with the findings of Nguyen et al (2006), who described the mannitol family apart from Fru (70).



**Figure 7. Functional, structural and taxonomical diversity of PTS.**

(a) Sequence Similarity Network (SSN) of PTS-EIIC domains from 125 proteins, retrieved from the Transporter Classification database, from the literature and present study, constructed with an initial score of  $10^{-5}$ . The nodes are connected by an edge if the pairwise

sequence identity is  $\geq 21.3\%$ . The corresponding substrates were annotated under each cluster of the 4.A.1 and 4.A.2 families. Unfilled nodes indicate sequences for which the functions were inferred by transcriptomics, while filled ones correspond to the functions validated biochemically or by gene deletion. Taxonomic distribution at genus (**b**) and phylum (**c**) levels of the characterized PTS transporters.

From these results, we conclude that there is no longer a link between the names of the families listed in the available PTS classifications and the function of their constituent members. Improvements in transporter classification will be proposed in the following section.

Since phylogenetic links are one of the bases of the existing transporter classifications, we also analyzed the taxonomical diversity of the 125 known PTS\_EIIC sequences. They represent 37 genera, classified into 12 distinct phyla (**Figure 7b** and **c**). For each PTS family, the taxonomic diversity is wide-ranging, with no family assigned to a specific phylum or genus. Of 125 sequences, more than half were assigned to the four genera *Escherichia* (20%), *Streptococcus* (10%), *Bacillus* (13%) and *Lactobacillus* (11%). Consequently, more than 80% belong to only two phyla, Firmicutes (48%) and Proteobacteria (35%). There is a similar distribution of characterized PTS between Gram-negative and Gram-positive bacteria. The present taxonomic distribution of PTS is highly biased by sequencing and biochemical characterization efforts which are unequal in their coverage of actual bacterial diversity. For example, characterized PTSs are absent from the Bacteroidetes phylum, while in the Pfam 32.0 database, we found 10 Bacteroidetes sequences harboring a PTS\_EIIC Pfam motif (PF13303) (72). In addition, the characterized PTSs appear to be very abundant in Firmicutes (48%) (**Figure 7c**), while this phylum accounts for just 29% of the sequenced bacterial genomes (versus 54% for Proteobacteria) referenced in the Genomes Online database (GOLD) (73). In any event, though, there are still many transporters to be discovered in the near future thanks to powerful tools such as functional metagenomics and novel generic biochemical approaches, such as those described here.

## Discussion

## Biochemical characterization of a FOS-specific PTS transporter

The PTS is a multiple component carbohydrate transport and phosphorylation system composed of two cytoplasmic phosphotransferases, enzyme I (EI) and histidine-containing phosphocarrier protein (HPr), which are common in most systems, and one multiprotein enzymes II complex (EII) including two phosphotransferase proteins (EIIA and EIIB) and one integral membrane protein (EIIC), which is usually specific for one substrate or a small group of closely related carbohydrates. For example, 21 different EII complexes were identified in *E. coli* as being involved in the transport of about 20 different carbohydrates, with only 1 EI and 1 HPr proteins, and five EI and six HPr paralogues, whose function are still to be uncovered (43). PTSs are thus major carbohydrate transport systems in certain bacteria, but their diversity and functionalities have been insufficiently investigated. Today, we have just 122 PTSs characterized (either biochemically or by transcriptomics and/or gene deletion), targeting a substrate range (mainly monosaccharides and some glucosides and N-acetylglucosides, fructosides, and galactosides) that is highly restricted compared to the possible diversity of oligosaccharides (74).

In the present study, we characterized a PTS\_EIIBCA complex involved in FOS utilization by both uncultured and cultured *Dorea* strains. Previously, several FOS PTSs have been identified from lactic acid bacteria and characterized by growth tests and transcriptomic analysis (75, 76), sometimes also with knock-out experiments (37), or by measurement of the extracellular composition by anionic exchange chromatography (45). However, the precise molecular mechanism and the exact length of the oligosaccharides that could be bound and internalized were not deeply investigated due to the technical bottlenecks inherent in transporter characterization. Here, we developed a highly generic, sensitive and quantitative method based on HPAEC-PAD analysis of oligosaccharides unbound to crude membranes or to entire cells (native or recombinant). This approach, which requires no protein complex dissociation or purification steps, is compatible with the use of unlabeled natural substrates to quantify binding, transport and metabolization rates. It thus effectively replaces the use of radiolabeled substrates, which have been widely used until now for transporter characterization. Here, we were able to determine the contributions of substrate binding and transport to the oligosaccharide metabolization process. We proved that *Dorea* PTS\_EIIBCA is able to transport sucrose and especially kestose, with an affinity for kestose binding five times better than its affinity for uptake.

Moreover, carbohydrate uptake through a PTS implies a phosphorylation cascade. PTS-based phosphorylation of oligosaccharides with a DP of more than 2 had never been shown. Here, we proved that a DP3 oligosaccharide can be phosphorylated by *Dorea* PTS\_EIIBCA and that its specificity for phosphorylation is the same as that for binding and translocation. We demonstrated this directly *in cellulo*, in contrast to the previous studies, in which phosphorylation reactions were performed *in vitro* by adding purified phosphotransferases and EIIAB to the EIIC component (77). By IC-ESI-HRMS analysis and use of *E. coli* single knock-out mutants, we proved the entire functionality of the *Dorea* PTS\_EIIBCA complex in *E. coli*, thanks to efficient complementation by the host's EI and HPr (or, alternatively, FPr) phosphotransferases.

Several PTS\_EII complexes involved in cellobiose utilization by Gram-positive bacteria have previously been successfully expressed in an *E. coli* host strain (78, 79), and proven *in vitro* to be complemented by the *E. coli* EI and HPr PTS components (80). The difference in cell membrane organization between Gram positives and *E. coli* thus appears not to have an impact on the functionality of recombinant PTS\_EIIBCA complexes, even though the localization of each component of Gram-positive PTS in *E. coli* is still an open question.

## Revisiting the PTS classification

In the last few years, a huge mass of omics data has accumulated (81), and microfluidic approaches have been developed to rapidly identify the substrates targeted by (meta)genomic loci (82). Meanwhile, an increasing number of tridimensional transporter structures has been solved (83), and the panel of biochemical methods for transporter characterization has been extended, as described in the present paper. It is thus likely that significant progress will be made over the next few years in the understanding of transporter structure-function relationships. In the present paper, we investigated this question by analyzing the diversity of sequences and substrate specificities of known PTS. We showed that, as the number of transporters with characterized specificities increases, there is no longer a link between the names of PTS families and their functions (for example, subfamily 4.A.1.1, named 'glucose subfamily' in the TCDB, contains PTSs specific to N-acetylglucosamine and to some glucosides, while the subfamily 4.A.1.2, named 'glucoside subfamily', contains PTS specific to N-acetylmuramic acid). The present classification is thus quite confusing. Instead, we propose a transporter classification based on similarities of structures and mechanism, with families containing members with one or more substrate specificities. This kind of

classification, with no specificity-related constraints, does exist for enzymes. For example, the CAZy database, based on this principle, is a reference for carbohydrate active enzyme classification (84), even though it is complementary to the EC classification (the EC numbers now appear in the CAZy database for each biochemically characterized enzyme). For transporter classification, we recommend the continued use of the TCDB, which is a mine of information but with caution in respect of the level of classification. Indeed, of the five components of the transporter classification identifier (TCID), only the first two relate to structure and mechanism: the first number corresponds to the transporter class (i.e., channel, carrier [porter], primary active transporter, or group translocator), and the following letter to the transporter subclass which, in the case of primary active transporters, refers to the energy source used to drive transport. The last three TCID components are, in most cases, related to substrates (at least for PTS) in the present TCDB. Classification into superfamilies, families and subfamilies will thus require further global/local sequence analyses in order to identify specific common structural traits within transporter groups. Of course, listing the known substrates for each functionally characterized transporter is still highly valuable, as it will provide a clear view of the functional diversity in each of the structurally defined groups. Thus, major advances can be made to elucidate, and, ultimately, predict, the structural determinants of transporter specificity.

### **Significance of unlocking prebiotic metabolic pathways from the human gut microbiome and implications for gut health**

FOS, one of the most commonly used and studied prebiotics, has been shown to stimulate the growth of beneficial gut bacteria such as bifidobacteria and lactic bacteria (30) but also of some pathogenic *E. coli* strains (13). Previously most of the FOS metabolization pathways have been discovered from cultured bacteria, but the emergence of high-throughput functional metagenomic approaches has made it possible to study more extensive sequence spaces and highlighted new genomic loci involved in FOS metabolism by uncultured bacteria (20). Here, we studied a locus assigned to *Dorea longicatena* DSM 13814 (99% identity). However, we could not exclude that it might come from a different bacterium, which would have acquired this DNA fragment by horizontal gene transfer.

Previously, *D. longicatena* was found to be stimulated by short-chain FOS, with a poor ability to use long-chain  $\beta$ -fructans (14, 19). This is consistent with the specificity of the *Dorea* locus characterized here for kestose, a DP3 FOS, which is by far the most abundant FOS in wheat

(85), a staple food for Westerners, and a major component of the FOS mixtures used for dietary supplementation (86). It does not exclude the possibility, however, that other *D. longicatena* loci could be involved into FOS metabolization, since we found three other loci in strain DSM13814 harboring an I9\_GH32 homolog (33 to 42% identity) with either an ABC transporter or a PTS. *Dorea* has been shown in several studies to be positively associated with intestinal permeability in individuals with alcohol dependence, with multiple episodes of sclerosis relapse (87, 88), and with IBS (15, 16). By contrast, in several studies, it was negatively associated with IBD. This dysbiosis, which is even more pronounced in a common complication in IBD, *Clostridium difficile* infection, is likely to be due to administration of antibiotics (89, 90). At the species level, *D. longicatena* is also negatively associated with Crohn's disease (91), and positively with remission after surgery (92). Here however, we showed that the *Dorea* FOS utilization locus is twice as abundant in the microbiome of patients suffering from IBD than in that of healthy individuals. It is thus difficult to correlate this to the results of genera and species abundance in the context of IBD, especially since the present study and previous ones showed that Gram-positive and -negative PULs are frequently exchanged between gut bacteria (20, 27), providing the microbiota with metabolic flexibility and robustness. Taking that into account, it appears more informative to focus on pathology biomarkers of the microbiome and their specific functions, rather than on bacterial species whose genomic composition and metabolic abilities are far from being fully characterized and which are, in addition, variable.

The focus should now be on the function of the PTS characterized here, and links to gut health. Based on the fact that prebiotics are characterized as stimulating the growth of beneficial gut bacteria, the effect of FOS dietary supplementation in decreasing Crohn's disease activity has been tested (93). Nevertheless, the latest results from a randomized, double-blind, placebo-controlled trial revealed no significant benefit of FOS supplementation in patients with active Crohn's disease, other than an induction of immunoregulatory dendritic cell responses (94). To date, only a small number of clinical trials have investigated prebiotic supplementation as treatment for chronic gastrointestinal disorders (including IBD and IBS). It remains a controversial issue with conflicting results, depending on the type and dose of the prebiotic used (95). In any case, the biochemical characterization of prebiotic utilization machineries, such as the FOS-targeting transport system highlighted here both as a biomarker of diet and pathology, is of strategic importance in formulating the mechanistic bases for

microbiome functional engineering, such as through the design of future specific probiotic-prebiotic combinations to maximize host benefits.

## Materials and methods

### *Escherichia coli* strains

The metagenomic clone I9 includes contigs I9a (GenBank accession number HE717008.1) and I9b (GenBank accession number HE717009.1). This *E. coli* clone comes from a metagenomic library constructed by sampling the ileum mucosal microbiota of a 51-year-old male patient undergoing colonoscopy and surgery for suspected lower colon cancer (20). The metagenomic DNA fragments (30 to 40 kbp) were cloned into the pCC1FOS fosmid and transformed into the EPI100 *E. coli* strain (Epicentre Technologies). All the I9a variants were constructed using the In-Fusion HD Cloning Plus kit (Clontech) following the instructions in the user manual. For each variant, amplification started 600 bp to 1 kb upstream of the first gene of the target locus, to involve potential promoter sequences which might be necessary for gene expression (52). The primers used in this study are listed in **Table S1**. Mutants were confirmed by new generation sequencing, performed by the GeT-Biopuces Platform (Toulouse) using the Ion Torrent S5 System. Read assembly was performed using Masurca (<http://www.genome.umd.edu/masurca.html>). The assembled contigs were cleaned from the pCC1FOS vector sequence using Crossmatch (<http://bozeman.mbt.washington.edu/phredphrapconsed.html>).

### Growth conditions

All the *E. coli* variants constructed in this study were grown in M9 medium ( $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ , 17.4 g/l;  $\text{KH}_2\text{PO}_4$ , 3.03 g/l; NaCl, 0.51 g/l;  $\text{NH}_4\text{Cl}$ , 2.04 g/l;  $\text{MgSO}_4$ , 0.49 g/l;  $\text{CaCl}_2$ , 4.38 mg/l;  $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$ , 15 mg/l;  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 4.5 mg/l;  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ , 0.3 mg/l;  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 1 mg/l;  $\text{H}_3\text{BO}_3$ , 1 mg/l;  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ , 0.4 mg/l;  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 3 mg/l;  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 0.3 mg/l; thiamine, 0.1 g/l; and leucine, 0.02 g/l) supplemented with 12.5 mg/l chloramphenicol. The carbohydrates were diluted in M9 medium at 0.5% (w/v). The cultures were inoculated at an initial optical density at 600 nm ( $\text{OD}_{600 \text{ nm}}$ ) of 0.05 from precultures in LB.

Growth dynamic were analyzed at 37 °C on various oligosaccharides and polysaccharides: FOS (Beghin Meiji; containing 5.3% (wt/wt) sucrose, 40.0% (wt/wt) 1-kestose, 46.0% (wt/wt) nystose, and 8.7% (wt/wt) fructosyl-nystose, quantified by HPAEC-PAD analysis, as described further in this section), sucrose (Sigma), 1-kestose (GF<sub>2</sub>; Wako Chemicals), nystose

(GF<sub>3</sub>; Wako Chemicals), fructofuranosyl-nystose (GF<sub>4</sub>; Wako Chemicals), inulotriose (Megazyme), inulin from dahlia tuber (Sigma) and levan from *Erwinia herbicola* (Sigma). The optical density of the cultures was measured at 600 nm using a plate reader (Infinite M200pro; TECAN).

To evaluate the role of endogenous *E. coli* PTS genes in the phosphorylation of kestose, *E. coli* K-12 BW25113 strain and its derived single-knockout mutants (from the KEIO library) were employed (68). I9min\_PTS/GH32 fosmid was electro-transferred into *E. coli* K-12 BW25113 single-knockout mutants for the *ptsI*, *ptsH* and *fruB* genes, respectively. Their growth abilities in M9 medium supplemented with FOS, glucose or fructose, were analyzed. The wild type KEIO strain transformed with the I9min\_PTS/GH32 fosmid and the empty fosmid pCC1FOS were used as positive and negative controls, respectively. The cultures were inoculated at an initial optical density at OD<sub>600</sub> of 0.05 from precultures in LB, and cell growth was monitored by measuring the OD<sub>600</sub> over 48 h at 37 °C using the FLUOStar Optima (BMG Labtech).

### **FOS utilization analysis**

In order to determine how the different oligosaccharides in the FOS mixture were utilized by *E. coli*, HPAEC-PAD was used to measure the concentration of residual carbohydrates in the culture media filtered at 0.20 µm. Supernatants were analyzed on a Dionex ICS-3000 system using a CarboPac PA100 4×250 column. Carbohydrates were eluted at 30 °C, at a flow rate of 1 ml/min with the following multistep gradient: 0-30 min (0-60% B), 30-31 min (60-0% B) and 31-36 min (0% B). Solvents were 150 mM NaOH (eluent A) and 150 mM NaOH, 500 mM CH<sub>3</sub>COONa (eluent B).

### **Enzymatic assays**

FOS hydrolysis assays were performed by using the cytoplasmic fraction of the I9min\_GH32 variant and the extracellular fraction. The I9min\_GH32 variant was grown overnight at 37 °C in LB supplemented with chloramphenicol 12.5 mg/l. The crude cytoplasmic extracts were obtained as follows: after centrifugation, cells were concentrated in 50 mM potassium phosphate buffer pH 7.0 to an optical density at 600 nm (OD<sub>600</sub>) of 80, containing 0.5 g/l lysozyme, and incubated at 37 °C for 1h. Cell lysis was completed with one freeze (at -80 °C) and thaw (at 37 °C) cycle. Cell debris were centrifuged at 15,000 g for 30 min, and the cytoplasmic extracts were filtered with a 0.20 mm Minisart RC4 syringe filter. Enzymatic

assays were performed at 37 °C with either the cytoplasmic extracts or the culture supernatant and 0.5% (w/v) FOS, sucrose, 1-kestose, nystose, fructofuranosyl-nystose, inulotriose, inulin or levan. Sampling was performed at initial time and at 24 h. Reactions were stopped by heating at 95 °C for 5 min. Samples were analyzed by HPAEC-PAD as described above.

## **Gene expression analysis**

Total RNAs were extracted using the RNeasy Mini Kit (Qiagen) from three independent cultures of the I9 clone in mid-exponential phase in LB. The RNase-Free DNase Set (Qiagen) was used to remove contaminating DNA. RNAs were quantified using a NanoDrop™ (ThermoFisher Scientific), and the Bioanalyzer RNA kit (Agilent Technologies) was used to control their quality. Then, 5 µg of total RNA were retrotranscribed using SuperScript™ II Reverse Transcriptase (ThermoFisher Scientific) and random primers (Invitrogen). Synthesized cDNA was purified using Illustra™ MicroSpin™ G-25 columns (GE Healthcare). The primers used for real-time quantitative PCR (qPCR) were designed using the Primer3Plus web interface with lengths from 18 to 22 bases, a GC content of more than 50%, melting temperatures between 55 and 65 °C and an amplification size from 83 to 148 bases (**Table S1**). Prior to RNA expression analysis, primer specificity was checked on the genomic DNA of the *E. coli* I9 clone. qPCR was carried out on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) and monitored using the CFX Maestro™ software. Data were normalized based on the expression of the housekeeping gene integration host factor β-subunit (*ihfB*), which is one of the commonly used reference gene in *E. coli* owing to its constant expression throughout growth (96). Relative expression was calculated as  $2^{-\Delta Ct}$ , in which  $\Delta Ct$  was obtained by subtracting the average cycle threshold values (Ct, corresponding to the number of cycles required for the fluorescent signal to cross the threshold) of the housekeeping gene from those of the gene of interest (97). For each of the three biological replicates, four technical replicates were performed, corresponding to four different DNA dilutions.

## **Carbohydrate uptake and binding rate quantification**

Overnight cultures of the I9min\_PTS variant in LB medium supplemented with 12.5 mg/l chloramphenicol were centrifuged. Cells were kept at 4 °C no longer than four hours before use. For uptake experiments, cell pellets were washed three times in 50 mM Tris-buffered saline (TBS) pH 7.0 and then resuspended in TBS to reach an OD<sub>600 nm</sub> of 40. Sucrose, fructose, glucose, kestose, nystose, fructofuranosyl-nystose, or inulotriose was added at a final concentration ranging from 0.5 to 100 mM. All reactions were incubated at 37 °C, with

shaking at 200 rpm. After different incubation times, from 5 min to 5 hours, samples of 0.2 ml were filtered with a 0.20 mm Minisart RC4 syringe filter. The concentrations of unbound carbohydrates in the supernatant were determined by HPAEC-PAD as described above. For binding experiments, cell pellets were re-suspended in TBS, containing 0.5 g/l lysozyme and incubated at 37 °C for 1h. Cell lysis was completed with one freeze (at -80 °C) and thaw (at 37 °C) cycle. Cell membrane fractions were isolated by centrifugation at 15,000 g for 30 min, washed three times with 50 mM Tris buffer, pH 7.0 and re-suspended in the same buffer to reach an OD<sub>600 nm</sub> of 40. The later steps were the same as for the uptake experiments. The EPI clone was used as negative control in all these experiments.

### **Kestose phosphorylation analysis**

One hundred µL of an overnight culture of the knocked-out *E. coli* strains in LB were used to inoculate 50 ml baffled shake flasks containing 10 ml of M9 medium supplemented with glucose or kestose as carbon sources. The flasks were incubated for 24 h at 37 °C with orbital shaking at 220 rpm. Cells were harvested by centrifugation for 10 min at 2,000g at room temperature, washed with diluted M9 and used to inoculate 50 ml baffled shake flasks at an OD<sub>600 nm</sub> of 0.1, containing M9 supplemented with the same carbon source. Cell growth, performed at 37 °C and 220 rpm, was monitored by measuring OD<sub>600 nm</sub> with a cell density meter (Ultrospect 10, Amersham BioSciences). When culture reached the exponential phase, 120 µl of culture medium were withdrawn and vigorously mixed with 1.25 ml of an acetonitrile/methanol/H<sub>2</sub>O (4:4:2) solution pre-cooled at -20 °C to rapidly quench metabolic activity and extract metabolites (67). The samples were kept at -20 °C for 20 minutes, and then centrifuged at 10,000 g at 4 °C for 10 minutes to remove cell debris. The supernatant was dried using a SpeedVac (SC110A SpeedVac Plus, ThermoSavant) under vacuum and then stored at - 80 °C until analysis.

The samples were dissolved in 120 µL of mQ water and then analyzed by ion chromatography (Thermo Scientific Dionex ICS-5000+ system, Dionex) coupled to an LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with an electrospray ionization probe (IC-ESI-HRMS). A revised ion chromatography method was used here (98). The KOH gradient was modified as follows: 0 min, 0.5 mM; 1 min, 0.5 mM; 9.5 min, 4.1 mM; 12.5 min, 30 mM; 24 min, 50 mM; 36 min, 60 mM; 36.1 min, 90 mM; 43 min, 90 mM; 43.5 min, 0.5 mM; 48 min, 0.5 mM. The gradients were all linear. For background suppression, an Anionic Electrolytically Regenerated Suppressor (AERS 300-2 mm, Dionex) was used. The

device was operated at a constant 87 mA electrolysis current, in external water mode with ultrapure water and regenerant delivered by an external AXP pump at a flow rate of 1 ml/min. The volume of the injected sample was 35  $\mu$ l. Mass spectrometry analysis was performed in the negative FTMS mode at a resolution of 30,000 in full scan mode, with the following source parameters: capillary temperature, 350  $^{\circ}$ C; source heater temperature, 300  $^{\circ}$ C; sheath gas flow rate, 50 a.u. (arbitrary unit); auxiliary gas flow rate, 5 a.u.; S-Lens RF level, 60%; and ion spray voltage, 3.5 kV. The data were acquired using the Xcalibur software (Thermo Fisher Scientific).

## **Bioinformatic analysis**

The LipoP 1.0 server was used to predict lipoprotein and signal peptides (<http://www.cbs.dtu.dk/services/LipoP/>) (99). Transmembrane domains were predicted using the TMHMM Server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) (100). Protein function was predicted by analyzing conserved domains in the NCBI's Conserved Domain Database (101). Sequence similarity network (SSN) were generated via the Enzyme Similarity Tool (EFI-EST, <https://efi.igb.illinois.edu/efi-est/>) (102) by inputting 125 amino-acid sequences of EIIC proteins extracted from the Transporter Classification Database (TCDB, <http://www.tcdb.org/>) (103), the literature (37, 45, 104) and present study. The SSNs were generated with edge values (E-values) of  $10^{-3}$  to  $10^{-20}$  and the alignment scores were further refined after the sequence networks were visualized in Cytoscape 3.6.1 (105). An E-value threshold of  $10^{-5}$  corresponding to 21.3% sequence identity was finally used to visualize the network.

Sequences of the I9a ORFs were searched by BLASTP analysis (E-value = 0, identity  $\geq$  90%, no coverage threshold) against the translated catalogue of 9.9 million reference genes constructed from the fecal metagenome of 1,267 subjects from USA, China, and Europe (46, 106). The microbial gene richness in the human gut microbiome was determined using the abundance and frequency matrix in the 1,267 subjects (<http://meta.genomics.cn/meta/dataTools>).

## **Data availability**

The nucleotide sequence of the I9 metagenomic DNA is available in the GenBank database under accession numbers HE717008.1 for contig I9a and HE717009.1 for contig I9b.

## Acknowledgments

We thank Matthieu Arlat for the fruitful discussions on this project.

The analytic work was carried out with the equipment of the TBI-ICEO facility, dedicated to enzyme screening, and that of Metatoul (Metabolomics & Fluxomics Facilities, Toulouse, France, [www.metatoul.fr](http://www.metatoul.fr)). ICEO is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund and the Institut National de la Recherche Agronomique (INRA), and MetaToul is part of the MetaboHUB-ANR-11-INBS-0010 national infrastructure ([www.metabohub.fr](http://www.metabohub.fr)). This research was funded by the European Union's framework programme Horizon 2020 (MSCA-IF-2015\_707457, CaTSYS, and LEIT-BIO-2015-685474, Metafluidics). ZW is supported by INSA Toulouse and China Scholarship Council in the frame of the “CSC – UT/INSA Program”.

## References

1. Bindels LB, Delzenne NM, Cani PD, Walter J. 2015. Towards a more comprehensive concept for prebiotics. *Nat Rev Gastroenterol Hepatol* 12:303–310.
2. Wan MLY, Ling KH, El-Nezami H, Wang MF. 2019. Influence of functional food components on gut health. *Crit Rev Food Sci Nutr* 59:1927–1936.
3. Gibson GR, Hutkins R, Sanders ME, Prescott SL, Reimer RA, Salminen SJ, Scott K, Stanton C, Swanson KS, Cani PD, Verbeke K, Reid G. 2017. Expert consensus document: the International Scientific Association for Probiotics and Prebiotic (ISAPP) consensus statement on the definition and scope of prebiotics. *Nat Rev Gastroenterol Hepatol* 14:491–502.
4. Bornet FRJ, Brouns F. 2002. Immune-stimulating and gut health-promoting properties of short-chain fructo-oligosaccharides. *Nutr Rev* 60:326–334.
5. Sangeetha PT, Ramesh MN, Prapulla SG. 2005. Recent trends in the microbial production, analysis and application of fructooligosaccharides. *Trends Food Sci Technol* 16:442–457.
6. Akkerman R, Faas MM, de Vos P. 2019. Non-digestible carbohydrates in infant formula as substitution for human milk oligosaccharide functions: effects on microbiota and gut maturation. *Crit Rev Food Sci Nutr* 9:1486–1497.
7. Grand View Research. 2015. Inulin market by application (food & beverage, dietary supplements, pharmaceuticals) is expected to reach USD 2.35 billion by 2020: Grand View Research, Inc.

8. Grand View Research. 2016. Fructooligosaccharides (FOS) market worth \$ 3.52 Billion By 2024.
9. Gentile CL, Weir TL. 2018. The gut microbiota at the intersection of diet and human health. *Science* 362:776–780.
10. Bruggencate SJM Ten, Bovee-oudenhoven IMJ, Lettink-wissink MLG, Katan MB, Meer R Van Der. 2006. Dietary fructooligosaccharides affect intestinal barrier function in healthy men. *J Nutr* 136:70–74.
11. Canfora EE, Jocken JW, Blaak EE. 2015. Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat Rev Endocrinol* 11:577–591.
12. Medina-Vera I, Sanchez-Tapia M, Noriega-López L, Granados-Portillo O, Guevara-Cruz M, Flores-López A, Avila-Nava A, Fernández ML, Tovar AR, Torres N. 2019. A dietary intervention with functional foods reduces metabolic endotoxaemia and attenuates biochemical abnormalities by modifying faecal microbiota in people with type 2 diabetes. *Diabetes Metab* 45:122–131.
13. Mao B, Li D, Zhao J, Liu X, Gu Z, Chen YQ, Zhang H, Chen W. 2015. *In vitro* fermentation of fructooligosaccharides with human gut bacteria. *Food Funct* 6:947–954.
14. Taras D, Simmering R, Collins MD, Lawson PA, Blaut M. 2002. Reclassification of *Eubacterium formicigenerans* Holdeman and Moore 1974 as *Dorea formicigenerans* gen . nov ., comb . nov ., and description of *Dorea longicatena* sp. nov. isolated from human. *Int J Syst Evol Microbiol* 52:423–428.
15. Rajilić-Stojanović M, Biagi E, Heilig GHJ, Kajander K, Kekkonen RA, Tims S, De Vos WM. 2011. Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* 141:1792–1801.
16. Saulnier D, Riehle K, Mistretta T, Diaz M, Mandal D, Raza S, Weidler E, Qin X, Coarfa C, Milosavljevic A, Petrosino J, Highlander S, Gibbs R, Lynch S, Shulman R, Versalovic J. 2011. Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome. *Gastroenterology* 141:1782–1791.
17. Harvie R, Walmsley R, Schultz M. 2017. “We are what our bacteria eat”: the role of bacteria in personalizing nutrition therapy in gastrointestinal conditions. *J Gastroenterol Hepatol* 32:352–357.

18. Chen T, Long W, Zhang C, Liu S, Zhao L, Hamaker BR. 2017. Fiber-utilizing capacity varies in *Prevotella*- versus *Bacteroides*-dominated gut microbiota. *Sci Rep* 7:1–7.
19. Respondek F, Gerard P, Bossis M, Boschhat L, Bruneau A, Rabot S, Wagner A, Martin JC. 2013. Short-chain fructo-oligosaccharides modulate intestinal microbiota and metabolic parameters of humanized gnotobiotic diet induced obesity mice. *PLoS One* 8:e71026.
20. Cecchini DA, Laville E, Laguerre S, Robe P, Leclerc M, Doré J, Henrissat B, Remaud-Simón M, Monsan P, Potocki-Veronèse G. 2013. Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One* 8:e72766.
21. Chanteloup NK, Porcheron G, Delaleu B, Germon P, Schouler C, Moulin-Schouleur M, Gilot P. 2011. The extra-intestinal avian pathogenic *Escherichia coli* strain BEN2908 invades avian and human epithelial cells and survives intracellularly. *Vet Microbiol* 147:435–439.
22. Schouler C, Taki A, Chouikha I, Moulin-Schouleur M, Gilot P. 2009. A genomic island of an extraintestinal pathogenic *Escherichia coli* strain enables the metabolism of fructooligosaccharides, which improves intestinal colonization. *J Bacteriol* 91:388–393.
23. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:490–495.
24. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* 6:121–131.
25. Koropatkin NM, Cameron EA, Martens EC. 2012. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 10:323–335.
26. Foley MH, Cockburn DW, Koropatkin NM. 2016. The *Sus* operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell Mol Life Sci* 73:2603–2617.
27. Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C, Cantarel BL, Coutinho PM, Henrissat B, Leclerc M, Doré J, Monsan P, Remaud-Simeon M, Potocki-Veronese G. 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res* 20:1605–1612.
28. Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, Henrissat B. 2018. PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res* 46:D677–D683.
29. Sheridan PO, Martin JC, Lawley TD, Browne HP, Harris HMB, Bernalier-Donadille A,

- Duncan SH, O'Toole PW, Scott KP, Flint HJ. 2016. Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic *Firmicutes*. *Microb Genomics* 2:e000043.
30. Goh YJ, Klaenhammer TR. 2015. Genetic mechanisms of prebiotic oligosaccharide metabolism in probiotic microbes. *Annu Rev Food Sci Technol* 6:137–156.
  31. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firkbank SJ, Bolam DN, Sonnenburg JL. 2010. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* 141:1241–1252.
  32. Joglekar P, Sonnenburg ED, Higginbottom SK, Earle KA, Morland C, Shapiro-Ward S, Bolam DN, Sonnenburg JL, Abbott W, Koropatkin N. 2018. Genetic variation of the SusC/SusD homologs from a polysaccharide utilization locus underlies divergent fructan specificities and functional adaptation in *Bacteroides thetaiotaomicron* strains. *mSphere* 3:e00185-18.
  33. Boger MCL, Lammerts van Bueren A, Dijkhuizen L. 2018. Cross-feeding amongst probiotic bacterial strains on prebiotic inulin involving the extracellular exo-inulinase of *Lactobacillus paracasei* strain W20. *Appl Environ Microbiol* 84:e01539-18.
  34. Vijayaraghavan K, Yamini D, Ambika V, Sravya Sowdamini N. 2009. Trends in inulinase production a-review. *Crit Rev Biotechnol* 29:67–77.
  35. Barrangou R, Altermann E, Hutkins R, Cano R, Klaenhammer TR. 2003. Functional and comparative genomic analyses of an operon involved in fructooligosaccharide utilization by *Lactobacillus acidophilus*. *Proc Natl Acad Sci USA* 100:8957–8962.
  36. Linke CM, Woodiga SA, Meyers DJ, Buckwalter CM, Salhi HE, King SJ. 2013. The ABC transporter encoded at the pneumococcal fructooligosaccharide utilization locus determines the ability to utilize long- and short-chain fructooligosaccharides. *J Bacteriol* 195:1031–1041.
  37. Chen C, Zhao G, Chen W, Guo B. 2015. Metabolism of fructooligosaccharides in *Lactobacillus plantarum* ST-III via differential gene transcription and alteration of cell membrane fluidity. *Appl Environ Microbiol* 81:7697–7707.
  38. Cockburn DW, Koropatkin NM. 2016. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *J Mol Biol* 428:3230–3252.
  39. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. 2017. Polysaccharide utilization loci: fuelling microbial communities. *J Bacteriol* 199:e00860-16.

40. van de Guchte M, Blottière HM, Doré J. 2018. Humans as holobionts: implications for prevention and therapy. *Microbiome* 6:81.
41. Deutscher J, AkéFMD, Derkaoui M, Zébré AC, Cao TN, Bouraoui H, Kentache T, Mokhtari A, Milohanic E, Joyet P. 2014. The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. *Microbiol Mol Biol Rev* 78:231–256.
42. Keyhani NO, Wang LX, Lee YC, Roseman S. 2000. The chitin disaccharide, N,N'-diacetylchitobiose, is catabolized by *Escherichia coli* and is transported/phosphorylated by the phosphoenolpyruvate: glycolate phosphotransferase system. *J Biol Chem* 275:33084–33090.
43. Tchieu JH, Norris V, Edwards JS, Saier MH. 2001. The complete phosphotransferase system in *Escherichia coli*. *J Mol Microbiol Biotechnol* 3:329–46.
44. Mo S, Kim BS, Yun SJ, Lee JJ, Yoon SH, Oh CH. 2015. Genome sequencing of *Clostridium butyricum* DKU-01, isolated from infant feces. *Gut Pathog* 7:8.
45. Saulnier DMA, Molenaar D, De Vos WM, Gibson GR, Kolida S. 2007. Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays. *Appl Environ Microbiol* 73:1753–1765.
46. Li J, Wang J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, Juncker AS, Manichanh C, Chen B, Zhang W, Levenez F, Wang J, Xu X, Xiao L, Liang S, Zhang D, Zhang Z, Chen W, Zhao H, Al-Aama JY, Edris S, Yang H, Wang J, Hansen T, Nielsen HB, Brunak S, Kristiansen K, Guarner F, Pedersen O, Doré J, Ehrlich SD, Consortium M, Bork P, Wang J. 2014. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 32:834–841.
47. Sun D. 2018. Pull in and push out: mechanisms of horizontal gene transfer in bacteria. *Front Microbiol* 9:2154.
48. O'Connor Á. 2012. Bread consumption in the UK: what are the main attitudinal factors affecting current intake and its place in a healthy diet? *Nutr Bull* 37:368–379.
49. Notarnicola B, Tassielli G, Renzulli PA, Monforti F. 2017. Energy flows and greenhouse gases of EU (European Union) national breads using an LCA (Life Cycle Assessment) approach. *J Clean Prod* 140:455–469.
50. Lipoeto NI, Geok Lin K, Angeles-Agdeppa I. 2012. Food consumption patterns and nutrition

- transition in South-East Asia. *Public Health Nutr* 16:1637–1643.
51. Hu EA, Pan A, Malik V, Sun Q. 2012. White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. *BMJ* 344:e1454.
  52. Tauzin AS, Laville E, Xiao Y, Nouaille S, Le Bourgeois P, Heux S, Portais JC, Monsan P, Martens EC, Potocki-Veronese G, Bordes F. 2016. Functional characterization of a gene locus from an uncultured gut *Bacteroides* conferring xylo-oligosaccharides utilization to *Escherichia coli*. *Mol Microbiol* 102:579–592.
  53. Locher KP. 2016. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* 23:487–493.
  54. Rice AJ, Park A, Pinkett HW. 2014. Diversity in ABC transporters: Type I, II and III importers. *Crit Rev Biochem Mol Biol* 49:426–437.
  55. Solopova A, Bachmann H, Teusink B, Kok J, Kuipers OP. 2018. Further elucidation of galactose utilization in *Lactococcus lactis* MG1363. *Front Microbiol* 9:1803.
  56. Cao Y, Jin X, Levin EJ, Huang H, Zong Y, Quick M, Javitch JA, Rajashankar KR. 2011. Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature* 473:50–54.
  57. Martens EC, Chiang HC, Gordon JI. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 5:447–457.
  58. Schwalm ND, Groisman EA. 2017. Navigating the gut buffet: control of polysaccharide utilization in *Bacteroides* spp. *Trends Microbiol* 25:1005–1015.
  59. Tauzin AS, Kwiatkowski KJ, Orlovsky NI, Smith CJ, Creagh AL, Haynes CA, Wawrzak Z, Brumer H, Koropatkin NM. 2016. Molecular dissection of xyloglucan recognition in a prominent human gut symbiont. *MBio* 7:e02134-15.
  60. Culurgioni S, Harris G, Singh AK, King SJ, Walsh MA. 2017. Structural basis for regulation and specificity of fructooligosaccharide import in *Streptococcus pneumoniae*. *Structure* 25:79–93.
  61. Jeckelmann JM, Harder D, Mari SA, Meury M, Ucurum Z, Müller DJ, Erni B, Fotiadis D. 2011. Structure and function of the glucose PTS transporter from *Escherichia coli*. *J Struct Biol* 176:395–403.
  62. Keyhani NO, Wang L-X, Lee YC, Roseman S. 1996. The chitin catabolic cascade in the

- marine bacterium *Vibrio furnissii*. *J Biol Chem* 271:33409–33413.
63. Chen L-Q, Cheung LS, Feng L, Tanner W, Frommer WB. 2015. Transport of sugars. *Annu Rev Biochem* 84:865–894.
  64. Miller DM, Olson JS, Pflugrathlt JW, Quiochol FA. 1983. Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. *J Biol Chem* 258:13665–13672.
  65. Lendenmann U, Snozzi M, Egli T. 1999. Growth kinetics of *Escherichia coli* with galactose and several other sugars in carbon-limited chemostat culture. *Can J Microbiol* 46:72–80.
  66. Senn H, Lendenmann U, Snozzi M, Hamer G, Egli T. 1994. The growth of *Escherichia coli* in glucose-limited chemostat cultures: a re-examination of the kinetics. *Biochim Biophys Acta - Gen Subj* 1201:424–436.
  67. Millard P, Massou S, Wittmann C, Portais JC, L  sise F. 2014. Sampling of intracellular metabolites for stationary and non-stationary <sup>13</sup>C metabolic flux analysis in *Escherichia coli*. *Anal Biochem* 465:38–49.
  68. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2:0008.
  69. Saier MH. 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* 34:D181–D186.
  70. Nguyen TX, Yen MR, Barabote RD, Saier MH. 2006. Topological predictions for integral membrane permeases of the phosphoenolpyruvate:sugar phosphotransferase system. *J Mol Microbiol Biotechnol* 11:345–360.
  71. Tarpan MA, De Cooman H, Sagstuen E, Waroquier M, Callens F. 2011. Identification of primary free radicals in trehalose dihydrate single crystals X-irradiated at 10 K. *Phys Chem Chem Phys* 13:11294–11302.
  72. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432.
  73. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Katta HY, Mojica A, Chen IMA, Kyrpides NC, Reddy TBK. 2019. Genomes OnLine database (GOLD) v.7: updates and new

- features. *Nucleic Acids Res* 47:D649–D659.
74. Lap ¢bie P, Lombard V, Drula E, Terrapon N, Henrissat B. 2019. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat Commun* 3:2043.
  75. O’ Donnell MM, Forde BM, Neville B, Ross PR, O’ Toole PW. 2011. Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microb Cell Fact* 10:S12.
  76. Barrangou R, Azcarate-Peril MA, Duong T, Connors SB, Kelly RM, Klaenhammer TR. 2006. Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays. *Proc Natl Acad Sci USA* 103:3816–3821.
  77. Veldhuis G, Broos J, Poolman B, Scheek RM. 2005. Stoichiometry and substrate affinity of the mannitol transporter, enzymeII<sup>mtl</sup>, from *Escherichia coli*. *Biophys J* 89:201–210.
  78. Lai X, Ingram LO. 1993. Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. *J Bacteriol* 175:6441–6450.
  79. Lai X, Davis F V., Hespell RB, Ingram LO. 1997. Cloning of cellobiose phosphoenolpyruvate-dependent phosphotransferase genes: functional expression in recombinant. *Appl Environ Microbiol* 63:355–363.
  80. Zheng Z, Jiang T, Zou L, Ouyang S, Zhou J, Lin X, He Q, Wang L, Yu B, Xu H, Ouyang J. 2018. Simultaneous consumption of cellobiose and xylose by *Bacillus coagulans* to circumvent glucose repression and identification of its cellobiose-assimilating operons. *Biotechnol Biofuels* 11:320.
  81. Kappelmann L, Kr ¨uger K, Hehemann JH, Harder J, Markert S, Unfried F, Becher D, Shapiro N, Schweder T, Amann RI, Teeling H. 2019. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J* 13:76–91.
  82. Dagkesamanskaya A, Langer K, Tauzin AS, Rouzeau C, Lestrade D, Potocki-Veronese G, Boitard L, Bibette J, Baudry J, Pompon D, Anton-Leberre V. 2018. Use of photoswitchable fluorescent proteins for droplet-based microfluidic screening. *J Microbiol Methods* 147:59–65.
  83. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Basl ¨e A, Firbank SJ, Zheng H, Robinson C V., Winterhalter M, Kleinekath ¨ofer U, Bolam DN, Van Den Berg B. 2017.

- Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* 541:407–411.
84. Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 280:309–316.
  85. Hussein HS, Campbell JM, Bauer LL, Fahey GC, Hogarth AJCL, Wolf BW, Hunter DE. 1998. Selected fructooligosaccharide composition of pet-food ingredients. *J Nutr* 128:2803S–2805S.
  86. Korakli M, Hinrichs C, Ehrmann MA, Vogel RF. 2003. Enzymatic determination of inulin and fructooligosaccharides in food. *Eur Food Res Technol* 217:530–534.
  87. Chen J, Chia N, Kalari KR, Yao JZ, Novotna M, Soldan MMP, Luckey DH, Marietta E V., Jeraldo PR, Chen X, Weinshenker BG, Rodriguez M, Kantarci OH, Nelson H, Murray JA, Mangalam AK. 2016. Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. *Sci Rep* 6:1–10.
  88. Leclercq S, Matamoros S, Cani PD, Neyrinck AM, Jamar F, Stärkel P, Windey K, Tremaroli V, Bäckhed F, Verbeke K, de Timary P, Delzenne NM. 2014. Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of alcohol-dependence severity. *Proc Natl Acad Sci USA* 111:E4485–E4493.
  89. Sokol H, Leducq V, Aschard H, Pham HP, Jegou S, Landman C, Cohen D, Liguori G, Bourrier A, Nion-Larmurier I, Cosnes J, Seksik P, Langella P, Skurnik D, Richard ML, Beaugerie L. 2017. Fungal microbiota dysbiosis in IBD. *Gut* 66:1039–1048.
  90. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward D V, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 13:R79.
  91. Gupta A, Kang S, Wagner J, Kirkwood CD, Morrison M, McSweeney C, Macrae Finlay. 2014. Analysis of mucosal microbiota in inflammatory bowel disease using a custom phylogenetic microarray. *Austin J Gastroenterol* 1:1020.
  92. Mondot S, Lepage P, Seksik P, Allez M, Tréton X, Bouhnik Y, Colombel JF, Leclerc M, Pochart P, Doré J, Marteau P. 2016. Structural robustness of the gut mucosal microbiota is associated with Crohn’s disease remission after surgery. *Gut* 65:954–962.
  93. Lindsay J, Whelan K, Stagg AJ, Gobin P, Al-Hassi HO, Rayment N, Kaamm MA, S C Knight,

- Forbes A. 2006. Clinical, microbiological, and immunological effects of fructo-oligosaccharide in patients with Crohn's disease. *Gut* 55:348–355.
94. Benjamin JL, Hedin CRH, Koutsoumpas A, Ng SC, McCarthy NE, Hart AL, Kamm MA, Sanderson JD, Knight SC, Forbes A, Stagg AJ, Whelan K, Lindsay JO. 2011. Randomised, double-blind, placebo-controlled trial of fructo-oligosaccharides in active Crohn's disease. *Gut* 60:923–929.
  95. Wilson B, Whelan K. 2017. Prebiotic inulin-type fructans and galacto-oligosaccharides: definition, specificity, function, and application in gastrointestinal disorders. *J Gastroenterol Hepatol* 32:64–68.
  96. Węgleńska A, Jacob B, Sirko A. 1996. Transcriptional pattern of *Escherichia coli ihfB (himD)* gene expression. *Gene* 181:85–88.
  97. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 25:402–408.
  98. Kiefer P, Nicolas C, Letisse F, Portais JC. 2007. Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry. *Anal Biochem* 360:182–188.
  99. Rahman O, Cummings SP, Harrington DJ, Sutcliffe IC. 2008. Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World J Microbiol Biotechnol* 24:2377–2382.
  100. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580.
  101. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:D200–D203.
  102. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. 2015. Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta - Proteins Proteomics* 1854:1019–1037.

103. Saier MH, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. 2016. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* 44:D372–D379.
104. Iyer R, Camilli A. 2007. Sucrose metabolism contributes to *in vivo* fitness of *Streptococcus pneumoniae*. *Mol Microbiol* 66:1–13.
105. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504.
106. Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.

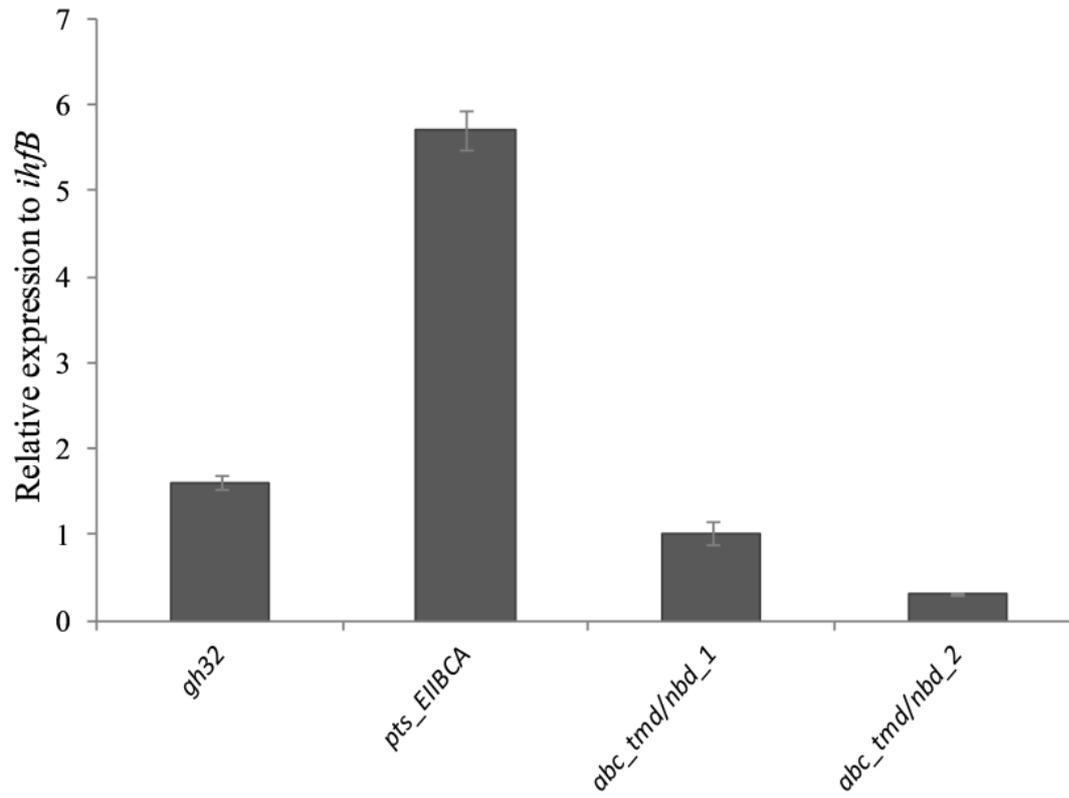
## Supplemental material

**Table S1. List of primers used in this study.**

| Name                            | Sequence 5'→3'                                     |
|---------------------------------|--|
| <i>I9 variant construction</i>  |  |
| Pcc1fos_f                       | GTGGGATCCTCTAGAGTCGACC                             |
| Pcc1fos_r                       | GTGGGATCCCCGGGTACCGAGC                             |
| I9min_F                         | ACCCGGGGATCCCACGCATCACAGGCGGACTTGCCAGAG            |
| I9min_R                         | TCTAGAGGATCCCACACCTCTCGGATCGTCGCAGACACA            |
| I9min_PTS/GH32_R                | TCTAGAGGATCCCACCAGAATCCATCAGCCGGATCACGG            |
| I9min_ABC/GH32_F                | GTGCATATAAACCTCCCATGGAATTAATCTTTGGTACAACCACCCCGATC |
| I9min_ABC/GH32_R                | CATGGGAGGTTTATATGCACCAACCGTAG                      |
| I9min_GH32_R                    | TCTAGAGGATCCCACCATGGGAGGTTTATATGCACCAACCGTAG       |
| I9min_PTS_F                     | ACCCGGGGATCCCACACCGCCTTCTGCATTAAACGGCG             |
| <i>Gene expression analysis</i> |  |
| I9a_gh32_f                      | ATCCCTCTCCTGCCAGCATA                               |
| I9a_gh32_r                      | ACGGCAAGATCTGTCAACGT                               |
| I9a_pts_EIIBC_A_f               | TACGATGGCCGGAATGATCG                               |
| I9a_pts_EIIBC_A_r               | TCGACATCGCAGGTATCACG                               |
| I9a_abc_tmd/nbd_1_f             | GCATGTCTGGATACCTCCCCG                              |
| I9a_abc_tmd/nbd_1_r             | CAGGATGCGGTAAGACGACA                               |
| I9a_abc_tmd/nbd_2_f             | TCAGTGTGTCTGTTCCGAGC                               |
| I9a_abc_tmd/nbd_2_r             | GACCAGTGTGACAGGCAA                                 |
| ihfB_f                          | GCCAAGACGGTTGAAGATGC                               |
| ihfB_r                          | CAAAGAGAACTGCCGAAACC                               |

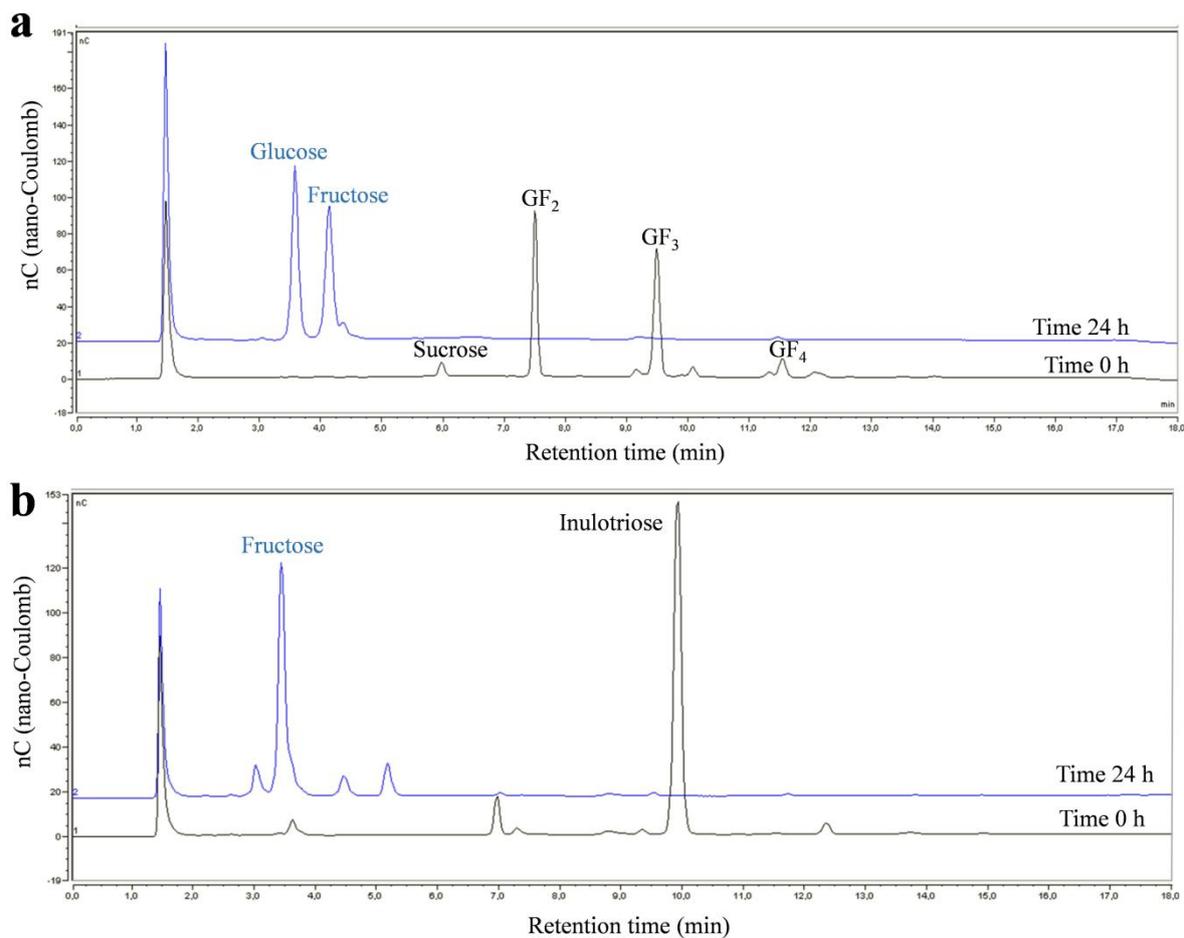
**Table S2. Ability of intracellular extracts of I9min\_GH32 cells to hydrolyze various FOS and  $\beta$ -fructans.**

| Substrate         | Initial concentration |      | Substrate consumption<br>after 24 h (%) |
|-------------------|-----------------------|------|---|
|                   | mg/ml                 | mM   |   |
| Sucrose           | 5                     | 14.6 | 99.4                                    |
| Kestose           | 5                     | 9.9  | 100                                     |
| Nystose           | 5                     | 7.5  | 100                                     |
| Fructosyl-nystose | 5                     | 6.0  | 92.2                                    |
| Inulotriose       | 5                     | 9.9  | 100                                     |
| Inulin            | 5                     | -    | 0                                       |
| Levan             | 5                     | -    | 0                                       |

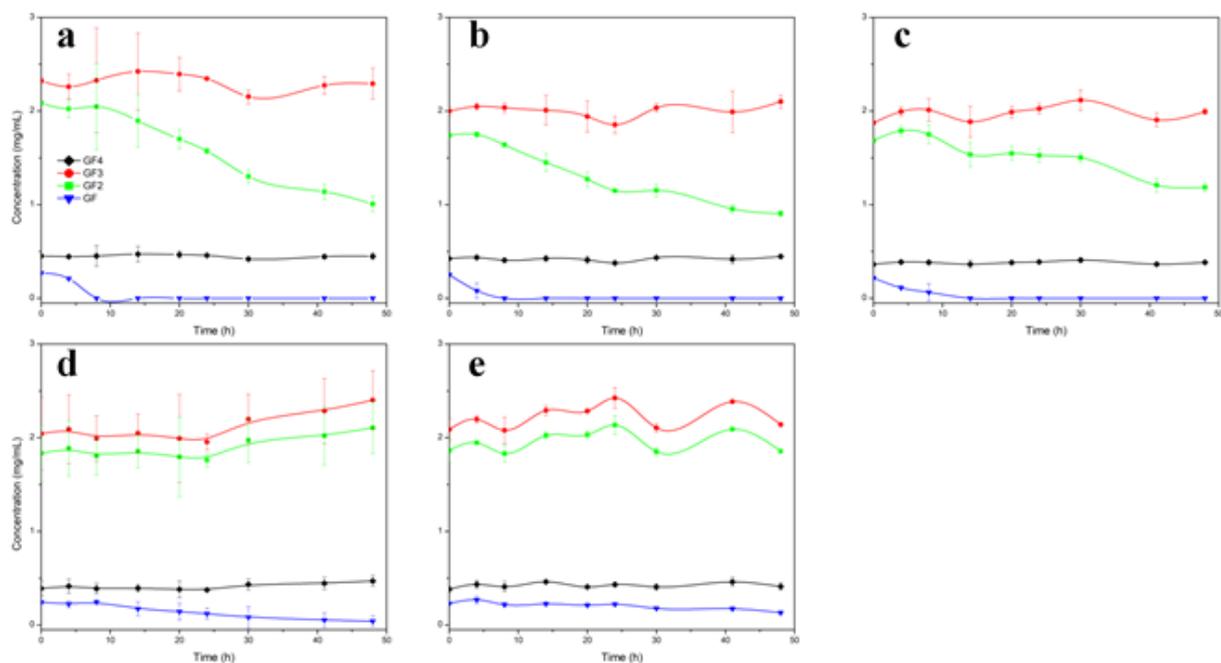


**Figure S1. Relative expression level of genes conferring the FOS metabolism phenotype to the I9 clone, vs that of the endogenous *E. coli* housekeeping gene (*ihfB*).**

The data represent four technical replicates from three biological replicates.

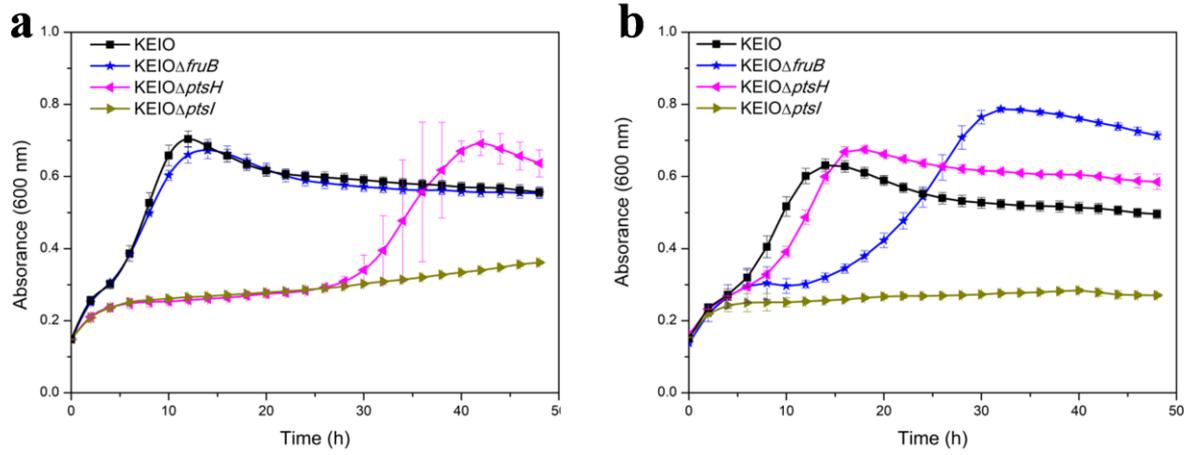


**Figure S2. HPAEC-PAD analysis of the reaction products resulting from (a) FOS and (b) Inulotriose hydrolysis by the I9min\_GH32 cytoplasmic extracts, after 24 h incubation at 37 °C.**



**Figure S3. Analysis of the uptake of each component of the FOS mixture by the (a) clone I9, and its variants (b) I9min, (c) I9min\_PTS/GH32, (d) I9min\_ABC/GH32 and (e) I9min\_GH32.**

The different FOS components are color-coded: sucrose (GF), blue; kestose (GF<sub>2</sub>), green; nystose (GF<sub>3</sub>), red; and 1<sup>F</sup>-fructofuranosyl-nystose (GF<sub>4</sub>), black. The data represent two biological replicates.



**Figure S4. Growth curves of knock-out mutants (*ptsI*, *ptsH* and *fruB*) of *E. coli* K-12 BW25113 (KEIO strain) on (a) glucose and (b) fructose.**

The data represent the mean of two biological replicates.



## Third article

As described in the first chapter of the “Results section”, two FOS metabolization pathways were identified, in the metagenomic clones F1 and I9, both issued from uncultured Firmicutes of the Clostridiales order. The FOS utilization locus has been characterized in the clone I9, revealing that the oligosaccharides were firstly transported and phosphorylated by the I9 PTS into the cell, and then hydrolyzed by the cytoplasmic GH32 enzyme. The clone F1 also harbors a PTS transporter, but lacking the EIIA domain compared with the I9 PTS. It was thus hereafter named F1 PTS\_EIIBC. In addition, the F1 locus contains a fructokinase encoding gene, but not the I9 one. Thus, we wondered i) if the F1 fructokinase would be involved in the utilization of FOS by the F1 clone, and ii) how could a PTS lacking the EIIA domain function in *Escherichia coli*. In this chapter, I present the work that I performed to address these questions. As explained below, further experiments will be required to confirm some of my hypotheses and to publish this article. I carried out all the experiments described hereafter, including the mass spectrometry analyses, which I performed thanks to the help of Pietro Tedesco and Fabien L  isse. They also provided the *E. coli* knockout mutants.



# Characterization of a fructooligosaccharide transporter from an uncultured human gut Clostridiales

## Abstract

Fructooligosaccharides (FOS) are increasingly being used as prebiotics in the human alimentation, to selectively stimulate the growth of beneficial human gut bacteria, such as *Bifidobacteria* and *Lactobacilli*. However, other bacteria, including non-beneficial species and uncultured ones from the Clostridiales order, can utilize FOS thanks to a phosphotransferase system to uptake them. Here, we describe the characterization of a FOS utilization locus belonging to an uncultured Clostridiales species, which involves a fructokinase and an original phosphotransferase system (PTS) lacking the carbohydrate-specific EIIA domain. Thanks to a genetic engineering approach combined with growth dynamic analyses of an *Escherichia coli* fosmid clone containing the Clostridiales locus, we demonstrated that the F1 PTS is critical for the uptake of FOS. By ion chromatography coupled with electrospray ion ionization high-resolution mass spectrometry (IC-ESI-HRMS), we proved that the F1 PTS expressed in *E. coli* is able to internalize kestose with its concomitant phosphorylation. The deletion of the genes encoding different *E. coli* PTS components showed that the missing components of the F1 PTS can be replaced by those of *E. coli*, highlighting the promiscuity of the carbohydrate PTS components EI, HPr and EIIA.

## Introduction

The human gastrointestinal tract contains a highly diverse and dynamic microbial community that exerts a profound impact on human health (Nayfach et al., 2019; Son et al., 2007; Valdes et al., 2018). The human gut microbiota indeed plays essential roles in the health and well-being of the host, since a disturbance or unbalance of the human gut microbiota is linked with obesity, inflammatory bowel diseases, liver diseases, colitis-associated cancers and so on (Chung et al., 2018; Gomes et al., 2018; Lavelle and Sokol, 2018; Marchesi et al., 2016). The molecular dialogue of the human gut microbiota with the host is typically communicated through the microbial metabolites including short-chain fatty acids (SCFAs), ammonia, phenols, certain amines and hydrogen sulfide (Marchesi et al., 2016). Notably, SCFAs are generated from dietary fiber fermentation by gut bacteria (Makki et al., 2018). Thus, dietary strategies, such as prebiotics, have been developed to modulate the metabolic activity and the composition of the human gut microbiota (DeMartino and Cockburn, 2020; Makki et al., 2018;

Wan et al., 2019). However, only few of the metabolic pathways involved in dietary fiber fermentation have been biochemically characterized, especially those of the yet uncultured species, which however represent the vast majority of the human gut ecosystem.

To efficiently capture and metabolize specific carbohydrates, bacteria have evolved different types of carbohydrate transport systems, including members of ATP-binding cassette (ABC), major facilitator superfamily (MFS), phosphotransferase system (PTS) and SusCD transporters (Saier, 2000; Terrapon et al., 2018). PTSs, which are one of the most ubiquitous transporters in bacteria, can transport a broad spectrum of monosaccharides, disaccharides, glycosides, and polyols. The PTS is a multiple component carbohydrate transport and phosphorylation system composed of two cytoplasmic phosphotransferases, Enzyme I (EI) and Histidine-containing phosphocarrier protein (HPr), which are common in most systems, and one carbohydrate-specific multiproteic Enzymes II complex (EII) consisting of two phosphotransferase domains (EIIA and EIIB), one integral membrane domain (EIIC) and sometimes an additional integral membrane domain EIID. These EII domains are often fused to each other, forming multifunctional polypeptides made up of two or more domains depending on bacterial species (Deutscher et al., 2006; Siebold et al., 2001). For the PTS transporters targeting carbohydrates, the phosphoryl group from phosphoenolpyruvate (PEP) is successively transferred from the EI to the HPr, then to the domains A and B of the Enzyme II (EII) and finally, to the carbohydrate bound to the EIIC domain (and sometimes EIID), which is responsible for selectively transporting carbohydrates across the inner bacterial membrane (McCoy et al., 2015).

Fructooligosaccharides (FOS) are oligosaccharides containing one or more  $\beta(2\rightarrow1)$  fructosyl units linked to sucrose. Mixtures of kestose (GF2), nystose (GF3) and fructosyl-nystose (GF4) have been ones of the firstly recognized prebiotics (Gibson et al., 2017). FOS are naturally present in a variety of plants such as wheat, barley, rye, onions, garlic, asparagus, Jerusalem artichoke and banana (Bornet and Brouns, 2002; Sangeetha et al., 2005). It has been reported that FOS can be selectively fermented by the human gut microbiota, including beneficial bacteria, such as *Bifidobacteria* and *Lactobacilli*, but also by commensal strains from various Clostridiales and Enterobacterales, including *Escherichia coli* strains isolated from humans with extraintestinal syndromes (meningitis or septicemia) (Goh and Klaenhammer, 2015; Mao et al., 2015; Schouler et al., 2009). However, the transporters specific for the uptake of FOS have been characterized only in a limited number of members of the gut microbiota. To date, FOS have been proved to be transported by PTS or ABC in the human gut Lactobacillales

including *Lactobacillus acidophilus* NCFM (Barrangou et al., 2003), *L. plantarum* ST-III (Chen et al., 2015) and *Streptococcus pneumoniae* TIGR4 (Linke et al., 2013), or by MFS in the gut pathogenic *Escherichia coli* BEN2908 (Schouler et al., 2009).

In addition, thanks to functional metagenomics, we recently discovered two FOS metabolization pathways issued from uncultured Clostridiales of the human fecal and ileal microbiota (Cecchini et al., 2013; Wang et al., 2020). One of them, identified in the metagenomic clone I9, is also present in *Dorea longicatena* DSM 13814. It involves a kestose specific PTS, consisting of the EIIA, EIIB and EIIC domains linked in one polypeptidic chain, that we deeply characterized by recombinant production in *E. coli* (Wang et al., 2020). The other one, from clone F1, differs from the I9 one because its PTS lacks a carbohydrate-specific EIIA domain.

Here, we biochemically characterized this original FOS PTS transporter, by using a genetic engineering approach combined with the study of the growth dynamic and FOS uptake ability of various *E. coli* knock-out mutants, transformed with the Clostridiales FOS utilization locus.

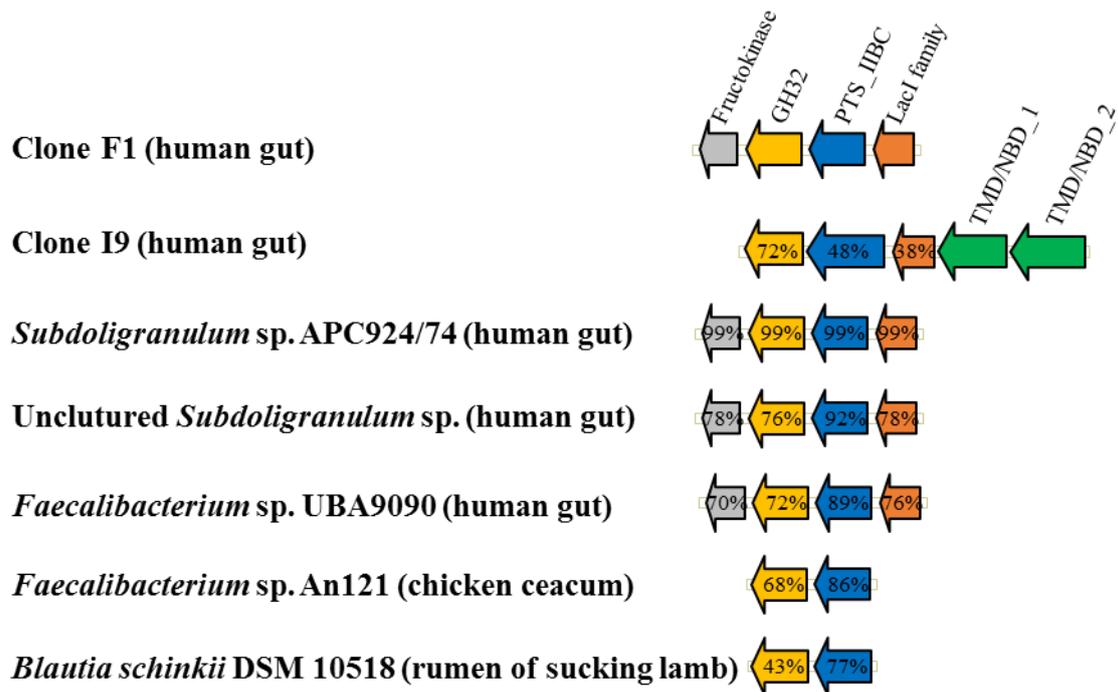
## Results

### Organization and genetic features of the FOS utilization locus from the clone F1

Previously, an activity-based screening of the human fecal metagenome led to the identification of a fosmid *E. coli* clone (clone F1) producing intracellular fructosidase activity responsible for the hydrolysis of a mixture of FOS, containing sucrose (GF), 1-kestose (GF<sub>2</sub>), nystose (GF<sub>3</sub>), and 1<sup>F</sup>-fructofuranosyl-nystose (GF<sub>4</sub>) (Cecchini et al., 2013). The F1 metagenomic sequence sizes 33,125 bp. Among its 23 genes, one FOS utilization locus was identified. It includes genes annotated as encoding a transcriptional regulator of the LacI family, a PTS component, a glycoside hydrolase belonging to the family GH32 of the carbohydrate-active enzymes (CAZy) classification, and a fructokinase (**Figure 1**). The absence of signal peptide implies that the GH32 enzyme and the fructokinase have a cytoplasmic location in the native organism. The RPS-BLAST analysis against the NCBI's Conserved Domain Database (CDD) indicated that the F1 PTS is a single multi-domain protein including the EIIB and EIIC domains, but lacking the EIIA domain of canonical PTS targeting carbohydrates (Morabbi Heravi and Altenbuchner, 2018; Tchieu et al., 2001; Wang et al., 2020). The F1 PTS was thus renamed F1 PTS\_EIIBC. As the I9 FOS utilization locus,

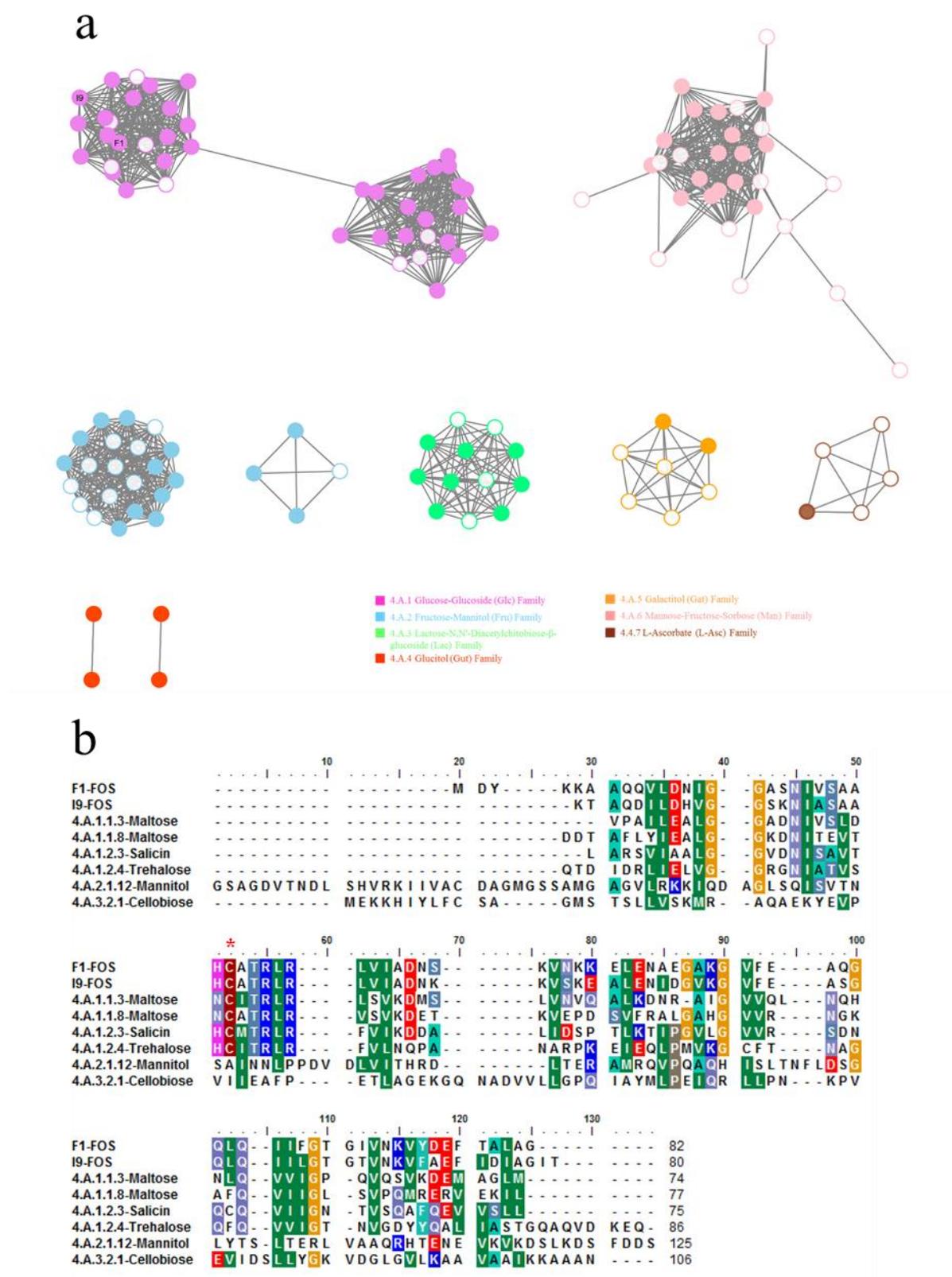
the genes encoding the EI and HPr are missing in the F1 clone, suggesting that the *E. coli* EI and HPr could complement the F1 PTS\_EIIBC to ensure its functionality in this recombinant host. The EIIC domain is membrane spanning, with ten predicted transmembrane helices, suggesting that the EIIC domain could bind extracellular substrates. By comparison with the I9 FOS utilization locus, F1 LacI, GH32 and PTS\_EIIBC share 37.8%, 47.7% and 72.4% amino acid sequence similarity with the I9 LacI, GH32 and EIIBC parts of the PTS, respectively (**Figure 1**). There is no other transporter sequence detected in clone F1, contrary to clone I9 which harbors an ABC transporter targeting sucrose (Wang et al., 2020). To further compare the F1 PTS with the other known carbohydrate PTSs, we constructed a sequence similarity network (SSN) with the EIIC domains from 126 PTSs, retrieved from the Transporter Classification database (TCDB, <http://www.tcdb.org/>), from the literature (Wang et al., 2020), and from the present study. In the SSN, the F1 PTS sequence clusters with the I9 EIIC domain, and with the members of the Glc family of the TCDB (**Figure 2a**), of which we recently revisited the functional diversity (Wang 2020). Besides, the IIB domain of the F1 PTS contains a particular cysteine residue conserved in the IIB domain of Glc PTS (**Figure 2b**). This residue can be phosphorylated by direct phosphoryl transfer from the IIA of the glucose PTS or one of its homologs (Eberstadt et al., 1996) suggesting the ability of the F1 IIB domain to be part of the phosphorylation cascade. The F1 sequence also contains a putative fructokinase encoding gene, which is absent in I9.

From a taxonomical point of view, the F1 nucleic acid sequence was assigned only to the order level, Clostridiales, as the sequence comes from a yet uncultured bacterium for which no genome was represented in the reference databases (Cecchini et al., 2013). F1 shows a microsynteny with other Clostridiales within the FOS locus, from the LacI to the fructokinase encoding genes (**Figure 1**). This locus is conserved in *Subdoligranulum* and *Faecalibacterium* species, which are abundant in the human (Visconti et al., 2019) and porcine (Lund et al., 2010) colon. Recently, the number of *F. prausnitzii* bacteria in the intestine of infants and adults was found to be increased by kestose consumption (Tochio et al., 2018) and the *Subdoligranulum variabile*-like bacteria were also abundant in the fecal bacteria of piglets inoculated with human fecal flora and supplemented with FOS (Shen et al., 2010). However, the molecular basis of FOS metabolization has never been studied in either of these species.



**Figure 1. Organization of the F1 FOS utilization locus, comparison with the I9 one, and synteny within other Clostridiales strains.**

Sequence identity with the F1 putative proteins is indicated in percentage. TMD and NBD represent respectively the transmembrane and nucleotide binding domains of the ABC transporter.



**Figure 2.** (a) Classification of the F1 PTS based on the sequence similarity network (SSN) of PTS-EIIC domains from 126 proteins, retrieved from the Transporter Classification database, from the literature, and from the present study, constructed with an initial score of  $10^{-5}$ . The

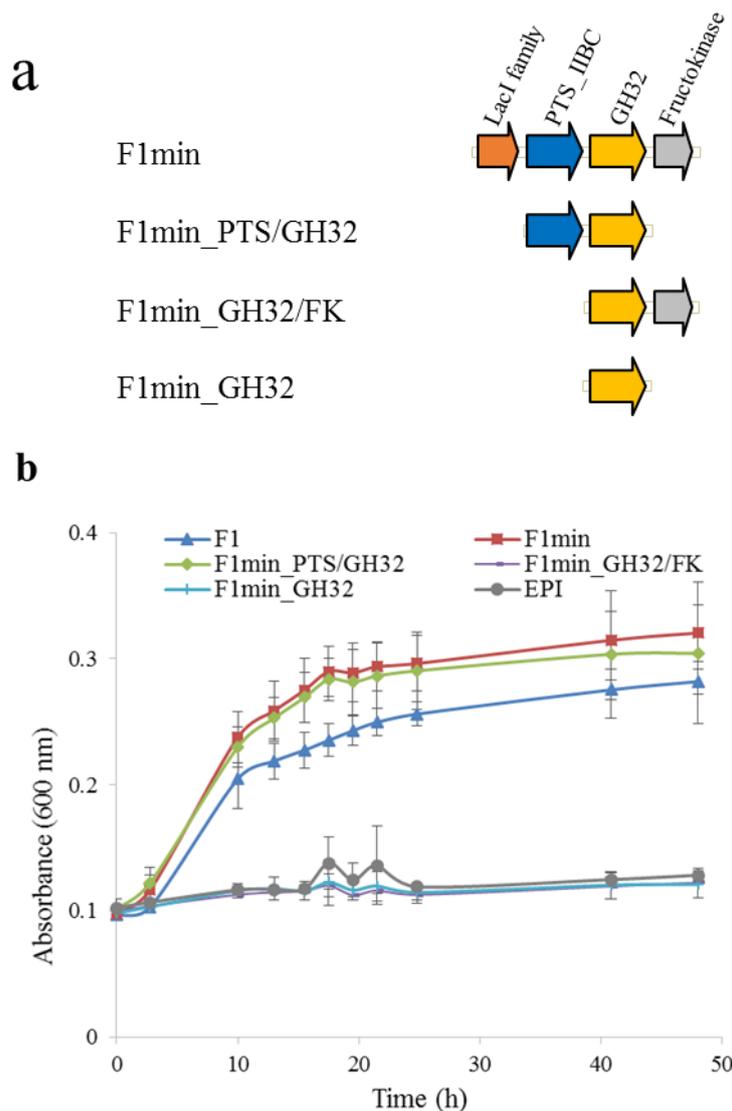
nodes are connected by an edge if the pairwise sequence identity is  $\geq 21.3\%$ . Unfilled nodes indicate sequences for which the functions were inferred by transcriptomics, while filled ones correspond to the functions validated biochemically or by gene deletion. (b) Multiple sequence alignment of the EIIB domains of the F1 and I9 PTSs, and their homologous belonging to the Glc, Fru and LacI families of the TCDB. The most conserved amino acids are indicated with colors. The conserved phosphorylation site (Cys residue) of EIIB is indicated by red asterisks above the alignment.

### **The F1 FOS locus is required for FOS utilization by *E. coli***

To prove the involvement of each gene of the F1 metagenomic locus in the uptake of FOS by *E. coli*, different truncated variants of the metagenomic DNA insert were generated (**Figure 3a**). As negative control, the *E. coli* host transformed with the empty fosmid pCC1FOS was used and hereafter named EPI.

Previously, the cytoplasmic cell extracts of clone F1 were tested on a mixture of FOS (including GF2 to GF4, with traces of sucrose) and inulin (GF<sub>n</sub>, with  $\beta$ -1,2-linked fructosyl units). After 24h of incubation, a complete hydrolysis of sucrose and FOS, and the release of shorter oligosaccharides and fructose from inulin were observed (Cecchini et al., 2013), while the *E. coli* EPI100 host strain was not able to hydrolyze FOS (Wang et al., 2020). Overall, these results suggest that the GH32 was responsible for the fructosidase activity of clone F1.

The uptake specificity of the F1 PTS\_EIIBC was investigated by testing the ability of truncated variants to use FOS as sole carbon source (**Figure 3b**). First, a reduced F1 consisting only of the genes encoding from the *LacI* to the *fructokinase* (F1min) was constructed. F1min grew more efficiently than clone F1 on FOS, demonstrating that the reduced locus was responsible for the growth on FOS. Interestingly, the clone missing the *LacI* and the fructokinase (F1min\_PTS/GH32) showed a similar growth curve to F1min, indicating that the *LacI* and the fructokinase have no crucial role in the FOS locus in *E. coli*. The slight difference observed between F1 and both variants, F1min and F1min\_PTS/GH32, might result from the higher number of recombinant proteins that could be produced by F1, which slowed down its growth compared to its variants. On the other hand, both variants harboring only the enzymes (F1min\_GH32/FK and F1min\_GH32) were not able to grow on FOS, confirming that the *E. coli* host strain does not possess a FOS transport system.



**Figure 3. Functional analysis of the F1 FOS utilization locus.**

(a) Gene organization in the F1 variants. (b) Growth curves of clone F1, its variants and the control clone EPI on a FOS mixture. The symbols indicate the sampling time points. The data represent the average of biological triplicates.

To characterize the transport specificity of the F1 PTS\_EIIBC, the percentage of consumption of each component of the FOS mixture was determined by HPAEC-PAD analysis after 48h of growth of clone F1 on FOS (**Table S1**). Clone F1 consumed sucrose and kestose but did not use nystose or 1<sup>F</sup>-fructofuranosyl-nystose. The growth of F1 was also tested on inulin, levan (GF<sub>n</sub>, with  $\beta$ -2,6-linked fructosyl units) and inulotriose (F<sub>3</sub>, with  $\beta$ -1,2-linked fructosyl units). No growth was observed with any of these carbohydrates (data not shown), while previously,

we showed that the F1 GH32 is able to hydrolyze the chains of  $\beta$ -1,2-linked fructosyl units of inulin with a degree of polymerization higher than 10 (Cecchini et al., 2013). These results demonstrate that the PTS\_EIIBC is specific for sucrose and kestose.

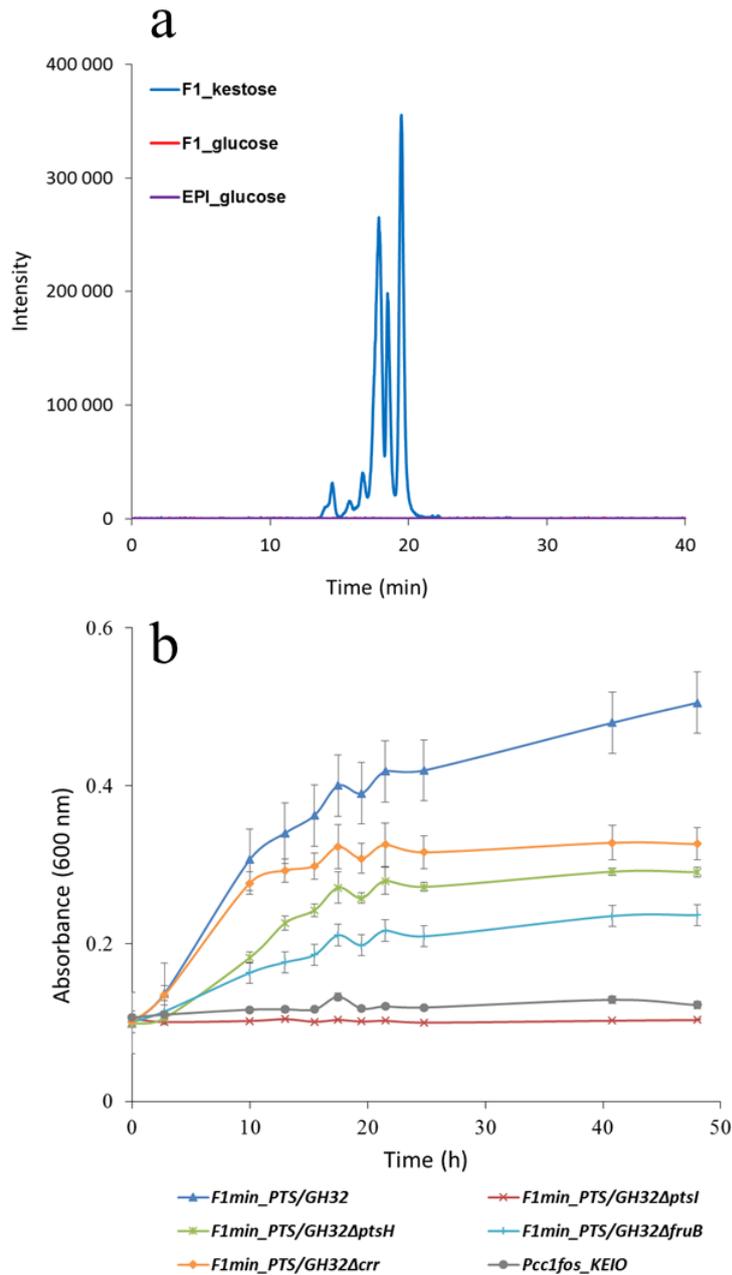
### **The functionality of the F1 PTS transporter in *E. coli* requires complementation by the components of the *E. coli* glucose or fructose PTSs**

In order to confirm that the F1 PTS\_EIIBC was able to couple transmembrane transport of its substrate to its concomitant phosphorylation, we analyzed the intracellular metabolites of clone F1 grown on kestose by ion chromatography coupled with electrospray ion ionization high-resolution mass spectrometry (IC-ESI-HRMS) (Millard et al., 2014), after quenching the metabolism (**Figure 4a**). Clone F1 and EPI grown on glucose was used as negative control. For all strains, we searched for an  $m/z$  value of 583.1270, which corresponds to the theoretically expected value in negative mode for the phosphorylated kestose. Using the eXtracted Ion Chromatogram (XIC), three main peaks with retention time between 14.5-19.6 min were observed only for the clone F1 grown on kestose, indicating the presence of three main forms of phosphorylated kestose.

As described in the introduction part, phosphorylation requires a multiproteic machinery to transfer the phosphoryl group from intracellular PEP to the carbohydrate bound to the EIIC domain of the PTS (Deutscher et al., 2014). Here, we demonstrated that the F1 PTS\_EIIBC is responsible for kestose internalization and phosphorylation by *E. coli*, even if the metagenomic F1 locus lacks the EI and HPr components of the PTS. Nevertheless, if produced by F1, we cannot exclude that the fructokinase may phosphorylate kestose once internalized by the F1 PTS\_EIIC. However, this is improbable, since the growth rate is identical for the F1<sub>min</sub> (including the fructokinase encoding gene) and the F1<sub>min</sub>\_PTS/GH32 (without the fructoskinase).

To investigate how the F1 PTS\_EIIC could be functional without the EI, HPr and EIIC components, we transformed the *E. coli* K-12 BW25113 (KEIO collection strains) single-knockout mutants for glucose and fructose-specific PTS genes with the fosmid F1<sub>min</sub>\_PTS/GH32. These mutants were grown on a mixture of FOS. The deletion of the *ptsI* (encoding for EI) completely abolish the growth on FOS (**Figure 4b**), confirming that the *E. coli* EI enzyme, which is involved in the first step of phosphorylation from PEP, is required to ensure F1 PTS functionality. The *ptsH* encodes the cytoplasmic HPr of the glucose PTS

transporter and the *fruB* encodes the HPr-like protein FPr of the fructose PTS transporter. As these proteins are not substrate specific, a single deletion of one of them could not abolish the growth (**Figure 4b**), because the remaining one could still transfer the phosphate group from EI to the carbohydrate-specific EIIA. Here, the variant deficient in the EIIA domain of the glucose PTS (encoded by gene *crr*) displayed a reduced growth compared with that of the wild type strain transformed with the F1min\_PTS/GH32 (**Figure 4b**). This indicates that the EIIA domain of the *E. coli* glucose PTS contributes to compensate the lack of EIIA domain in the F1 PTS, to transfer the phosphate group from the HPr or FPr proteins to the F1 EIIB. The fact that the *crr* knockout mutant is still able to grow on FOS indicates that the EIIA domain of other *E. coli* PTSs may also be involved in complementation of the F1 PTS. All in all, these results, which are in agreement with the results obtained previously for the clone I9, indicate that the F1 PTS\_EIIBC requires the *E. coli* PTS components including EI, HPr and EIIA proteins from different PTS systems that could be alternatively used for efficient substrate uptake.



**Figure 4. Functionality of the F1 PTS transporter in *E. coli*.**

(a) Extracted Ion Chromatogram (XIC) from negative ionization mode analysis at  $m/z$  583.1270, corresponding to the exact mass of the  $[M-H]^-$  ion of phosphorylated kestose (calculated from the molecular formula  $C_{18}H_{33}O_{19}P$ ) F1 clone grown on kestose and F1 clone grown on glucose. (b) Growth curves on FOS of *E. coli* K-12 1036 BW25113 (KEIO) single-knockout mutants for PTS genes (*ptsI*, *ptsH*, *FruB* and *crr*) transformed with fosmid F1min\_PTS/GH32. The KEIO wild-type strain transformed with the empty fosmid pCC1FOS

(pCC1FOS\_KEIO) was used as negative control. These data represent the mean of three biological replicates.

## Discussion

In this study, we characterized an original FOS utilization locus from a human gut uncultured Clostridiales by heterologous expression in *E. coli*. FOS, a natural component in some edible plants and a common prebiotic, has been demonstrated to be fermented mainly by *Bifidobacteria* and *Lactobacilli* in the human gut (Goh and Klaenhammer, 2015). To date, two mechanisms of FOS utilization have been described for human gut bacteria: (i) the oligosaccharide is internalized and then hydrolyzed in the cytoplasm, such as in *Lactobacillus plantarum* WCFS1 (Saulnier et al., 2007), *Lactobacillus acidophil* NCFM (Barrangou et al., 2003), and *Dorea longicatena* DSM 13814 (Wang et al., 2020); or (ii) the oligosaccharide is firstly hydrolyzed by cell surface-associated enzymes to monosaccharides, which are then internalized and metabolized, as in *Lactobacillus paracasei* W20 and 1195 (Boger et al., 2018; Yong et al., 2007). With this second mechanism, the release of monosaccharides may result in a cross-feeding among gut bacteria (Boger et al., 2018). Here, we showed that the reduced F1 locus harboring two genes encoding a cytoplasmic GH32 and a short PTS transporter is enough to confer *E. coli* the ability to use FOS. The F1 FOS utilization mechanism involves a specific PTS transporter capable to uptake kestose which then will be degraded intracellularly, maximizing the utilization of FOS compared with the mechanism involving extracellular hydrolysis.

Besides, the F1 locus exhibits high conservation in the prominent gut strains *Faecalibacterium* and *Subdoligranulum* spp. Several isolates from these species revealed their ability to produce short-chain fatty acids such as butyric acid and formic acid (Bjerrum et al., 2006; R ós-Covi án et al., 2016), which may have an important function in both growth performance and protection against pathogens (Holscher, 2017). Thus, the identification of the F1 FOS utilization pathway may facilitate the profiling of the metabolic properties of these beneficial species. It also highlights the potential of these strains for the design of FOS containing symbiotics. Nevertheless, the involvement of the FOS utilization locus in the ability of these strains to compete with other FOS utilization bacteria in the gut remains to be demonstrated. Moreover, similarity in gene organization between the F1 locus and other species indicates that horizontal gene transfer probably led to the dispersion of the FOS utilization ability among gut bacteria. Indeed, in the human gut, bacteria would greatly benefit

from acquisition of such gene clusters involved in transport and catabolism of fibers, especially if they confer a competitive edge toward successful colonization (Sun, 2018), in case of long-term dietary supplementation with a particular type of fibers.

Previously, the PTS\_EIIBCA transporter in clone I9 has been characterized to transport and phosphorylate kestose in *E. coli* with the complementation of *E. coli* HPr and EI like components. The EI and HPr are PTS components which are common to all PTS transporters, regardless of their targeted substrate. For example, the cellobiose PTS\_EII complex from the *Bacillus coagulans* NL01 Gram-positive bacterium has been reported to be complemented by the *E. coli* EI and HPr PTS components (Zheng et al., 2018). Similarly, we demonstrated in a recent study that the I9 PTS transporter, which is identical to that of *Dorea longicatena* DSM 13814, requires the complementation of the *E. coli* HPr and EI proteins to be functional in this recombinant host (Wang et al., 2020). We observed the same phenomena with the F1 PTS. In addition, we showed here that the lack of the EIIB domain in this original PTS could be compensated by the EIIB domain of the *E. coli* glucose PTS. Similar complementations between EIIB domains have previously been described in *E. coli* and *B. subtilis*, which also have PTSs lacking the EIIB. Complementation always happened between EIIB components from the Glc TCDB family, and from the same strain (Morabbi Heravi and Altenbuchner, 2018; Tchieu et al., 2001). Complementation of PTS lacking an EIIB domain from a Gram-negative strain by the EIIB domain the *E. coli* Glc PTS has already been described too (Pikis et al., 2006). Phylogenetic analysis within the F1 PTS in SSN also revealed that the F1 PTS\_EIIBC belongs to the Glc family. This is also in agreement with the conservation of the cysteine residue in the F1 PTS EIIB domain, which indicates that the F1 PTS\_EIIB domain can be phosphorylated by direct phosphoryl transfer from IIB of glucose PTS or one of its homologues. The results of our study highlight the promiscuity of the EIIB domains of PTS even inter Gram- positive and negative species. In the native strain of which the F1 FOS utilization locus originates, kestose uptake would thus rely on the EIIB domain of another PTS system which would be concomitantly expressed. This hypothesis will have to be tested by transcriptomic analysis of *Subdoligranulum* sp. APC924/74 grown on FOS, which is the sequenced strain which contains a locus sharing more than 99 % sequence identity with the F1 FOS utilization locus. In addition, the role played by the fructokinase will have to be investigated to fully understand the mechanism of FOS internalization by the Clostridiales species equipped with homologs of the F1 FOS utilization locus.

In conclusion, we functionally characterized the role of the GH32 enzyme and of the original PTS encoded by a FOS utilization locus from a human gut uncultured bacteria, which is highly similar to that of prominent gut species of which the FOS utilization ability has, to our knowledge, not hitherto been reported. Furthermore, in this study we used an approach based on the use of knockout mutants of the *E. coli* host to study the functionality of transporters issued from uncultured bacteria. This approach allowed us to demonstrate the cooperation between three different components of PTS with various specificities and from different bacterial species, highlighting the promiscuity of carbohydrate PTS transporters.

## Materials and methods

### Cloning

The metagenomic clone F1 (Genbank accession number HE663537) comes from a metagenomic library issued from a human fecal sample of a healthy 30-year-old male, as previously described (Cecchini et al., 2013). In this library, the metagenomic DNA fragments (33,125 bp) were cloned into the pCC1FOS fosmid and transformed in EPI100 *E. coli* cells (Epicentre Technologies). The minimal variants of clone F1 were constructed using the In-Fusion HD Plus cloning kit (Clontech) following the instructions in the user manual. Amplification was started 600 bp to 1 kb upstream of the first gene of the target locus, to involve potential promoter sequences which might be necessary for gene expression. The primers used in this study are listed in **Table S2**. Mutants were confirmed by new generation sequencing, performed by the GeT-Biopuces Platform (Toulouse) using the Ion Torrent S5 System. Read assembly was performed using Masurca (<http://www.genome.umd.edu/masurca.html>). The assembled contigs were cleaned from the pCC1FOS vector sequence using Crossmatch (<http://bozeman.mbt.washington.edu/phredphrapconsed.html>).

### Growth conditions

For growth experiment, clone F1 and its variants were firstly pre-cultured in LB medium supplemented with 12.5 mg/l chloramphenicol. Then, cells were grown in M9 medium supplemented with 12.5 mg/l chloramphenicol with an initial optical density at 600 nm ( $OD_{600\text{ nm}}$ ) of 0.05, obtained by inoculating the culture with the appropriate volume of the overnight preculture in the LB medium. The M9 medium contains  $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$  17.4 g/l,  $\text{KH}_2\text{PO}_4$  3.03 g/l, NaCl 0.51 g/l,  $\text{NH}_4\text{Cl}$  2.04 g/l,  $\text{MgSO}_4$  0.49 g/l,  $\text{CaCl}_2$  4.38 mg/l,

Na<sub>2</sub>EDTA 2H<sub>2</sub>O 15 mg/l, ZnSO<sub>4</sub> ·7H<sub>2</sub>O 4.5 mg/l, CoCl<sub>2</sub> 6H<sub>2</sub>O 0.3 mg/l, MnCl<sub>2</sub> 4H<sub>2</sub>O 1 mg/l, H<sub>3</sub>BO<sub>3</sub> 1 mg/l, Na<sub>2</sub>MoO<sub>4</sub> 2H<sub>2</sub>O 0.4 mg/l, FeSO<sub>4</sub> ·7H<sub>2</sub>O 3 mg/l, CuSO<sub>4</sub> 5H<sub>2</sub>O 0.3 mg/l, thiamine 0.1 g/l and leucine 0.02 g/. Solutions of FOS (Beghin Meiji), 1-kestose (GF<sub>2</sub>, Wako Chemicals), inulotriose (Megazyme), inulin from dahlia tuber (Sigma) or levan from *Erwinia herbicola* (Sigma) were added in the M9 medium as sole carbon source, at a final concentration of 0.5% (w/v) , after sterilization by filtration at 0.2 µm (Sartorius). The growth was followed by measuring the OD<sub>600 nm</sub> over 48 h at 37 °C using a FLUOStar Optima plate reader (BMG Labtech).

### **FOS uptake analysis**

To assay the uptake of the different oligosaccharides in the FOS mixture by the clone F1, we analyzed the concentration of residual carbohydrates in the culture media when cells were growing in M9 medium supplemented with XOS at 37 °C. Growth was monitored at regular point time, meanwhile samples were collected and centrifuged. The supernatants were filtered at 0.20 µm and analyzed by HPAEC-PAD on a Dionex ICS-3000 system using a CarboPac PA100 4×250 column. Carbohydrates were eluted at 30 °C, at a flow rate of 1 ml/min with the following multistep gradient: 0-30 min (0-60% B), 30-31 min (60-0% B) and 31-36 min (0% B). Solvents were 150 mM NaOH (eluent A) and 150 mM NaOH, 500 mM CH<sub>3</sub>COONa (eluent B).

### **Kestose phosphorylation analysis**

The presence of intracellular phosphorylated kestose in clone F1 cells grown on kestose was examined by ion chromatography coupled with electrospray ion ionization high-resolution mass spectrometry (IC-ESI-HRMS) as previously described (Millard et al., 2014; Wang et al., 2020)(Millard et al., 2014; Wang et al., 2020). Briefly, 100 µL of an overnight preculture of the *E. coli* clone F1 in LB were used to inoculate 50 ml baffled shake flasks containing 10 ml of M9 medium supplemented with glucose or kestose as carbon sources. The flasks were incubated for 24 h at 37 °C. Cells, washed with diluted M9, were used to inoculate 50 ml baffled shake flasks at an OD<sub>600 nm</sub> of 0.1, containing M9 supplemented with the same carbon source. Cell growth was monitored regularly at 37 °C with orbital shaking at 220 rpm by measuring OD<sub>600 nm</sub> with a cell density meter (Ultrospect 10, Amersham BioSciences). When culture reached the middle of the lag phase, 120 µl of culture medium were withdrawn and vigorously mixed with 1.25 ml of an acetonitrile/methanol/H<sub>2</sub>O (4:4:2) solution pre-cooled at -20 °C to rapidly quench metabolic activity and extract metabolites (Millard et al., 2014). The

samples were kept at -20 °C for 20 minutes, and then centrifuged at 10,000 g at 4 °C for 10 minutes to remove cell debris. The supernatant was dried using a SpeedVac (SC110A SpeedVac Plus, ThermoSavant) under vacuum and then stored at - 80 °C until analysis. The samples were dissolved in 120 µL of mQ water and then analyzed by ion chromatography (Thermo Scientific Dionex ICS-5000+ system, Dionex) coupled to an LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with an electrospray ionization probe (IC-ESI-HRMS). A revised ion chromatography method was used here (Kiefer et al., 2007). The KOH gradient was modified as follows: 0 min, 0.5 mM; 1 min, 0.5 mM; 9.5 min, 4.1 mM; 14.6 min, 4.1 mM; 24 min, 9.65 mM; 36 min, 60 mM; 36.1 min, 90 mM; 43 min, 90 mM; 43.5 min, 0.5 mM; 48 min, 0.5 mM. The gradients were all linear. For background suppression, an Anionic Electrolytically Regenerated Suppressor (AERS 300-2 mm, Dionex) was used. The device was operated at a constant 87 mA electrolysis current, in external water mode with ultrapure water and regenerant delivered by an external AXP pump at a flow rate of 1 ml/min. The volume of the injected sample was 35 µl. Mass spectrometry analysis was performed in the negative FTMS mode at a resolution of 30,000 in full scan mode, with the following source parameters: capillary temperature, 350 °C; source heater temperature, 300 °C; sheath gas flow rate, 50 a.u. (arbitrary unit); auxiliary gas flow rate, 5 a.u.; S-Lens RF level, 60%; and ion spray voltage, 3.5 kV. The data were acquired using the Xcalibur software (Thermo Fisher Scientific).

## **Bioinformatic analyses**

The conserved domains of the protein were identified by using RPS-BLAST against the NCBI's Conserved Domain Database (CDD) (Marchler-Bauer et al., 2017). Transmembrane helices in the F1 PTS protein were predicted with the TMHMM Server v.2.0 (Krogh et al., 2001). The presence and location of lipoprotein and other signal peptide cleavage sites were predicted by the LipoP v1.0 server (Juncker and Willenbrock, 2003).

Conservation of the FOS locus in sequenced genomes was analyzed by BLAST comparison of the amino acid sequence of the genes encoding the F1 PTS, GH32, LacI, and fructokinase proteins with the non-redundant protein sequences in NCBI database (Boratyn et al., 2013). Sequence similarity networks (SSNs) were generated similarly as previously described (Wang et al., 2020).

## References

- Barrangou, R., Altermann, E., Hutkins, R., Cano, R., Klaenhammer, T.R., 2003. Functional and comparative genomic analyses of an operon involved in fructooligosaccharide utilization by *Lactobacillus acidophilus*. Proc. Natl. Acad. Sci. USA 100, 8957–8962. <https://doi.org/10.1073/pnas.1332765100>
- Bjerrum, L., Engberg, R.M., Leser, T.D., Jensen, B.B., Finster, K., Pedersen, K., 2006. Microbial community composition of the ileum and cecum of broiler chickens as revealed by molecular and culture-based techniques. Poult. Sci. <https://doi.org/10.1093/ps/85.7.1151>
- Boger, M.C.L., Lammerts van Bueren, A., Dijkhuizen, L., 2018. Cross-feeding amongst probiotic bacterial strains on prebiotic inulin involving the extracellular exo-inulinase of *Lactobacillus paracasei* strain W20. Appl. Environ. Microbiol. 84, e01539-18. <https://doi.org/10.1128/AEM.01539-18>
- Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y., Raytselis, Y., Sayers, E.W., Tao, T., Ye, J., Zaretskaya, I., 2013. BLAST: a more efficient report with usability improvements. Nucleic Acids Res. 41, 29–33. <https://doi.org/10.1093/nar/gkt282>
- Bornet, F.R.J., Brouns, F., 2002. Immune-stimulating and gut health-promoting properties of short-chain fructo-oligosaccharides. Nutr. Rev. 60, 326–334. <https://doi.org/10.1301/002966402320583442>
- Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré J., Henrissat, B., Remaud-Simón, M., Monsan, P., Potocki-Véronèse, G., 2013. Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. PLoS One 8, e72766. <https://doi.org/10.1371/journal.pone.0072766>
- Chen, C., Zhao, G., Chen, W., Guo, B., 2015. Metabolism of fructooligosaccharides in *Lactobacillus plantarum* ST-III via differential gene transcription and alteration of cell membrane fluidity. Appl. Environ. Microbiol. 81, 7697–7707. <https://doi.org/10.1128/AEM.02426-15>
- Chung, H.-J., Nguyen, T.T.B., Kim, H.-J., Hong, S.-T., 2018. Gut microbiota as a missing

- link between nutrients and traits of human. *Front. Microbiol.* 9, 1510. <https://doi.org/10.3389/FMICB.2018.01510>
- DeMartino, P., Cockburn, D.W., 2020. Resistant starch: impact on the gut microbiome and health. *Curr. Opin. Biotechnol.* 61, 66–71. <https://doi.org/10.1016/j.copbio.2019.10.008>
- Deutscher, J., Ak  F.M.D., Derkaoui, M., Z r A.C., Cao, T.N., Bouraoui, H., Kentache, T., Mokhtari, A., Milohanic, E., Joyet, P., 2014. The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. *Microbiol. Mol. Biol. Rev.* 78, 231–256. <https://doi.org/10.1128/MMBR.00001-14>
- Deutscher, J., Francke, C., Postma, P.W., 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* 70, 939–1031. <https://doi.org/10.1128/MMBR.00024-06>
- Eberstadt, M., Grdadolnik, S.G., Gemmecker, G., Kessler, H., Buhr, A., Erni, B., 1996. Solution structure of the IIB domain of the glucose transporter of *Escherichia coli*. *Biochemistry* 35, 11286–11292. <https://doi.org/10.1021/bi9604921>
- Gibson, G.R., Hutkins, R., Sanders, M.E., Prescott, S.L., Reimer, R.A., Salminen, S.J., Scott, K., Stanton, C., Swanson, K.S., Cani, P.D., Verbeke, K., Reid, G., 2017. Expert consensus document: the International Scientific Association for Probiotics and Prebiotic (ISAPP) consensus statement on the definition and scope of prebiotics. *Nat. Rev. Gastroenterol. Hepatol.* 14, 491–502. <https://doi.org/10.1038/nrgastro.2017.75>
- Goh, Y.J., Klaenhammer, T.R., 2015. Genetic mechanisms of prebiotic oligosaccharide metabolism in probiotic microbes. *Annu. Rev. Food Sci. Technol.* 6, 137–156. <https://doi.org/10.1146/annurev-food-022814-015706>
- Gomes, A.C., Hoffmann, C., Mota, J.F., 2018. The human gut microbiota: Metabolism and perspective in obesity. *Gut Microbes* 0976, 1–18. <https://doi.org/10.1080/19490976.2018.1465157>
- Holscher, H.D., 2017. Dietary fiber and prebiotics and the gastrointestinal microbiota. *Gut Microbes* 8, 172–184. <https://doi.org/10.1080/19490976.2017.1290756>

- Juncker, A., Willenbrock, H., 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662. <https://doi.org/10.1110/ps.0303703>.
- Kiefer, P., Nicolas, C., Letisse, F., Portais, J.C., 2007. Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry. *Anal. Biochem.* 360, 182–188. <https://doi.org/10.1016/j.ab.2006.06.032>
- Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. <https://doi.org/10.1006/jmbi.2000.4315>
- Lavelle, A., Sokol, H., 2018. Gut microbiota: Beyond metagenomics, metatranscriptomics illuminates microbiome functionality in IBD. *Nat. Rev. Gastroenterol. Hepatol.* 15, 193–194. <https://doi.org/10.1038/nrgastro.2018.15>
- Linke, C.M., Woodiga, S.A., Meyers, D.J., Buckwalter, C.M., Salhi, H.E., King, S.J., 2013. The ABC transporter encoded at the pneumococcal fructooligosaccharide utilization locus determines the ability to utilize long- and short-chain fructooligosaccharides. *J. Bacteriol.* 195, 1031–1041. <https://doi.org/10.1128/JB.01560-12>
- Lund, M., Bjerrum, L., Pedersen, K., 2010. Quantification of *Faecalibacterium prausnitzii*- and *Subdoligranulum variabile*-like bacteria in the cecum of chickens by real-time PCR. *Poult. Sci.* 89, 1217–1224. <https://doi.org/10.3382/ps.2010-00653>
- Makki, K., Deehan, E.C., Walter, J., Bäckhed, F., 2018. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* 705–715. <https://doi.org/10.1016/j.chom.2018.05.012>
- Mao, B., Li, D., Zhao, J., Liu, X., Gu, Z., Chen, Y.Q., Zhang, H., Chen, W., 2015. *In vitro* fermentation of fructooligosaccharides with human gut bacteria. *Food Funct.* 6, 947–954. <https://doi.org/10.1039/C4FO01082E>
- Marchesi, J.R., Adams, D.H., Fava, F., Hermes, G.D.A., Hirschfield, G.M., Hold, G., Quraishi, M.N., Kinross, J., Smidt, H., Tuohy, K.M., Thomas, L. V., Zoetendal, E.G., Hart, A., 2016. The gut microbiota and host health: A new clinical frontier. *Gut* 65, 330–339. <https://doi.org/10.1136/gutjnl-2015-309990>

- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., Bryant, S.H., 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. <https://doi.org/10.1093/nar/gkw1129>
- McCoy, J.G., Levin, E.J., Zhou, M., 2015. Structural insight into the PTS sugar transporter EIIC. *Biochim. Biophys. Acta - Gen. Subj.* 1850, 577–585. <https://doi.org/10.1016/j.bbagen.2014.03.013>
- Millard, P., Massou, S., Wittmann, C., Portais, J.C., L  isse, F., 2014. Sampling of intracellular metabolites for stationary and non-stationary <sup>13</sup>C metabolic flux analysis in *Escherichia coli*. *Anal. Biochem.* 465, 38–49. <https://doi.org/10.1016/j.ab.2014.07.026>
- Morabbi Heravi, K., Altenbuchner, J., 2018. Cross-talk among transporters of the phosphoenolpyruvate-dependent phosphotransferase system in *Bacillus subtilis*. *J. Bacteriol.* JB.00213-18. <https://doi.org/10.1128/JB.00213-18>
- Nayfach, S., Shi, Z.J., Seshadri, R., Pollard, K.S., Kyrpides, N.C., 2019. New insights from uncultivated genomes of the global human gut microbiome. *Nature* 568, 505–510. <https://doi.org/10.1038/s41586-019-1058-x>
- Pikis, A., Hess, S., Arnold, I., Erni, B., Thompson, J., 2006. Genetic requirements for growth of *Escherichia coli* K12 on methyl- $\alpha$ -D-glucopyranoside and the five  $\alpha$ -D-glucosyl-D-fructose isomers of sucrose. *J. Biol. Chem.* 281, 17900–17908. <https://doi.org/10.1074/jbc.M601183200>
- R  s-Covi  n, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., De los Reyes-Gavil  n, C.G., Salazar, N., 2016. Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* 7, 1–9. <https://doi.org/10.3389/fmicb.2016.00185>
- Saier, M.H., 2000. Families of transmembrane sugar transport proteins. *Mol. Microbiol.* 35, 699–710. <https://doi.org/10.1046/j.1365-2958.2000.01759.x>
- Sangeetha, P.T., Ramesh, M.N., Prapulla, S.G., 2005. Recent trends in the microbial production, analysis and application of fructooligosaccharides. *Trends Food Sci. Technol.*

16, 442–457. <https://doi.org/10.1016/j.tifs.2005.05.003>

Saulnier, D.M.A., Molenaar, D., De Vos, W.M., Gibson, G.R., Kolida, S., 2007. Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays. *Appl. Environ. Microbiol.* 73, 1753–1765. <https://doi.org/10.1128/AEM.01151-06>

Schouler, C., Taki, A., Chouikha, I., Moulin-Schouleur, M., Gilot, P., 2009. A genomic island of an extraintestinal pathogenic *Escherichia coli* strain enables the metabolism of fructooligosaccharides, which improves intestinal colonization. *J. Bacteriol.* 91, 388–393. <https://doi.org/10.1128/JB.01052-08>

Shen, J., Zhang, B., Wei, H., Che, C., Ding, D., Hua, X., Bucheli, P., Wang, L., Li, Y., Pang, X., Zhao, L., 2010. Assessment of the modulating effects of fructo-oligosaccharides on fecal microbiota using human flora-associated piglets. *Arch. Microbiol.* 192, 959–968. <https://doi.org/10.1007/s00203-010-0628-y>

Siebold, C., Flükiger, K., Beutler, R., Erni, B., 2001. Carbohydrate transporters of the bacterial phosphoenolpyruvate: Sugar phosphotransferase system (PTS). *FEBS Lett.* 504, 104–111. [https://doi.org/10.1016/S0014-5793\(01\)02705-3](https://doi.org/10.1016/S0014-5793(01)02705-3)

Son, D., Le, N.T., Pannacciulli, N., Chen, K., Salbe, A.D., Hill, J.O., Rena, R., Reiman, E.M., Krakoff, J., 2007. Diversity of the Human Intestinal Microbial Flora. *Science* 308, 1635–1638. <https://doi.org/10.1109/TMI.2012.2196707>.

Sun, D., 2018. Pull in and push out: mechanisms of horizontal gene transfer in bacteria. *Front. Microbiol.* 9, 2154. <https://doi.org/10.3389/fmicb.2018.02154>

Tchieu, J.H., Norris, V., Edwards, J.S., Saier, M.H., 2001. The complete phosphotransferase system in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 3, 329–46.

Terrapon, N., Lombard, V., Drula, É., Lap bie, P., Al-Masaudi, S., Gilbert, H.J., Henrissat, B., 2018. PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* 46, D677–D683. <https://doi.org/10.1093/nar/gkx1022>

Tochio, T., Kadota, Y., Tanaka, T., Koga, Y., 2018. 1-Kestose, the smallest fructooligosaccharide component, which efficiently stimulates faecalibacterium

- prausnitzii as well as bifidobacteria in humans. *Foods* 7, 1–11. <https://doi.org/10.3390/foods7090140>
- Valdes, A.M., Walter, J., Segal, E., Spector, T.D., 2018. Role of the gut microbiota in nutrition and health. *BMJ* 361, 36–44. <https://doi.org/10.1136/bmj.k2179>
- Visconti, A., Le Roy, C.I., Rosa, F., Rossi, N., Martin, T.C., Mohny, R.P., Li, W., de Rinaldis, E., Bell, J.T., Venter, J.C., Nelson, K.E., Spector, T.D., Falchi, M., 2019. Interplay between the human gut microbiome and host metabolism. *Nat. Commun.* 10. <https://doi.org/10.1038/s41467-019-12476-z>
- Wan, M.L.Y., Ling, K.H., El-Nezami, H., Wang, M.F., 2019. Influence of functional food components on gut health. *Crit. Rev. Food Sci. Nutr.* 59, 1927–1936. <https://doi.org/10.1080/10408398.2018.1433629>
- Wang, Z., Tauzin, A.S., Lavillea, E., Tedesco, P., L'èissea, F., Terraponb, N., Lepercqa, P., Mercadea, M., Potocki-Veronese, G., 2020. Harvesting of prebiotic fructooligosaccharides by non-beneficial human gut bacteria. *mSphere* 5, e00771-19. <https://doi.org/https://doi.org/10.1128/mSphere.00771-19>
- Wang, Z., Tauzin, A.S., Potocki-Veronesea, G., n.d. Identification of functional glycoside transporters from uncultured bacteria. in preparation.
- Yong, J.G., Lee, J.H., Hutkins, R.W., 2007. Functional analysis of the fructooligosaccharide utilization operon in *Lactobacillus paracasei* 1195. *Appl. Environ. Microbiol.* 73, 5716–5724. <https://doi.org/10.1128/AEM.00805-07>
- Zheng, Z., Jiang, T., Zou, L., Ouyang, S., Zhou, J., Lin, X., He, Q., Wang, L., Yu, B., Xu, H., Ouyang, J., 2018. Simultaneous consumption of cellobiose and xylose by *Bacillus coagulans* to circumvent glucose repression and identification of its cellobiose-assimilating operons. *Biotechnol. Biofuels* 11, 320. <https://doi.org/10.1186/s13068-018-1323-5>

## Supplemental information

**Table S1. Analysis of the consumption of the carbohydrates by HPAEC-PAD by clone F1 grown for 48 hours in M9 supplemented with FOS as sole carbon source.**

| <b>Components of FOS mixture</b>        | <b>Initial concentration (mM)</b> | <b>Percentage of consumption after 48 h</b> |
|---|-----------------------------------|---|
| Sucrose                                 | 0.77                              | 100   |
| Kestose                                 | 3.96                              | 15.2  |
| Nystose                                 | 3.45                              | 0   |
| 1 <sup>F</sup> -fructofuranosyl-nystose | 0.52                              | 0   |

**Table S2. List of primers used for the construction of clone F1 variants.**

| <b>Name</b>              | <b>Sequence 5'→3'</b>                    |
|--------------------------|--|
| pCC1FOS_F                | GTGGGATCCTCTAGAGTCGACC                   |
| pCC1FOS_R                | GTGGGATCCCCGGGTACCGAGC                   |
| F1min_LacI/PTS/GH32/FK_F | ACCCGGGGATCCCACCTTGGCTTTGGGCACACTGTAAAAG |
| F1min_LacI/PTS/GH32/FK_R | TCTAGAGGATCCCACGCAGCCAGTTCATCAATGATCTTG  |
| I9min_PTS/GH32_F         | ACCCGGGGATCCCACAACCTACCAGCTGCTGCTGATCAAC |
| I9min_PTS/GH32_R         | TCTAGAGGATCCCACGCAGCCAGTTCATCAATGATCTTG  |
| I9min_GH32/FK_F          | ACCCGGGGATCCCACCTCAAGACGCTGGGCGATATCTTTG |
| I9min_GH32/FK_R          | TCTAGAGGATCCCACGCAGCCAGTTCATCAATGATCTTG  |
| I9min_GH32_F             | ACCCGGGGATCCCACCTCAAGACGCTGGGCGATATCTTTG |
| I9min_GH32_R             | TCTAGAGGATCCCACGCAGCCAGTTCATCAATGATCTTG  |

## Fourth article

Previously, a metagenomic clone (F5) containing a PUL specific for xylooligosaccharides (XOS) issued from an uncultured gut *Bacteroides* strain, was identified and characterized in the team (Cecchini et al., 2013; Tauzin et al., 2016). This locus encodes a SusCD and a MFS transport system, dedicated to XOS uptake up to a degree of polymerization 3 and 4, respectively. The deletion of the MFS transporter and SusD protein from the clone F5 (variant F5min\_SusΔSusD) abolished the growth ability of the recombinant *Escherichia coli* cells on XOS, which shows that the presence of SusD is essential for the functionality of the SusC transporter and for the uptake of XOS. In other characterized PULs, the physical presence of the SusD protein was also found to be essential for strain growth on the polysaccharides targeted by the PUL.

The F5\_SusD protein shares very low sequence similarities with those from characterized PULs, and its binding specificity has never been studied. In this context, the following questions arised:

- Does the binding specificity of SusD reflect the uptake specificity of the SusCD transport system towards oligosaccharides of defined structures? This is a generic question for glycoside transporters, which, thanks to our recombinant approach, we could answer to.
- What are the relationships between the tridimensional structure of the F5\_SusD and its function?

In this chapter, I present the work that we collectively performed to address these questions. The SusD production and purification, the isothermal titration calorimetry and affinity gel electrophoresis analyses were performed by Alexandra Tauzin. I produced and purified the SusD protein to perform the complementation experiments, and prepared the samples for the nuclear magnetic resonance experiments performed by Guy Lippens. I carried out the crystallization assays under the supervision of Gianluca Cioci, who solved the SusD structure.



# **Non-binding SusD protein is essential for xylooligosaccharide utilization in a human gut uncultured *Bacteroides***

## **Abstract**

In the human gut microbiota, the Bacteroidetes is a dominant phylum, whose members can breakdown dietary and endogenous glycosides by involving highly specific polysaccharide utilization loci (PULs), which encode a variety of sensor/regulators, binding proteins, transporters and carbohydrate-active enzymes (CAZymes). Surface glycan-binding proteins (SGBPs) are essential for an efficient capture of the glycosides present nearby the cells, conferring a competitive advantage to Bacteroidetes to colonize their habitats. Here, we present the functional and structural characterization of a SusD-like protein encoded by a xylooligosaccharide utilization locus (XOS PUL) from an uncultured human gut *Bacteroides*, which is also conserved in *B. vulgatus*, thereby providing new mechanistic insights into the role of SGBPs in the metabolism of dietary fibers of importance for gut health. Complementation experiments and various *in vitro* and *in cellulo* analyses, including saturation transfer difference NMR spectroscopy, revealed that the SusD protein cannot bind to the cognate substrate of the XOS PUL, whereas its presence is essential for the PUL functioning. The analysis of the crystal structure of the SusD protein revealed an unfolded binding surface and the absence or inappropriate orientation of several key residues compared with other solved SusD-like structures. These results highlight the critical role of the SusD protein in the oligosaccharides transport, and provide new fundamental knowledge about the structure-function of SusCD transporters, revealing that the binding specificity of the SusD-like SGBP does not necessarily reflect the uptake specificity of the transporter.

## **Introduction**

Human rely on their microbiota to breakdown dietary fibers, which are oligosaccharides and polysaccharides that cannot be digested by the host carbohydrate-active enzymes (CAZymes (Lombard et al., 2014)). To do so, Bacteroidetes harbor polysaccharide utilization loci (PULs) encoding for proteins involved in sensing, binding, transporting and degrading the target

glycosides into monosaccharides. According to the Sus paradigm (Cameron et al., 2014; Cho and Salyers, 2001), polysaccharides are first degraded by enzymes attached at the cell surface. Then SusD-like proteins, sometimes together with the other surface glycan binding proteins, SusE and/or F, selectively recognize and capture the oligosaccharides which will be transported into the cell through the SusC-like protein.

In-depth characterization of the first described starch utilization locus from *Bacteroides thetaiotaomicron* showed that the SusD protein has a critical role in addition to its binding ability (Foley et al., 2016). Indeed, deletion of SusD resulted in a loss of growth, which could be restored by adding the SusD recombinant protein within the culture medium, suggesting that the presence of SusD is a prerequisite for efficient transport (Koropatkin et al., 2008; Sonnenburg et al., 2010; Tamura et al., 2019; Tauzin et al., 2016a). SusD and SusC interact together to form a complex (Cho and Salyers, 2001). The recent solving of the structure of SusC/D complexes provided new insights to identify the molecular determinants of the interactions protein/protein (Glenwright et al., 2017). The SusC/D pair is well conserved among Bacteroidetes, pointing out their crucial role in glycan uptake, and the presence of genes encoding SusD is crucial to predict PULs (Terrapon et al., 2018).

Dozens of Bacteroidetes strains have been shown to use xylan, as main carbon source (Despres et al., 2016; Dodd et al., 2010; Martens et al., 2011; McNulty et al., 2013; Mendis et al., 2018; Mirande et al., 2010; Rogowski et al., 2015; Salyers et al., 1981; Tang et al., 2017; Zhang et al., 2014). Xylan is a major constituent of hemicellulose, abundantly available in plant cell walls, particularly in cereals, fruits and vegetables (Leth et al., 2018; Selvendran, 1984). The conserved backbone of xylan is composed of a linear  $\beta$ -(1-4)-D-xylopyranosyl chain which could be decorated with arabinofuranosyl, glucopyranosyl or uronic acid derivatives (Singh et al., 2015). Xylooligosaccharides (XOS) are hydrolyzed from xylan or naturally found in fruits, vegetable bamboo, and honey (Singh et al., 2015). They have recently shown a range of health benefits even at a lower dose compared to fructooligosaccharides, and are thus considered as prebiotic candidates (Amorim et al., 2019; Gibson et al., 2017; Goh and Klaenhammer, 2015; Singh et al., 2015). To date, only two different PULs specific for xylan utilization by *B. ovatus* ATCC 8483 (PUL-XylS and PUL-XylL) have been fully biochemically characterized, including SusD-like proteins and demonstrating different specificities according to the xylan structural complexity (Rogowski et al., 2015). In addition, in 2016 we identified a PUL specific for XOS, issued from an uncultured gut *Bacteroides* strain, and characterized its recombinant expression in *E. coli*

(Cecchini et al., 2013; Tauzin et al., 2016b). This locus encodes a SusC/D and an MFS transport system, dedicated to XOS uptake up to a degree of polymerization (DP) 3 and 4, respectively. We also demonstrated that the deletion of the SusD protein completely abolished the growth ability of the recombinant *E. coli* cells on XOS, which is in agreement with the results obtained with other SusD proteins. Furthermore, this PUL shares 99 % sequence identity at the DNA level with a *Bacteroides vulgatus* ATCC 8482 PUL (BVU\_0037 to BVU\_0043), that we proved to be involved in linear XOS utilization by this strain (Tauzin et al., 2016b). Its growth is very robust on XOS, while it is delayed and at a lower rate on arabino-xylo-oligosaccharides (AXOS). In addition, *B. vulgatus* is unable to utilize complex heavily decorated xylans or arabinoxylan, although marginally it could utilize simpler xylans from Beechwood and Birchwood (Tauzin et al., 2016b).

In the present study, we investigated how the SusD protein from this XOS targeting PUL contributes to the utilization of its cognate substrate. Using complementation experiments, various *in vitro* and *in cellulo* techniques to investigate the binding ability of SusD, and analysis of its crystallographic structure, we provided new mechanistic insights into the XOS uptake by *Bacteroides* strains.

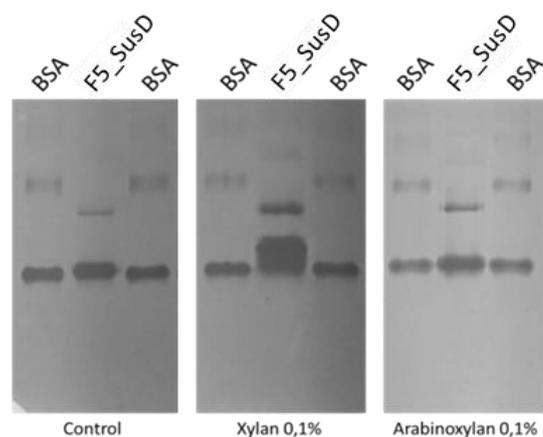
## Results

### SusD cannot bind to XOS

#### *Determination of the binding abilities by AGE and ITC*

Previously we demonstrated that a locus from an uncultured *Bacteroides* conferred the ability to *E. coli* to utilize xylo-oligosaccharides (XOS). The locus encodes a SusD-like protein (hereafter named F5\_SusD) sharing 99.4% identity with the SusD protein (BVU\_0037) from *Bacteroides vulgatus* ATCC 8482 (with four differences in positions N249, D444, A446, and A516 which correspond to D, G, P and V residues in BVU\_0037, respectively). The recombinant protein was expressed in *E. coli* without the predicted signal peptide and N-terminal lipidation site. The binding ability of F5\_SusD towards soluble xylan and arabinoxylan was proved by affinity gel electrophoresis (AGE) (**Figure 1**). The two bands observed correspond to the dimer and monomer forms of the F5\_SusD. The dimer did not show any difference of migration in presence and in absence of glycan. In contrast, the migration of the monomer was weakly delayed, by around 12 %, in the gel containing xylan. The obtained smear suggests that the binding of the monomer to xylan could be unstable due

to a really weak affinity. The retention of F5\_SusD in the gel containing xylan is much less than the one observed for all the other characterized SusD proteins which efficiently bind polysaccharides (Tamura et al., 2019; Tauzin et al., 2016a). No migration delay was observed with wheat arabinoxylan. These results indicate a weak affinity of F5\_SusD for the xylan backbone, further impaired by the presence of side-chains. Isothermal titration calorimetry (ITC) was attempted using XOS as ligand, but no binding could be detected (data not shown).

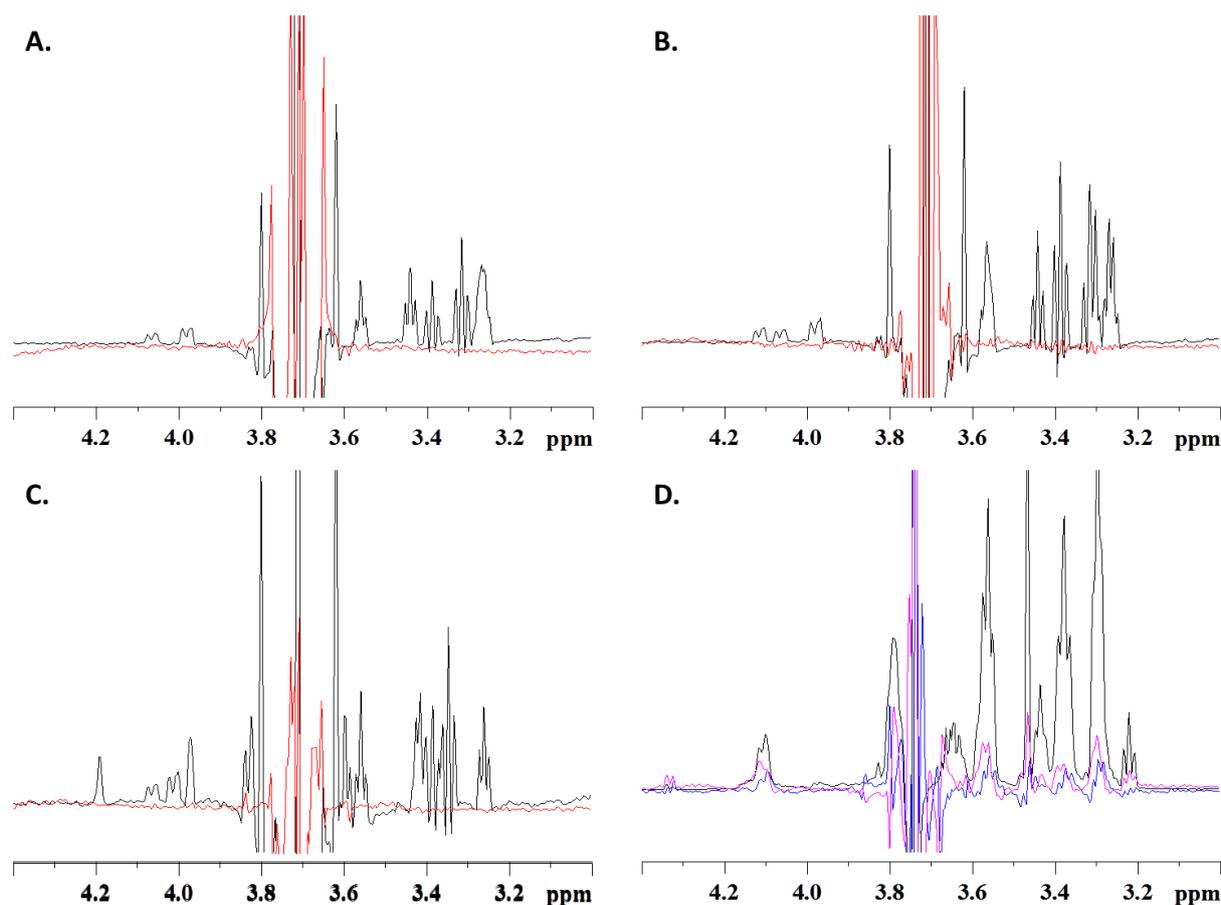


**Figure 1. Affinity gel electrophoresis of the F5\_SusD on a gel containing no polysaccharide (negative control) and on gels with 0.1% (w/v) beechwood xylan and wheat arabinoxylan, respectively.**

#### ***Determination of the binding ability by NMR***

The interest of saturation transfer difference (STD) NMR to analyze directly protein-ligand interactions even with weak affinities was previously shown (Mayer and Meyer, 1999). However, to our knowledge, this approach has only rarely been used to analyze the binding ability of carbohydrate-binding proteins associated to CAZymes or to carbohydrate transporters (Bell et al., 2019). We first confirmed by 1D NMR that the protein was correctly folded in solution (**Figure S1**). The STD experiments with xylobiose, xylotriose or 3<sup>2</sup>- $\alpha$ -L-arabinofuranosyl-xylobiose (AX2) against this F5\_SusD protein were negative (**Figure 2A, B and C**), thereby confirming that small xylo-oligosaccharides do not binding to the protein. A beechwood xylan sample was then prepared, but its white shine indicated that even at 500  $\mu$ M, the xylan chains do associate into macroscopic structures. The NMR experiment in the presence of the F5\_SusD protein gave a robust STD signal, but such was equally the case when we probed the same xylan sample in the absence of the protein (**Figure S2**). We hence

could not distinguish if binding occurred between xylan and the transporter or between the macroscopic xylan superstructures. To reduce signal saturation due to these macroscopic xylan structures, we saturated the protein at the level of its indole protons (at 10.1ppm), where xylan signals should be reduced. The STD effect with the xylan alone was indeed reduced (**Figure S3**), and was not only more important in the presence of F5\_SusD, but also increased when we added more protein (**Figure 2D**). We thus concluded that F5\_SusD does not bind short xylo-oligosaccharides, but weakly binds to the macroscopic xylan chains.

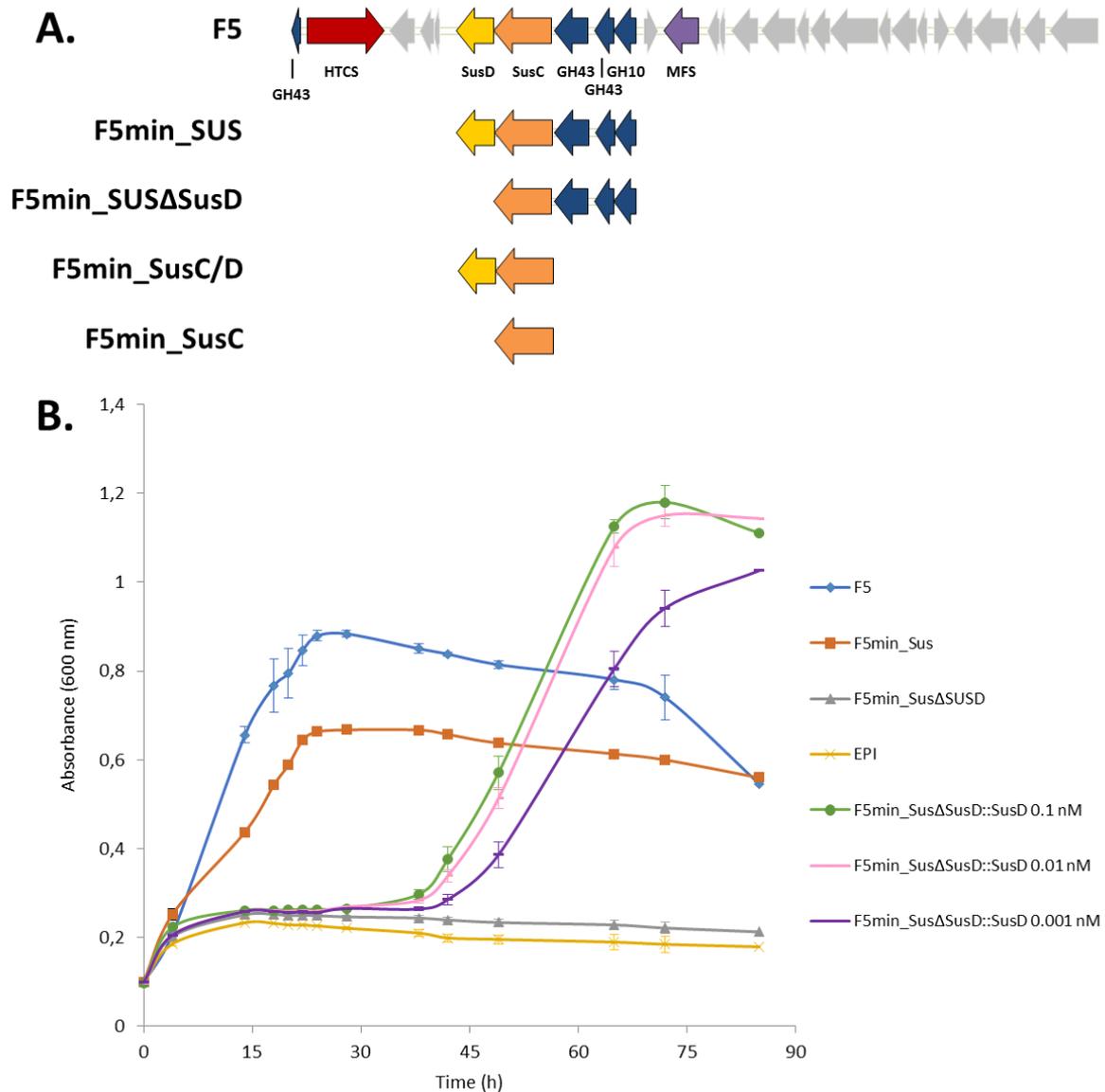


**Figure 2. Characterization of the ligand binding abilities of F5-SusD using STD-NMR.**

(A.) Zoom on the 1D xylobiose proton spectrum centered on the signals at 3.7ppm (black) and STD signal of the ligand upon saturation of the protein signals at 0ppm (red). (B.) Idem for the xylotriose. (C.) Idem for the branched arabino-xylo-oligosaccharides (AX2,  $3^2\text{-}\alpha\text{-L}$ -arabinofuranosyl-xylobiose). (D.) Zoom on the 1D xylan spectrum (black), STD signal upon saturation at 10.1ppm of the isolated xylan (blue) or the xylan in presence of F5\_SusD (pink). The increased STD effect in the presence of F5\_SusD indicates a molecular interaction.

## **SusD is essential for an efficient transport of XOS**

Previously, we demonstrated that the growth on XOS was completely abolished when the *SusD* encoding gene was deleted (clone F5min\_SUS $\Delta$ SusD) (Tauzin et al., 2016b). This finding was in agreement with the results obtained after deletion of *SusD* of the loci involved in the uptake of starch from *Bacteroides thetaiotaomicron* VPI-5482 (Cameron et al., 2014), of xyloglucan from *Bacteroides ovatus* ATCC 8483 (Tauzin et al., 2016a) and of mixed linkage  $\beta$ -glucans from *B. ovatus* ATCC 8483 (Tamura et al., 2019). In all the cases, the addition of the recombinant protein in the growth medium restored the growth of the deleted strain on the cognate substrate. Similarly, we performed preliminary complementation experiments by adding the recombinant F5\_SusD in the growth medium of the F5min\_SUS strain at different concentrations (**Figure 3**). We observed that between a range of 0.001 to 0.1 nM of recombinant F5\_SusD present in the growth medium, the transport system was efficiently complemented, resulting in the uptake of XOS by the *SusC* transporter. Thus, the presence of F5\_SusD is essential for functionality of the XOS locus.



**Figure 3. Complementation of XOS locus with recombinant F5\_SusD.**

(A.) Representation of the XOS loci from the metagenomics clone F5 and its variants used in the present study. (B.) Growth test of F5 and its variants on XOS with or without addition of recombinant SusD at different concentrations (0.001 to 0.1 nM). The data represent the average of at least three biological replicates.

### Crystallographic structure of F5\_SusD

In order to better understand why the F5\_SusD does not bound the XOS substrate targeted by the PUL, its three-dimensional structure has been solved by X-ray crystallography. The most similar sequence listed in the Protein Data Bank (PDB, <https://www.rcsb.org/>) (Berman et al.,

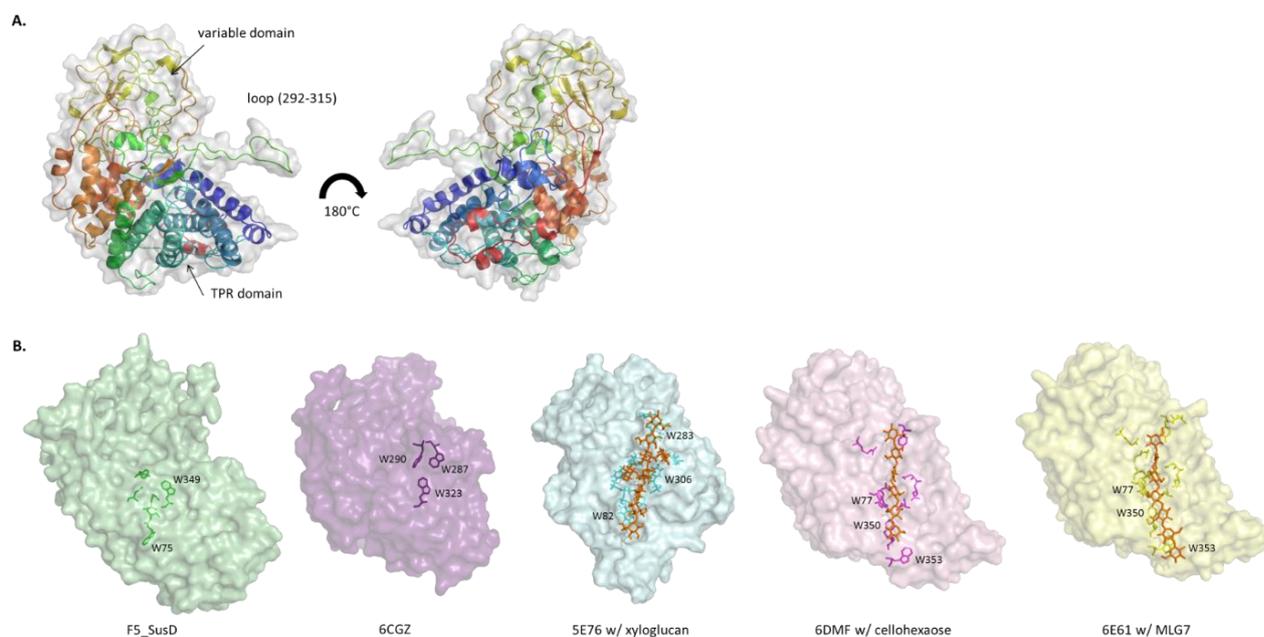
2000)), is that of a SusD protein (PDB-ID: 3OTN) from *Parabacteroides distasonis* ATCC 8503, with which the F5 SusD only shares 34% identity on a short stretch of sequence (128 length). Thus, the F5\_SusD structure, which could not be solved by molecular replacement, has been solved by a MR-SAD approach (Skubak et al., 2018) and refined at a 2.6Å resolution. Data collection and refinement statistics are shown in **Table 1**. Despite the very low sequence similarity with the structurally characterized SusD proteins, the F5\_SusD protein displays a canonical ‘RagB/SusD’ fold which includes the conserved tetratricopeptide repeat (TPR) units. Overall, the F5\_SusD structure is similar to that of the other SusD proteins, with the presence of a convex surface, made by the alpha helical bundle and a flat surface on the other side, where the putative sugar binding platform should be localized. The closest structural homolog identified with the Dali server (<http://ekhidna2.biocenter.helsinki.fi/dali/> (Holm, 2019)) is the Laminarin Binding protein GM\_SusD (PDB-ID:6GCZ), with an RMSD of about 2.9Å (Mystkowska et al., 2018). The presence of a large variable domain (D292-P480) inserted between the two conserved  $\beta$ -strands in the core of the helical bundle, and essentially made of random coils and some  $\beta$ -sheets (**Figure 4A**), is one of the particular features of F5\_SusD. This domain does not have any homologs in the PDB. Another interesting feature is the presence of several loops (*e.g.* K225-A246, R552-S538 and M622-Q631) that are inserted between the conserved secondary structural elements. In particular, the loop 292-315 that is protruding away from the folded domain and which is partly stabilized by interactions with glycine and serine residues of one of the neighboring SusD molecules, is probably completely disordered in solution (**Figure S4**). Overall, these features make the F5\_SusD slightly bigger than the other structurally characterized members of the SusD family (**Figure 4B**).

**Table 1. Data collection and refinement statistics for the crystal structure of F5\_SusD**

| <i>Data collection</i>                      | Native                     | S-SAD                      |
|---|----------------------------|----------------------------|
| <b>Wavelength (Å)</b>                       | 0.979                      | 2.066                      |
| <b>Space group</b>                          | I4 <sub>1</sub> 22         | I4 <sub>1</sub> 22         |
| <b>Molecules per asymmetric unit</b>        | 1                          | 1                          |
| <b>Cell constants a, b, c (Å)</b>           | 179.08<br>179.08<br>182.23 | 179.84<br>179.84<br>183.10 |
| <b>Resolution (Å)</b>                       | 50.00-2.65                 | 50.00-3.30                 |
| <b>Measured reflections</b>                 | 307172                     | 2114678                    |
| <b>Unique reflections</b>                   | 43113                      | 22897                      |
| <b>Data completeness %</b>                  | 100.0                      | 100.0                      |
| <b>Multiplicity</b>                         | 7.1                        | 92.4                       |
| <b>Rmerge</b>                               | 0.062 (0.736)              | 0.10 (0.41)                |
| <b>&lt; I/σ (I) &gt;</b>                    | 8.4 (1.0)                  | 5.2 (1.5)                  |
| <b>CC<sub>1/2</sub></b>                     | 0.99 (0.86)                | 1.00 (0.99)                |
| <b>Wilson B-factor (Å<sup>2</sup>)</b>      | 65.3                       | 73.7                       |
| <b>Refinement</b>                           |                            |                            |
| <b>R<sub>work</sub>/R<sub>free</sub></b>    | 0.187/0.219                |                            |
| <b>RMSD bonds (Å)</b>                       | 0.070                      |                            |
| <b>RMSD angles (°)</b>                      | 1.191                      |                            |
| <b>Ramachandran's</b>                       |                            |                            |
| <b>Favored/Allowed/Outliers (%)</b>         | 96/4/0                     |                            |
| <b>Number of atoms</b>                      |                            |                            |
| <b>Protein</b>                              | 4829                       |                            |
| <b>Water</b>                                | 6                          |                            |
| <b>Other</b>                                | 6                          |                            |
| <b>Average B-factor (Å<sup>2</sup>)</b>     | 84.6                       |                            |
| <b>CC<sub>all</sub> / CC<sub>free</sub></b> | 0.96/0.95                  |                            |
| <b>Clashscore (percentile)</b>              | 4 (100 <sup>th</sup> )     |                            |

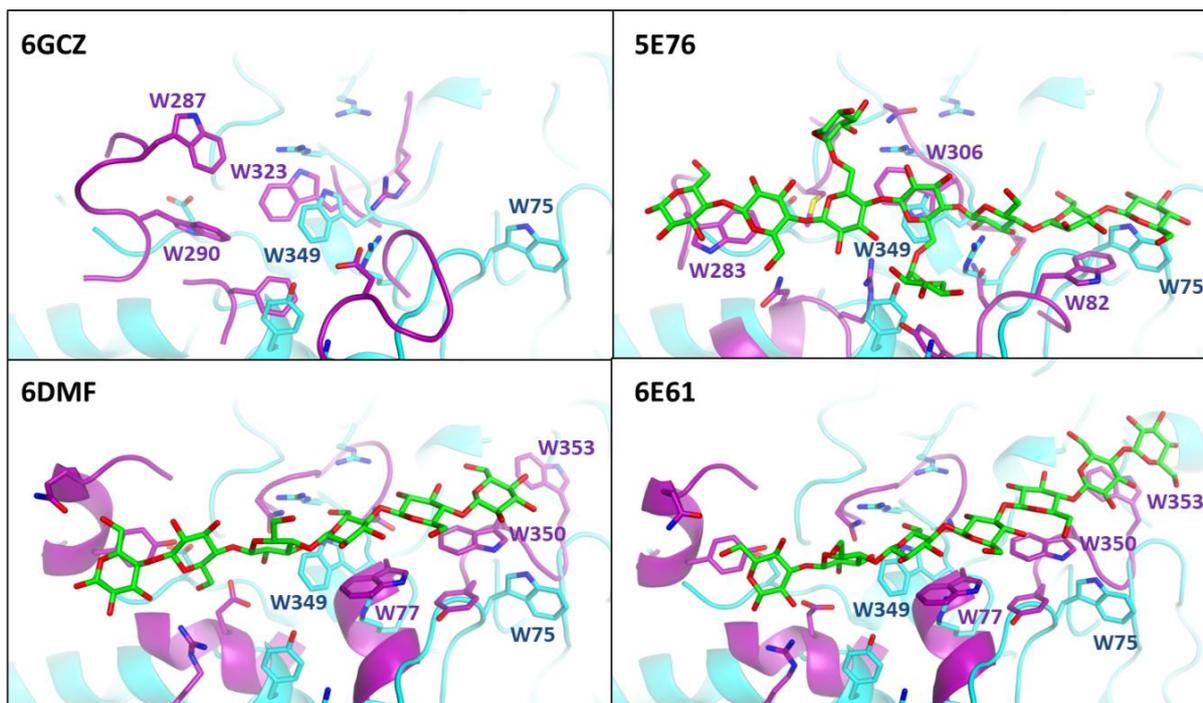
We attempted to co-crystallize and also soak F5\_SusD crystals with XOS ranging from DP2 to DP4, but no complex structure was obtained, confirming the lack of binding to xylo-oligosaccharides. Comparison with other SusD structures with demonstrated sugar binding ability, showed that two of the three essential aromatic residues involved in substrate binding

are present in similar positions in F5\_SusD, such as W75 (equivalent to W82 in the XyG\_SusD used as positive control in this study, PDB ID: 5E76) and W349 (equivalent to W306 in 5E76 and to W323 in 6GCZ) (**Figure 5**). However, while these residues are well exposed and perfectly oriented to bind oligosaccharides in the canonical binding SusD proteins, their localization in F5\_SusD seems either to set back in the structure in a way that makes them less accessible for ligand binding (W349) or the orientation of their side chain is radically different (W75). Other residues that play a role in binding long oligosaccharides (DP higher than 5) in the SusD, such as W353 in the  $\beta$ -glucan SusD (PDB-ID: 6E61) in *B. ovatus* ATCC 8483 (Tamura et al., 2019) or the W283 in the xyloglucan SusD (PDB-ID: 5E76) in *B. ovatus* ATCC 8483 (Tauzin et al., 2016a), lack in F5\_SusD. In addition, the highly flexible loop 292-315 seems to further restrict the surface area for substrate binding in F5\_SusD. Importantly, the four residues N249, D444, A446, and A516 corresponding to D, G, P and V residues in BVU\_0037, respectively, are not located in the region around W75 and W349.



**Figure 4. F5\_SusD crystal structure.**

(A.) Overall structure of F5\_SusD. (B.) Comparison of F5\_SusD with other SusD from *Gramella* sp. MAR\_2010\_102 specific of laminarin (6GCZ), *B. ovatus* specific of xyloglucan (5E76) and mixed linkage glucan (6DMF and 6E61). The stacking platforms identified in 6GCZ, 5E76, 6DMF and 6E61 are showed, with the corresponding residues in F5\_SusD, which are non-effective to bind XOS.



**Figure 5. Comparison between F5\_SusD putative binding area and the homologous structures (same as Figure 4).**

In each panel, F5\_SusD is in cyan, the homologous structure is purple and the ligand is green.

## Discussion

Despite the low sequence identity shared between SusD-like proteins, the SusD/C pair is used as a reference for PULs identification in genomes (Martens et al., 2008; Terrapon et al., 2018). Indeed, substrate recognition and capture is a key step for glycoside utilization and colonization of their habitat by bacteria. So far, most of the SusD-like proteins that have been biochemically characterized are all able to bind the polysaccharide targeted by the PUL, but also to some extent oligosaccharides resulting from the hydrolysis of polysaccharide by cell-exposed enzymes. Only rare examples of SusD-like proteins have been demonstrated as non-binding proteins but their function has been investigated only using AGE or ITC (D'jean et al., 2020; Luis et al., 2018). Here, we fully characterized a SusD (F5\_SusD) from a XOS PUL isolated from an uncultured bacteria, which shares 99.4 % sequence identity with the *B. vulgatus* SusD protein BVU\_0037. Previously, we demonstrated that the PUL containing BVU\_0037 was involved in the ability of *B. vulgatus* to grow on XOS, and to a lesser extent, on branched arabino-xylo-oligosaccharides, while it poorly utilizes xylans (Tauzin et al., 2016b). In the present study, we used a multidisciplinary approach to investigate the binding

ability of the F5\_SusD. Surprisingly, F5\_SusD was not able to bind XOS or AXOS but exhibited a very low affinity for xylan. In contrast, the glycan-binding SusD-like proteins characterized to date display a strong binding affinity, in the millimolar to micromolar range, depending on the length of the tested substrate. In most cases, the values obtained for complex polysaccharides are in the same range than the values obtained with the oligosaccharides resulting from the polysaccharide hydrolysis (Tauzin et al., 2016a; Temple et al., 2017). Rogowski et al. characterized two SusD-like proteins involved in two distinct xylan PULs from *B. ovatus* (BACOVA\_04392 and BACOVA\_03427). Contrary to the F5\_SusD which belongs to a PUL specific for XOS, these SusD-like proteins showed a strong binding ability against xylan, measurable by ITC, but only BACOVA\_04392 binds xylo-oligosaccharides no smaller than DP6 ( $K_D=8.0 \times 10^2 \text{ M}^{-1}$  for xylohexaose) (Rogowski et al., 2015).

Several previous studies showed that SusD-like proteins are essential for the uptake of substrate via the SusC transporter (Koropatkin et al., 2008; Tamura et al., 2019; Tauzin et al., 2016a, 2016b). It is also the case for the F5 SusC/D system, since when F5\_SusD is deleted, the clone F5min\_SUS $\Delta$ SusD cannot grow on XOS anymore. Interestingly, complementation with the F5\_SusD recombinant protein restored the growth, as previously observed for the starch and the xyloglucan systems from *B. thetaiotaomicron* and *B. ovatus*, respectively (Koropatkin et al., 2008; Tauzin et al., 2016a). The F5\_SusC/D pair is thus able to transport XOS, but not to bind them significantly. If the presence of the SusD-like proteins is crucial for glycoside transport, in Bacteroidetes their binding ability is not always required, for the reasons explained hereafter. Indeed, both the presence and binding ability of the SusD homolog from the *B. ovatus* transport system are essential for growth on mixed-linkage  $\beta$ -glucans (Tamura et al., 2019). In contrast, the complementation using non-binding SusD-like mutants is enough to restore the functionality of the starch transport system in *B. thetaiotaomicron* and the xyloglucan one in *B. ovatus* (Koropatkin et al., 2008; Tauzin et al., 2016a). As explained by the authors, it is probably because in Bacteroidetes, the cell surface glycan-binding proteins (such as the SusD-like SGBP\_A protein, and the SusE-like SGBP\_B protein, from *B. ovatus*) functionally complement each other in the capture of the polysaccharide. Indeed, for the rare examples of non-binding SusD-like proteins that have been identified so far from loci involved in pectin and  $\beta$ -1,3-glycan utilization, there are additional SGBP or SusD-like proteins in the same locus, suggesting that they ensured complementation of the system for glycan binding (D'jean et al., 2020; Luis et al., 2018). By using truncated forms of the F5 PUL in *E. coli*, we demonstrated here, for the first time, that

a SusC/D transport system can be functional to uptake the substrate targeted by the XOS PUL without any binding ability of SusD or any other SGBP.

In order to better understand the molecular basis of the non-binding ability of the F5\_SusD, we determined its crystal structure. By comparison with the solved SusD-like structures, we hypothesized that the low binding affinity could be due to unfolded binding surface and the absence or not appropriate orientation of several key residues. Indeed, recognition of oligosaccharides might require more than the two conserved tryptophan residues W75 and W349 that we have identified in the F5\_SusD, although not in a configuration which would allow them to bind XOS. This is not surprising, since SusD affinity is driven by the surface complementation between the protein and the substrate, rather than by affinity for individual chemical groups towards the substrate, as proposed in the description of the very first structure of SusD (Koropatkin et al., 2008) and further confirmed by the later crystal structures (Glenwright et al., 2017; Koropatkin et al., 2009; Larsbrink et al., 2016; Mystkowska et al., 2018; Phansopa et al., 2014; Tamura et al., 2019; Tauzin et al., 2016a). Interestingly, in the putative binding area of F5\_SusD, we can identify several arginines (R339, R347, R351) that form a positively charged patch surrounding the W349, which might serve to bind negatively charged carbohydrates, such as glucuronic acids (**Figure S5**). Maybe, the F5\_SusD would thus have some affinity to particular regions of the xylan polymer that are rich in negatively charged sugars or which present a very particular arrangement of the branching points. These regions are very rare in beechwood xylan so this would explain why the protein has so low affinity. However, in the absence of further binding experiments with glucouronic acids, we don't know whether this would have some biological significance. Finally, the fact that complementation with F5\_SusD restores the growth on XOS shows that despite the F5\_SusD shows no detectable affinity for XOS, it is nevertheless an essential protein for the nutrient uptake machinery. Thus, the role of the “non-binding” F5\_SusD would be to effectively complement the SusC transporter and activate its functionality. This has been proposed for the first time by Glenwright et al. (Glenwright et al., 2017), who solved the crystal structure of a quaternary complex from *B. thetaiotamicron* BT2261-BT2264 which includes the SusD (BT2263) and the SusC transporter (BT2264) (**Figure S6A**). Using a variety of techniques, the authors proposed a “pedal bin” mechanism in which the SusD moves away from SusC in a hinge-like fashion in the absence of ligand to expose the substrate binding site to the extracellular milieu. According to this model, SusC carries out its function only when it is in complex with SusD. To investigate whether this could be the case for the F5\_SusC/D, we have constructed an homology model for the F5\_SusC/D

interaction and compared it to the crystallized complex (**Figure S6B**). Overall the well-defined surface complementarity between the SusD and SusC supports the fact that the SusD is indeed the structural partner of the SusC, and it could therefore provide the same capping function. As already mentioned in the results section, one interesting feature of the F5\_SusD protein without any carbohydrate or protein ligand is the conformation of its N-terminal loop, which is folded back on the protein, with the M21 inserted in a small hydrophobic pocket. Conversely, in the crystal structure of the BT2261-BT2263 complex, the N-term of SusD is pointing towards the membrane. This region is called the “hinge” and is a central region for the interaction SusC-SusD as we can see from the crystal structure of the complex where a small helix from SusC perfectly fills the cavity creating a high complementary interaction. In the F5\_SusD structure, this interaction would not be possible, as it would create an important steric clash. It is probable that, when bound to SusC, the N-terminal loop of F5\_SusD will displace and adopt another conformation, flanking the SusC and pointing towards the membrane. Nevertheless, only the resolution of the crystallographic structure of the SusC/D complex would allow us to demonstrate this hypothesis.

Herein, non-binding SusDs can be found in PULs including i) a SusC/D pair and additional SGBPs, ii) two SusC/D pairs and a SGBP, or iii) only one SusC/D pair (Luis et al., 2018; Dégéan et al., 2020; Present work). The presence of an additional SGBP or a SusC/D pair in the same locus as a non-binding SusD suggests a complementation of the binding function lacking in the related SusD. In contrast, we hypothesize that the PULs containing only a SusC/D pair for which SusD is a non-binding protein, and no other SGBP, might be involved in oligosaccharide utilization relying only on the SusC ability to recognize and transport its cognate substrate, although SusD is required to confer SusC its full functionality.

## **Materials and methods**

### **Heterologous protein expression and purification**

The gene fragment encoding SusD protein was amplified from metagenomic DNA of clone F5 (Genbank accession number HE717017) from metagenomic library issued of human fecal sample by PCR with forward primer, including NdeI restriction site (F5\_SusD\_Cloning\_F), and reverse primer, including XhoI (F5\_SusD\_Cloning\_R) listed in **Table S1**. The first thirty amino acids of SusD corresponding to signal peptide/lipotag as predicted by SignalP 4.1 and LipoP 1.0 servers were not included in the PCR amplicon (Juncker and Willenbrock, 2003;

Petersen et al., 2011). The gene product was cloned into the NdeI and XhoI restriction sites of pET-28a(+) (Novagen, Darmstadt, Germany) preceding an N-terminal 6-His tag for affinity purification and transformed into *E. coli* DH5 $\alpha$  (Invitrogen). The pET-28a(+) expressing SusD was fully sequenced and transformed into *E. coli* BL21(DE3) for expression. A single bacterial colony was inoculated in 5 ml of LB medium containing kanamycin (50  $\mu$ g/ml) at 37  $^{\circ}$ C. Overnight preculture was used to inoculate 200 ml of LB containing 50  $\mu$ g/ml kanamycin with an initial optical density at 600 nm ( $OD_{600\text{ nm}}$ ) of 0.05. The culture was induced by the addition of 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) in the mid-exponential phase ( $OD_{600\text{ nm}}$  of  $\approx$ 0.6), and then incubated for 4 h at 37 $^{\circ}$ C. XyG\_SusD was produced as described by Tauzin et al. (Tauzin et al., 2016a).

For purification of the recombinant F5\_SusD protein, cells were harvested by centrifugation (15 min, 6,000 g), and sonicated in Tris buffer (20 mM Tris-HCl, 300 mM NaCl, pH 8.0). Bacterial debris were cleared by centrifugation at 11,000 g for 30 min at 4  $^{\circ}$ C, then lysates were passed through a column of 2 ml of Talon metal affinity resin (Clontech, USA). F5\_SusD protein was eluted using a gradient of 50 to 200 mM imidazole and then was resuspended in Tris buffer (20 mM Tris-HCl, 150 mM NaCl, pH 7.5) using an Amicon Ultra Filters (Sigma-Aldrich) to eliminate the imidazole. The purity of protein was confirmed by SDS-PAGE and the concentration was determined by the absorbance at 280 nm using Thermo Scientific<sup>TM</sup> Nanodrop 2000 with an extinction coefficient of 153,350  $M^{-1}cm^{-1}$ . F5\_SusD protein was kept at 4  $^{\circ}$ C for later analysis. Protein used for crystallization experiments was purified using an ÄKTExpress system (GE Healthcare) with an affinity step (Talon Crude 1ml, GE Healthcare) using the same loading buffer (20 mM Tris-HCl, 300 mM NaCl, pH 8.0), eluted with a 250mM step of imidazole followed by a gel filtration step (HiPrep 16/60 Sephacryl S-200 HR) in 20mM TRIS-HCl, 0.15M NaCl, pH 7.5. The fractions containing the monomeric F5\_SusD protein were pooled, concentrated and stored at 4  $^{\circ}$ C.

### **Isothermal titration calorimetry**

Isothermal titration calorimetry (ITC) of glycan binding by F5\_SusD was tested using the ITC200 (Malvern) calorimeter calibrated to 25  $^{\circ}$ C. Proteins (40  $\mu$ M) were prepared in 20 mM HEPES-100 mM NaCl (pH 7.0), and oligosaccharides were prepared using the same buffer. The F5\_SusD was placed in the sample cell and the syringe was loaded with 2-10 mM XOS. The resulting heat of reaction was recorded.

## **Affinity gel electrophoresis**

The ability of F5\_SusD to bind polysaccharides was assayed by affinity gel electrophoresis. Continuous native polyacrylamide gels were prepared consisting of 10% (w/v) acrylamide in 25 mM Tris, 250 mM glycine buffer, pH 8.3. Polysaccharides of interest were added prior to polymerization at final concentration of 0.1% (w/v). Two and half  $\mu\text{g}$  of purified F5\_SusD were loaded on the gels. Electrophoresis was carried out for 90 min at room temperature. BSA was used as non-interacting negative control protein.

## **NMR spectroscopy**

All experiments were done on a Bruker Avance II 800MHz NMR spectrometer equipped with a QCPI cryogenically cooled probe head. Spectra were recorded at 298K, and all samples were in a Tris buffer (Tris 20 mM and NaCl 150 mM at pH 7.5). 1D proton spectra were acquired with a Watergate water suppression (Piotto et al., 1992), with 16 scans for XG\_SusD (at 327  $\mu\text{M}$ ) and 256 scans for P5\_SusD (at 54  $\mu\text{M}$ ). STD experiments were performed with saturation of the protein resonances at 0 ppm through a train of 5ms Gaussian 180° pulses (Ley et al., 2014; Mayer and Meyer, 1999). Typical experiments were run with 256 scans, 2048 acquisition points and a 5s relaxation delay (including the 3 second presaturation train). STD experiments were run with 256 or 512 scans. Spectra were transformed after one level of zero filling and apodization with a  $\pi/3$  shifted square sine bell.

## ***E. coli* growth experiments**

All the *E. coli* cells were grown in M9 medium (Na<sub>2</sub>HPO<sub>4</sub> · 12H<sub>2</sub>O 17.4 g/l, KH<sub>2</sub>PO<sub>4</sub> 3.03 g/l, NaCl 0.51 g/l, NH<sub>4</sub>Cl 2.04 g/l, MgSO<sub>4</sub> 0.49 g/l, CaCl<sub>2</sub> 4.38 mg/l, Na<sub>2</sub>EDTA 2H<sub>2</sub>O 15 mg/l, ZnSO<sub>4</sub> · 7H<sub>2</sub>O 4.5 mg/l, CoCl<sub>2</sub> · 6H<sub>2</sub>O 0.3 mg/l, MnCl<sub>2</sub> · 4H<sub>2</sub>O 1 mg/l, H<sub>3</sub>BO<sub>3</sub> 1mg/l, Na<sub>2</sub>MoO<sub>4</sub> · 2H<sub>2</sub>O 0.4 mg/l, FeSO<sub>4</sub> · 7H<sub>2</sub>O 3 mg/l, CuSO<sub>4</sub> · 5H<sub>2</sub>O 0.3 mg/l, thiamine 0.1 g/l and leucine 0.02 g/l) supplemented with 12.5 mg/l chloramphenicol. The carbohydrates were diluted in M9 medium at 0.5% (w/v). The cultures were inoculated at an initial optical density at 600 nm (OD<sub>600 nm</sub>) of 0.05 from precultures in LB.

## **Glycosides**

The mixture of XOS are with DP from 2 to 7 (WAKO and IOR- TAIHE,). Simple xylans, with sparsely decorated structures, were purchased from Sigma for beechwood xylan and from Megazyme for wheat arabinoxylan. Xylobiose, xylotriose, 3<sup>2</sup>- $\alpha$ -L-arabinofuranosyl-

xylobiose (AX2), xyloglucan (DP 7-9) and higher DP xyloglucan oligosaccharides (XGHDP) were purchased from Megazyme.

## **Crystallization and data collection**

Purified F5\_SusD was concentrated to 20mg/ml and initial crystallization conditions were screened using a Mosquito robot (TPP Labtech) and the JCSG I-IV commercial screens (Qiagen) from which one starting condition was identified (1.26M Sodium Citrate, 0.09M HEPES-HCl, 10% Glycerol, pH 7.5). After manual condition optimization, F5\_SusD crystals were grown using the hanging drop method by mixing 1 $\mu$ L of protein solution with 1 $\mu$ L of precipitant solution (1.7M Na<sub>3</sub>Citrate, 10% Glycerol, 0.1M HEPES pH 7.5) and incubated on a 24 well crystallization plate at 12 °C. After harvesting, the crystals were soaked in cryo-protectant buffer (1.7M Na<sub>3</sub>Citrate, 15% Glycerol, 0.1M HEPES pH 7.5) with or without the presence of 100mM Xylotriose and flash-frozen in liquid N<sub>2</sub>. Data collection was performed at XALOC beamline of the ALBA synchrotron (Barcelona, Spain). A long wavelength dataset has also been collected with a wavelength of 2.0Å, by using a special setup of the XALOC in which the crystal is continuously flowed in Helium atmosphere during the data collection to reduce absorption from air.

## **Structure Solution and refinement**

First using the native dataset only, the structure has been solved using the webserver MORDA (Hungler et al., 2016) from which a low homology molecular replacement solution was identified. However, the very low sequence homology resulted in poor electron density maps that were very difficult to interpret. The starting MR model has been combined with the long wavelength dataset as implemented in the program Phaser and used to calculate improved phases, also allowing to correctly position all the sulphur atoms in the sequence. The structure has been traced by several rounds of Arp/Warp webserver (Perrakis et al., 1997) alternated to manual rebuilding using Coot (Emsley et al., 2010) and refined by restrained refinement using Refmac (Murshudov et al., 2011).

## **References**

Amorim, C., Silv ério, S.C., Prather, K.L.J., Rodrigues, L.R., 2019. From lignocellulosic residues to market: Production and commercial potential of xylooligosaccharides. *Biotechnol. Adv.* 0–1. <https://doi.org/10.1016/j.biotechadv.2019.05.003>

- Bell, A., Brunt, J., Crost, E., Vaux, L., Nepravishta, R., Owen, C.D., Latousakis, D., Xiao, A., Li, W., Chen, X., Walsh, M.A., Claesen, J., Angulo, J., Thomas, G.H., Juge, N., 2019. Elucidation of a sialic acid metabolism pathway in mucus-foraging *Ruminococcus gnavus* unravels mechanisms of bacterial adaptation to the gut. *Nat. Microbiol.* 4, 2393–2404. <https://doi.org/10.1038/s41564-019-0590-7>
- Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., Zardecki, C., 2000. The protein data bank. *Nucleic Acids Res.* 28, 235–242. <https://doi.org/10.1107/S0907444902003451>
- Cameron, E.A., Kwiatkowski, K.J., Lee, B., Hamaker, B.R., Koropatkin, N.M., Martens, C., 2014. Multifunctional nutrient-binding proteins adapt human symbiotic enhanced sensing and catalysis. *MBio* 5, e01441-14. <https://doi.org/10.1128/mBio.01441-14>.
- Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré J., Henrissat, B., Remaud-Simón, M., Monsan, P., Potocki-Véronèse, G., 2013. Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One* 8, e72766. <https://doi.org/10.1371/journal.pone.0072766>
- Cho, K.H., Salyers, A.A., 2001. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* 183, 7224–7230. <https://doi.org/10.1128/JB.183.24.7224-7230.2001>
- Déjean, G., Tamura, K., Cabrera, A., Jain, N., Pudlo, N.A., Pereira, G., Viborg, A.H., Van Petegem, F., Martens, E.C., Brumer, H., 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, 1–21. <https://doi.org/10.1128/mBio.00095-20>
- Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Chambon, C., Yeoman, C.J., Berg Miller, M.E., Fields, C.J., Martens, E., Terrapon, N., Henrissat, B., White, B.A., Mosoni, P., 2016. Xylan degradation by the human gut *Bacteroides xylanisolvens* XB1AT involves two distinct gene clusters that are linked at the transcriptional level. *BMC Genomics* 17, 1–14. <https://doi.org/10.1186/s12864-016-2680-8>

- Dodd, D., Moon, Y.H., Swaminathan, K., Mackie, R.I., Cann, I.K.O., 2010. Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic bacteroidetes. *J. Biol. Chem.* 285, 30261–30273.  
<https://doi.org/10.1074/jbc.M110.141788>
- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 486–501.  
<https://doi.org/10.1107/S0907444910007493>
- Foley, M.H., Cockburn, D.W., Koropatkin, N.M., 2016. The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell. Mol. Life Sci.* 73, 2603–2617.  
<https://doi.org/10.1007/s00018-016-2242-x>
- Gibson, G.R., Hutkins, R., Sanders, M.E., Prescott, S.L., Reimer, R.A., Salminen, S.J., Scott, K., Stanton, C., Swanson, K.S., Cani, P.D., Verbeke, K., Reid, G., 2017. Expert consensus document: the International Scientific Association for Probiotics and Prebiotic (ISAPP) consensus statement on the definition and scope of prebiotics. *Nat. Rev. Gastroenterol. Hepatol.* 14, 491–502. <https://doi.org/10.1038/nrgastro.2017.75>
- Glenwright, A.J., Pothula, K.R., Bhamidimarri, S.P., Chorev, D.S., Baslé A., Firbank, S.J., Zheng, H., Robinson, C. V., Winterhalter, M., Kleinekathöfer, U., Bolam, D.N., Van Den Berg, B., 2017. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* 541, 407–411. <https://doi.org/10.1038/nature20828>
- Goh, Y.J., Klaenhammer, T.R., 2015. Genetic mechanisms of prebiotic oligosaccharide metabolism in probiotic microbes. *Annu. Rev. Food Sci. Technol.* 6, 137–156.  
<https://doi.org/10.1146/annurev-food-022814-015706>
- Holm, L., 2019. Benchmarking fold detection by DaliLite v . 5. *Bioinformatics* 35, 5326–5327. <https://doi.org/doi: 10.1093/bioinformatics/btz536>.
- Hungler, A., Momin, A., Diederichs, K., Arold, S.T., 2016. ContaMiner and ContaBase: A webserver and database for early identification of unwantedly crystallized protein contaminants. *J. Appl. Crystallogr.* 49, 2252–2258.  
<https://doi.org/10.1107/S1600576716014965>

- Juncker, A., Willenbrock, H., 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662. <https://doi.org/10.1110/ps.0303703>.
- Koropatkin, N., Martens, E.C., Gordon, J.I., Smith, T.J., 2009. The structure Of a SusD homolog, BT1043, involved in mucin O- glycan utilization in a prominent human gut symbiont. *Biochemistry* 23, 1–7. <https://doi.org/10.1038/jid.2014.371>
- Koropatkin, N.M., Martens, E.C., Gordon, J.I., Smith, T.J., 2008. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* 16, 1105–1115. <https://doi.org/10.1016/j.str.2008.03.017>
- Larsbrink, J., Zhu, Y., Kharade, S.S., Kwiatkowski, K.J., Eijsink, V.G.H., Koropatkin, N.M., McBride, M.J., Pope, P.B., 2016. A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnol. Biofuels* 9, 1–16. <https://doi.org/10.1186/s13068-016-0674-z>
- Leth, M.L., Ejby, M., Workman, C., Ewald, D.A., Pedersen, S.S., Sternberg, C., Bahl, M.I., Licht, T.R., Aachmann, F.L., Westereng, B., Hachem, M.A., 2018. Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* 3, 570–580. <https://doi.org/10.1038/s41564-018-0132-8>
- Ley, N.B., Rowe, M.L., Williamson, R.A., Howard, M.J., 2014. Optimising selective excitation pulses to maximise saturation transfer difference NMR spectroscopy. *RSC Adv.* 4, 7347–7351. <https://doi.org/10.1039/c3ra46246c>
- Lombard, V., Ramulu H.G., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. <https://doi.org/10.1093/nar/gkt1178>
- Luis, A.S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., Baslé A., Cartmell, A., Terrapon, N., Stott, K., Lowe, E.C., McLean, R., Shearer, K., Schückel, J., Venditto, I., Ralet, M.C., Henrissat, B., Martens, E.C., Mosimann, S.C., Abbott, D.W., Gilbert, H.J., 2018. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat. Microbiol.* 3, 210–219. <https://doi.org/10.1038/s41564-017-0079-1>

- Martens, E.C., Chiang, H.C., Gordon, J.I., 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 5, 447–457. <https://doi.org/10.1016/j.chom.2008.09.007>
- Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., Gordon, J.I., 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* 9. <https://doi.org/10.1371/journal.pbio.1001221>
- Mayer, M., Meyer, B., 1999. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chemie - Int. Ed.* 38, 1784–1788. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990614\)38:12<1784::AID-ANIE1784>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q)
- McNulty, N.P., Wu, M., Erickson, A.R., Pan, C., Erickson, B.K., Martens, E.C., Pudlo, N.A., Muegge, B.D., Henrissat, B., Hettich, R.L., Gordon, J.I., 2013. Effects of Diet on Resource Utilization by a Model Human Gut Microbiota Containing *Bacteroides cellulosilyticus* WH2, a Symbiont with an Extensive Glycobiome. *PLoS Biol.* 11. <https://doi.org/10.1371/journal.pbio.1001637>
- Mendis, M., Martens, E.C., Simsek, S., 2018. How Fine Structural Differences of Xylooligosaccharides and Arabinoxylooligosaccharides Regulate Differential Growth of *Bacteroides* Species. *J. Agric. Food Chem.* 66, 8398–8405. <https://doi.org/10.1021/acs.jafc.8b01263>
- Mirande, C., Kadlecikova, E., Matulova, M., Capek, P., Bernalier-Donadille, A., Forano, E., B éra-Maillet, C., 2010. Dietary fibre degradation and fermentation by two xylanolytic bacteria *Bacteroides xylanisolvens* XB1AT and *Roseburia intestinalis* XB6B4 from the human intestine. *J. Appl. Microbiol.* 109, 451–460. <https://doi.org/10.1111/j.1365-2672.2010.04671.x>
- Murshudov, G.N., Skub ák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 355–367. <https://doi.org/10.1107/S0907444911001314>

- Mystkowska, A.A., Robb, C., Vidal-Melgosa, S., Vanni, C., Fernandez-Guerra, A., H öhne, M., Hehemann, J.H., 2018. Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan-binding protein of a marine Bacteroidetes. *FEBS J.* 285, 4465–4481. <https://doi.org/10.1111/febs.14674>
- Perrakis, A., Sixma, T.K., Wilson, K.S., Lamzin, V.S., 1997. wARP: Improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 53, 448–455. <https://doi.org/10.1107/S0907444997005696>
- Petersen, T.N., Brunak, S., Von Heijne, G., Nielsen, H., 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. <https://doi.org/10.1038/nmeth.1701>
- Phansopa, C., Roy, S., Rafferty, J.B., Douglas, C.W.I., Pandhal, J., Wright, P.C., Kelly, D.J., Stafford, G.P., 2014. Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling Bacteroidetes species. *Biochem. J.* 458, 499–511. <https://doi.org/10.1042/BJ20131415>
- Piotto, M., Saudek, V., Sklenář, V., 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661–665. <https://doi.org/10.1007/BF02192855>
- Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Basl é A., Morland, C., Day, A.M., Zheng, H., Rogers, T.E., Thompson, P., Hawkins, A.R., Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J., Bolam, D.N., 2015. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* 6. <https://doi.org/10.1038/ncomms8481>
- Salyers, A.A., Gherardini, F., O'Brien, M., 1981. Utilization of Xylan by two species of human colonic *Bacteroides*. *Appl. Environ. Microbiol.* 41, 1065–1068. <https://doi.org/10.1128/aem.41.4.1065-1068.1981>
- Selvendran, R.R., 1984. The plant cell wall as a source of dietary fiber: chemistry and structure. *Am. J. Clin. Nutr.* 39, 320–337. <https://doi.org/10.1093/ajcn/39.2.320>

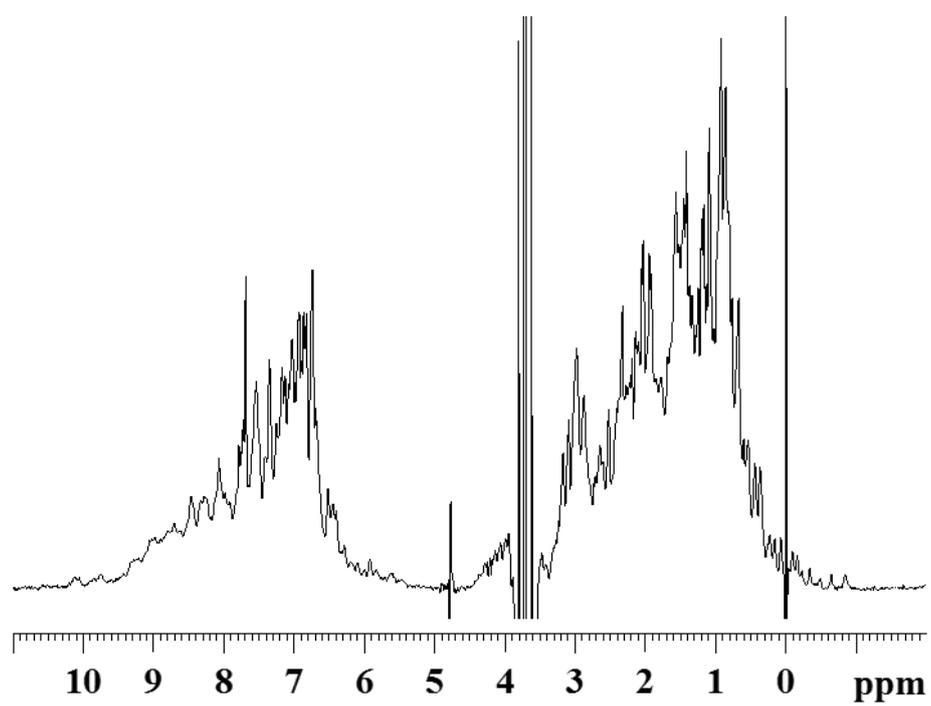
- Singh, R.D., Banerjee, J., Arora, A., 2015. Prebiotic potential of oligosaccharides: A focus on xylan derived oligosaccharides. *Bioact. Carbohydrates Diet. Fibre* 5, 19–30.  
<https://doi.org/10.1016/j.bcdf.2014.11.003>
- Skubak, P., Arac, D., Bowler, M.W., Correia, A.R., Hoelz, A., Larsen, S., Leonard, G.A., McCarthy, A.A., McSweeney, S., Mueller-Dieckmann, C., Otten, H., Salzman, G., Pannua, N.S., 2018. A new MR-SAD algorithm for the automatic building of protein models from low-resolution X-ray data and a poor starting model. *IUCrJ* 5, 166–171.  
<https://doi.org/10.1107/S2052252517017961>
- Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N., Sonnenburg, J.L., 2010. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* 141, 1241–1252.  
<https://doi.org/10.1016/j.cell.2010.05.005>
- Tamura, K., Foley, M.H., Gardill, B.R., Dejean, G., Schnizlein, M., Bahr, C.M.E., Louise Creagh, A., van Petegem, F., Koropatkin, N.M., Brumer, H., 2019. Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*. *Cell. Mol. Life Sci.* <https://doi.org/10.1007/s00018-019-03115-3>
- Tang, K., Lin, Y., Han, Y., Jiao, N., 2017. Characterization of potential polysaccharide utilization systems in the marine *Bacteroidetes Gramella flava* JLT2011 using a multi-omics approach. *Front. Microbiol.* 8, 1–13. <https://doi.org/10.3389/fmicb.2017.00220>
- Tauzin, A.S., Kwiatkowski, K.J., Orlovsky, N.I., Smith, C.J., Creagh, A.L., Haynes, C.A., Wawrzak, Z., Brumer, H., Koropatkin, N.M., 2016a. Molecular dissection of xyloglucan recognition in a prominent human gut symbiont. *MBio* 7, e02134-15.  
<https://doi.org/10.1128/mBio.02134-15>
- Tauzin, A.S., Laville, E., Xiao, Y., Nouaille, S., Le Bourgeois, P., Heux, S., Portais, J.C., Monsan, P., Martens, E.C., Potocki-Veronese, G., Bordes, F., 2016b. Functional characterization of a gene locus from an uncultured gut *Bacteroides* conferring xylo-oligosaccharides utilization to *Escherichia coli*. *Mol. Microbiol.* 102, 579–592.  
<https://doi.org/10.1111/mmi.13480>

- Temple, M.J., Cuskin, F., Baslé A., Hickey, N., Speciale, G., Williams, S.J., Gilbert, H.J., Lowe, E.C., 2017. A Bacteroidetes locus dedicated to fungal 1,6--glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6--glucanase. *J. Biol. Chem.* 292, 10639–10650. <https://doi.org/10.1074/jbc.M117.787606>
- Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H.J., Henrissat, B., 2018. PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* 46, D677–D683. <https://doi.org/10.1093/nar/gkx1022>
- Wang, Z., Tauzin, A.S., Lavillea, E., Tedescoa, P., L'aisea, F., Terraponb, N., Lepercqqa, P., Mercadea, M., Potocki-Veronese, G., 2020. Harvesting of prebiotic fructooligosaccharides by non-beneficial human gut bacteria. *mSphere* 5, e00771-19. <https://doi.org/https://doi.org/10.1128/mSphere.00771-19>
- Zhang, M., Chekan, J.R., Dodd, D., Hong, P.Y., Radlinsk, L., Revindran, V., Nair, S.K., Mackie, R.I., Cann, I., 2014. Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 111. <https://doi.org/10.1073/pnas.1406156111>

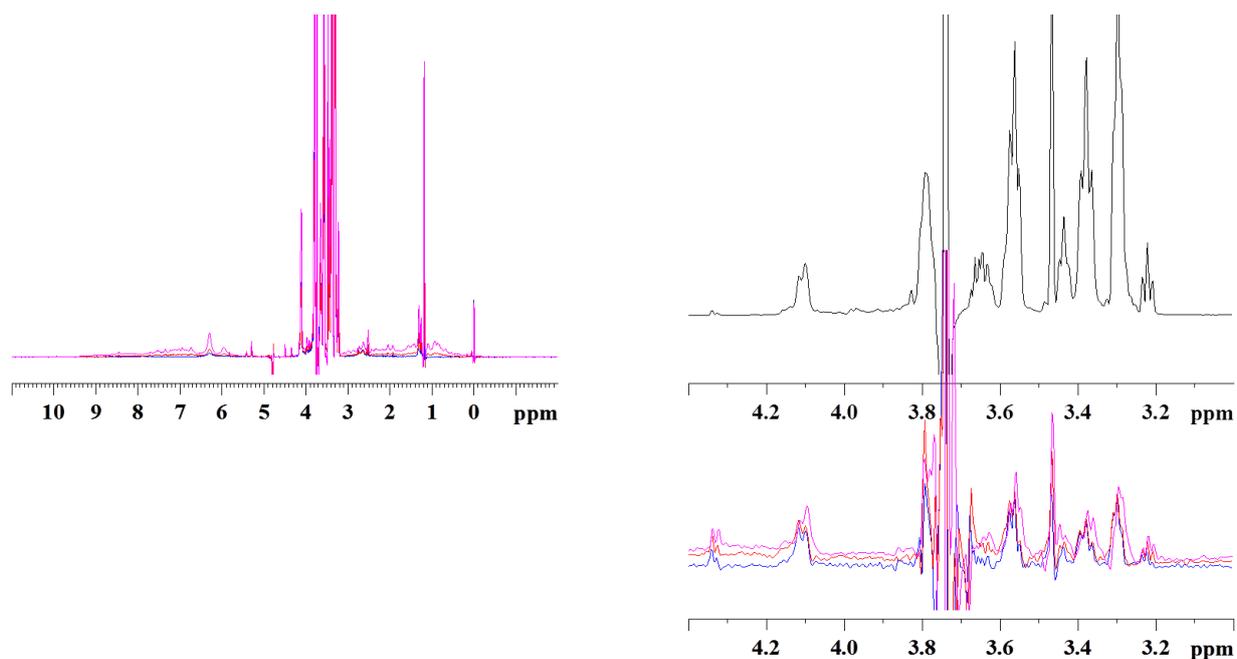
## Supplemental data

Table S1. List of primers for F5 used in this study.

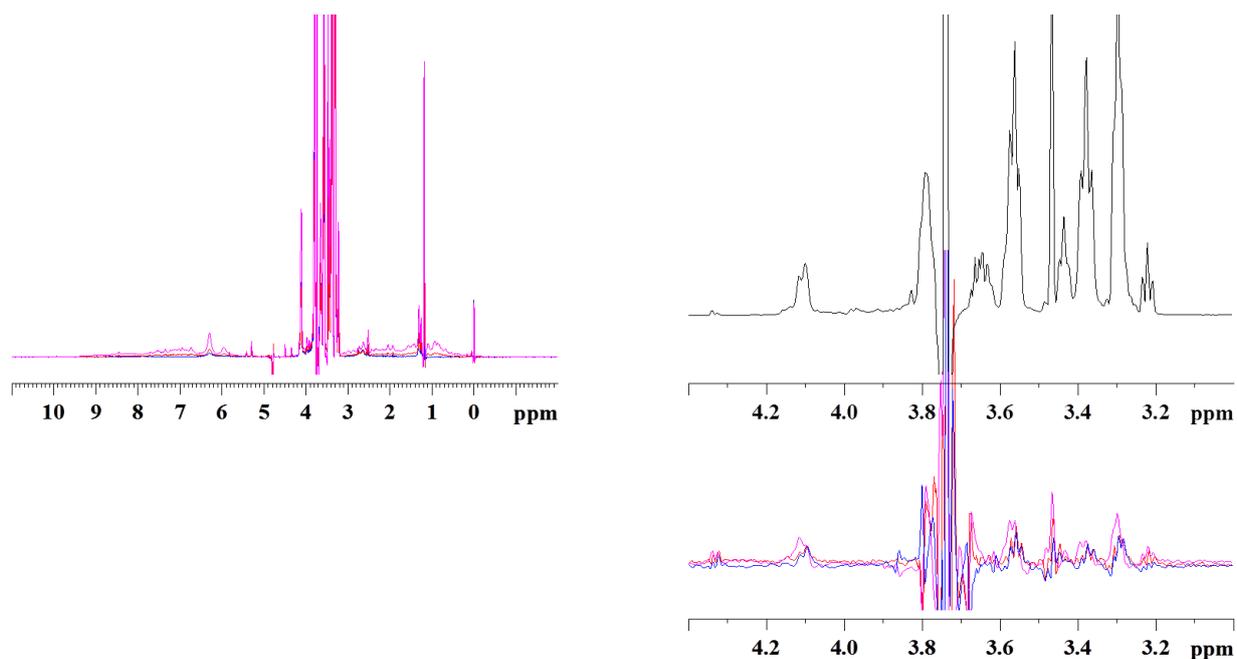
| Name                     | Sequence 5'→3'                              |
|--------------------------|---|
| F5 variants construction |   |
| Pcc1fos_F                | GTGGGATCCTCTAGAGTCGACC                      |
| Pcc1fos_R                | GTGGGATCCCCGGGTACCGAGC                      |
| F5min_SusC/D_F           | ACCCGGGGATCCCACCGATAAGTTTATCAGGAGGAAGCCAC   |
| F5min_SusC_F             | ACCCGGGGATCCCACAATGTCTGAACAGGATGCTGTCAGG    |
| F5min_SusC/D_R           | TCTAGAGGATCCCACACAAGCCTGAATACAGCGGACATC     |
| SusD cloning             |   |
| F5_SusD_Cloning_F        | CCGCGCGGCAGCCATATGGACGAGCAGCCGCGCAGCAGTTATG |
| F5_SusD_Cloning_R        | GTGGTGGTGGTGCTCGAGTCAATCTCTGTATCCCGGATTCTG  |



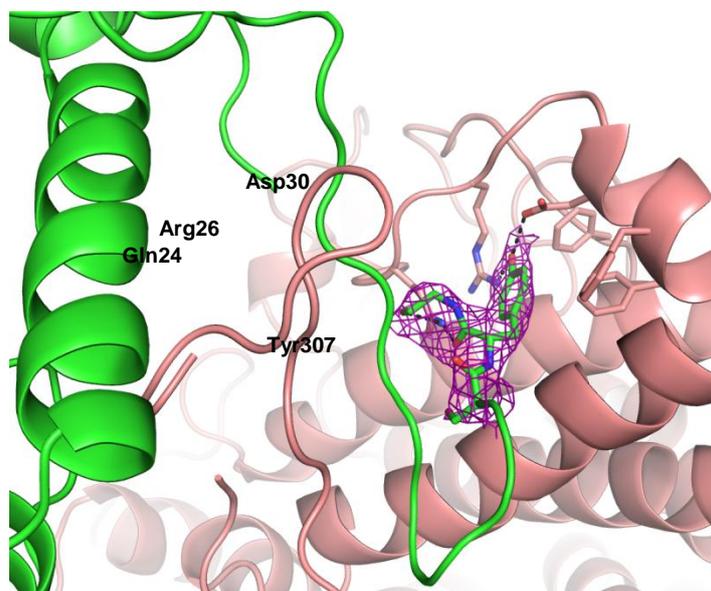
**Figure S1. 1D proton spectrum at 298K of 500  $\mu$ l of P5\_SusD protein  $\sim$ 3.8mg/ml (=54  $\mu$ M) in Tris buffer.**



**Figure S2. (left) 1D proton spectrum at 298K of 500  $\mu$ l of xylan@500  $\mu$ M with P5-SusD at 0 (blue), 15 (red) or 27  $\mu$ M (magenta) in Tris buffer. (right) STD spectra of the different samples with saturation at 0ppm for the isolated xylan (blue), in presence of 15 (red) or 27  $\mu$ M (magenta) of P5-SusD. The reference xylan spectrum is shown in black.**

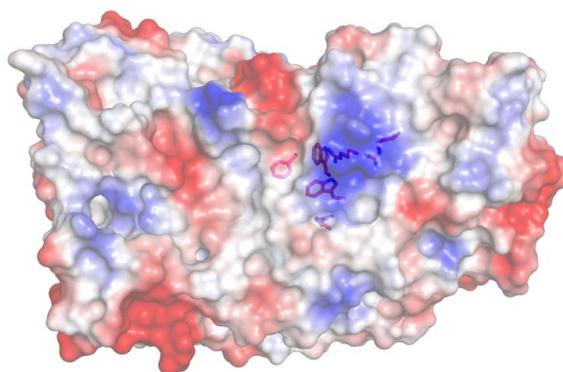


**Figure S3. (left) 1D proton spectrum of 500 μl of xylan@500 μM with P5-SusD at 0 (blue), 15 (red) or 27 μM (magenta) in Tris buffer  $T_e= 298K$  (right) STD spectra of the different samples with saturation at 10.1ppm for the isolated xylan (blue), in presence of 15 (red) or 27 μM (magenta) of P5-SusD. The reference xylan spectrum is shown in black.**



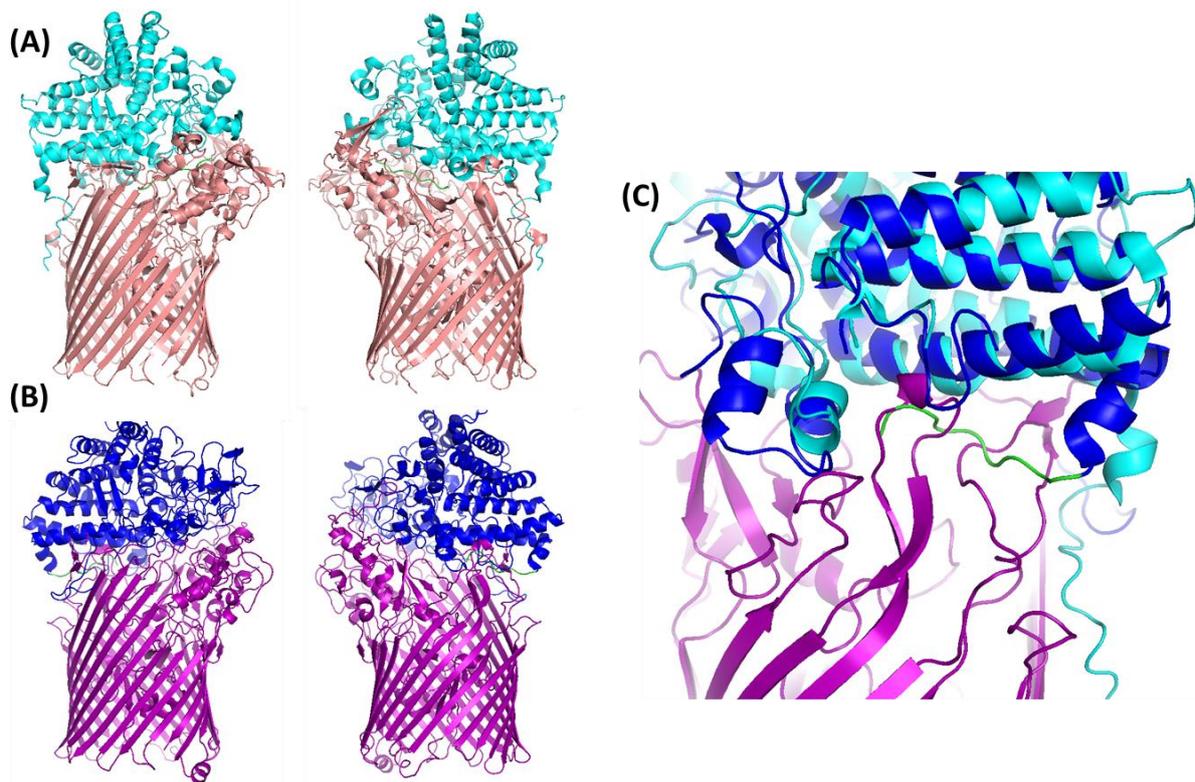
**Figure S4. Close-up view of the interaction between the loop 292-315 and a symmetry related molecule.**

Tyr307 is bound in a small amphipathic cavity and stabilized by hydrogen bonds with Asp30, Arg26 and Gln24.



**Figure S5. Overall structure of the F5\_SusD with the charged area colored according to a gradient from positive red to negative blue.**

The amino acids of the highly positively charged area are shown in sticks.



**Figure S6. Study of the possible interaction between SusC and SusD.**

(A) Crystal structure of the BT2261-BT2264 complex (PDBID: 5FQ8). (B) Homology model of the F5\_SusD\_SuC complex. The F5\_SusC model has been constructed using Phyre2 server. (C) Close view of the N-terminal region of F5\_Sud (green) to show the steric clash with the SusC (purple). In cyan, the N-terminal region in the crystallized complex BT2261-BT2264 (PDBID: 5FQ8).



# **Conclusion and perspectives**



In the framework of the work presented in this thesis manuscript, we developed novel, highly generic methods to fasten the identification of transporters, and to profile their specificities. We applied these techniques to the discovery of oligosaccharide transporters involved in the utilization of dietary fibers by uncultured human gut bacteria, and to the deciphering of their binding and translocation mechanisms.

Identification of glycoside transporters from uncultured bacteria was one of the main objectives of this project. The first part of this PhD work was dedicated to the evaluation of the efficiency and genericity of positive selection to screen *E. coli* metagenomic libraries for functional transporters. To establish the proof of concept of this approach, we scrutinized the sequences of 102 metagenomic clones issued from the human intestinal and bovine rumen microbiomes, which were previously identified by screening four libraries totalizing 216,320 clones for CAZyme activity. We selected 45 clones which contain metagenomic loci encoding at least one putative glycoside transporter and one CAZyme. Then, 129 growth tests were performed in micro-plates on minimal medium containing oligosaccharides which could be hydrolyzed by the CAZymes identified in each clone. Finally, 13 hit clones showed growth ability on FOS, XOS, aldouronic acids, cello-oligosaccharides, mixed-linkage type  $\beta$ -glucooligosaccharides, laminatriose, mannoooligosaccharides, galactomannooligosaccharides, and lactulose, thanks to the co-expression of genes encoding transporters and CAZymes. These assays provided direct evidence of the functionality of transporters from various families, despite their differences in specificity, taxonomical origin and mechanism. Thanks to these results, we evidenced a novel function for MFS transporter that is the uptake of galacto-mannooligosaccharides. We also proved that it is possible to produce functional transporters by expressing PULs' genes from Gram-positive bacteria in the Gram-negative bacterium *Escherichia coli*, highlighting the genericity of this strategy, although the precise location of each of the transporter domains in the *E. coli* cell is still not determined. We calculated that the probability to find functional glycoside-transporters from large-insert metagenomic libraries using the *E.coli* host is only 0.7%, and the probability to produce both a functional glycosidase and a functional transporter from a (meta)genomic PUL cloned into a fosmid is only 1.8%. This highlights the importance of automated screening to identify functional transporters from metagenomes, and more generally, from bacterial genomes, using PUL cloning in fosmids and expression in *E. coli*. The use of other bacterial hosts and of other vectors to control the expression of each of the PUL genes might allow us to increase these hit yields. The throughput and sensitivity of this screening method could also be increased using

microfluidics, with which each assay could be miniaturized in pico-liter droplets. Such a miniaturization by several orders of magnitude will allow screening for the metabolization of highly costly substrates, such as human glycosides for instance. Finally, by using the approach presented here, we cannot definitively prove that a transporter is responsible for the uptake of a target glycoside nor describe its mechanism. This indeed requires rational truncation of the PUL, in order to eliminate any potential bias due to the presence in the metagenomic DNA fragment of several transporter encoding genes. Because this cannot be done at high-throughput, we had to select the targets for in-depth biochemical characterization. We thus chose some of the most original glycoside transporters systems, in particular those of which the function could be important for human health.

This is the reason why, in the frame of this thesis, we were particularly interested in two fructooligosaccharide utilization loci, identified in the clones I9 and F1, both issued from uncultured Clostridiales strains. Indeed, FOS are increasingly being used as prebiotics in human alimentation and livestock feeding. The F1 and I9 FOS utilization loci present almost perfect syntenies with parts of the genomes of *Subdoligranulum* sp. APC924/74 and *Dorea longicatena* DSM 13814, respectively. Several isolates from *Subdoligranulum* species are short-chain fatty acid producers, and thus, could be considered as rather beneficial to the human host. On the contrary, *Dorea longicatena* was previously described as non-beneficial, which correlates to the fact that we found the I9 locus overabundant and over prevalent in the metagenome of patients suffering from inflammatory bowel diseases. The F1 locus contains only one transporter from the PTS family, but which lacks the IIA domain compared with the I9 PTS and other canonical PTSs, which include the three domains specific for glycosides (the EIIB, C and A domains). The I9 locus contains both a PTS transporter and an ABC transporter. After the screening step, it was thus impossible to determine which one of the two transporters was responsible for the substrate internalization in the growth assays. First, the I9 PUL truncation revealed that the ABC transporter, which lacks a solute binding protein, confers a weak but significant ability to *E. coli* to use sucrose. How this transporter could uptake this glycoside without solute binding protein is still an open question. Then, by combining locus engineering, growth dynamic analyses and mass spectrometry, we demonstrated that both PTS are responsible for the binding, transport and phosphorylation of kestose. The F1 and I9 PTSs are the first FOS PTSs of which the precise molecular mechanism and the exact specificity towards oligosaccharides of precise lengths have been determined. To do so, we developed a highly generic method to quantify the binding and

uptake rates of transporters, based on the use of crude cell debris and intact cells, and of non-labelled substrates of which very weak concentration variations (less than one nanomole) are monitored by HPAEC-PAD. This method will allow one to circumvent the use of radio-labelled substrates, of which the use requires dedicated facilities, which are less and less current in biochemistry and microbiology labs. Furthermore, specific mutations in the *E.coli* genome revealed that the F1 and I9 recombinant PTSs are functional in *E.coli* thanks to the complementarity of several non-specific and specific components of the glucose and fructose *E. coli* PTSs, highlighting the promiscuity of the PTS components of Gram-positive and negative bacteria. How could the EIIA-lacking F1 PTS properly function in the native strains from which it is issued is still an open question, that could be answered to by examining the transcriptome of *Subdoligranulum* sp. APC924/74 during growth on FOS (if this strain could grow on this substrate). Besides, in this study focused on PTSs, we used a sequence similarity network to analyze the functional, structural and taxonomical diversity of PTS glycoside transporters. We revisited the classification of the TCDB, and recommend one to be cautious with respect to the level of classification. The last three TCID components which are, in most cases, related to substrates, should indeed not be used to predict the substrate specificity of PTSs, and more generally, of glycoside transporters.

Another part of this thesis project was dedicated to the characterization of a SusD protein of a polysaccharide utilization locus (PUL) specific for xylooligosaccharides (XOS), identified in the metagenomic clone F5. This XOS PUL, issued from an uncultured gut *Bacteroides* strain, and which is quasi-identical to that of *Bacteroides vulgatus* ATCC 8482, was previously characterized in *E. coli* and in the native strain by the DiscOmics group, in collaboration with that of Eric Martens. Here, using a large variety of methods, we revealed that very surprisingly, the F5\_SusD protein cannot bind XOS, which are however efficiently uptaken by the SusCD system. These methods include those usually used to characterize surface glycan binding proteins, such as isothermal titration calorimetry and affinity gel electrophoresis. In addition, we used STD-NMR, which has never been used for the SusD-like proteins. Once again, these techniques are generic, extending the panel of techniques that can be used, from now, to characterize the specificity of glycoside-transporters. In addition, complementation experiments confirmed that the presence of the F5\_SusD is essential for XOS uptake. Moreover, the crystallographic structure of the F5\_SusD was solved at a 2.6Å resolution in *apo* form, which represents the first crystal structure of SusD protein of a XOS PUL. Compared with other solved structures of SusD-like proteins, the F5\_SusD revealed an

unfolded binding surface and the absence or inappropriate orientation of several key residues such as tryptophan residues known to be involved in binding. These results highlight that the binding specificity of SusD does not necessarily reflect the uptake specificity of the transporter. They also constitute the first evidence that SusCD transporters can uptake oligosaccharides without any functional surface glycan binding protein produced in the cell to capture the uptaken oligosaccharides. The term ‘surface glycan binding protein’ used for SusD-like proteins is thus not generic, being not adequate to describe some of these proteins. Given that most PUL systems that have been characterized up to now target polysaccharides and not oligosaccharides, contrary to the XOS PUL characterized here, we still do not know if non-binding SusDs are a specific trait of oligosaccharide SusCD transporters. We also wondered if this protein could bind other oligo- or polysaccharides, especially since we noticed a positively charged surface that could, hypothetically, accommodate negatively charged oligosaccharides like glucouronic acids. Although we still have to test this hypothesis with binding assays, it is very improbable, since this PUL (with a quasi-identical SusCD pair) has previously been shown in *Bacteroidetes vulgatus* to target linear XOS, but much less efficiently, or not at all, ramified oligosaccharides and xylans. Furthermore, we showed, with *E. coli* cells harboring the SusCD pair, that once assembled, the SusC and SusD components do not bind XOS. This indicates that no structural reorganization of SusD inducing binding occurred, once assembled to the SusC component. Rational engineering of the F5\_SusD will be required i) to definitively prove that the *B. vulgatus* SusD (with four different residues) does not bind XOS; ii) to identify the molecular determinants of the SusD/SusC interaction which induces the functionality of the transporter. Crystallization of the SusCD tandem could also be attempted to highlight putative structural rearrangements of SusD when bound to its cognate proteic ligand, in presence and in absence of XOS. Finally, the binding ability of the other SusD identified in this work (in clones 5 and 41O1), which are involved in oligosaccharide uptake, will have to be assessed. Using our strategy, we could indeed have selected several *Bacteroidetes* PULs preferentially targeting oligosaccharides rather than polysaccharides, to provide the native strain a competitive advantage to feed on the short dietary fibers which are naturally present in some dietary compounds, or on the leftovers of other microbes.

To sum up, in this PhD work we performed several technological improvements to unlock the identification of glycoside transporters and the characterization of their specificity towards oligosaccharides of defined structures. The analysis of the mechanism of some of them has

expanded our vision of how dietary fibers, in particular prebiotics, are selectively metabolized by uncultured and cultured species of the two dominant phyla in the human gut microbiome. Nevertheless, the work required to characterize as many glycoside transporters as CAZymes is huge. This domain can still be considered almost unexplored, given the great diversity of glycoside transporter sequences which is revealed in genomes and metagenomes.



## Résumé de la thèse

Les glycosides sont des molécules très abondantes et de structures très diverses. Ils sont présents dans toutes sortes d'organismes vivants dans lesquels ils assurent des rôles importants tels que le stockage d'énergie, la structuration cellulaire, la signalisation et la différenciation. Ils représentent le plus grand réservoir de carbone métaboliquement accessible, et donc, une source d'énergie importante pour la plupart des bactéries dans de nombreux environnements. En particulier, dans l'intestin des mammifères, les glycosides d'origines végétale, mammifère et bactérienne sont très abondants. Leur utilisation par les bactéries implique de multiples effets sur l'hôte influençant positivement le confort digestif, mais également la progression de diverses maladies intestinales chroniques, et potentiellement, d'infections bactériennes. En outre, le contrôle de l'utilisation des glycosides est important pour développer des usines cellulaires et des consortiums microbiens capables de convertir efficacement la biomasse non raffinée en produits de haute valeur ajoutée pour des applications biotechnologiques. Pour se nourrir de glycosides présentant des structures complexes, les bactéries ont développé des stratégies originales faisant intervenir des loci spécifiques de l'utilisation des polysaccharides (Polysaccharide Utilization Loci, PUL), qui codent à la fois les transporteurs et les enzymes actives sur les glucides (Carbohydrate Active Enzymes, CAZymes) qui sont nécessaires pour importer les substrats d'intérêt dans les cellules et les dégrader en monosaccharides. Le transport est donc une étape cruciale dans le métabolisme des glycosides, qui est lui-même un élément clé du fonctionnement et de l'équilibre du microbiote. Cependant, l'identification et la caractérisation des transporteurs se heurtent à plusieurs verrous technologiques. Malgré les avancées majeures réalisées pour la culture de certaines bactéries au cours de la dernière décennie, les difficultés de manipulation génétique, souvent associées à une redondance fonctionnelle, et le fait que les protéines transmembranaires ne peuvent pas être étudiées aussi facilement que les protéines solubles, rendent en effet difficile la caractérisation fonctionnelle des systèmes de transport. Cela est même impossible dans le cas des micro-organismes non cultivés, qui dominent les écosystèmes microbiens. Afin d'explorer directement la diversité taxonomique et fonctionnelle des microbes non cultivés, la métagénomique a été développée à la fin des années 1980. Selon la définition de Jo Handelsman, la métagénomique (également appelé génomique des communautés, génomique environnementale ou génomique des populations) est «l'analyse génomique d'une population de micro-organismes».

Au cours de la dernière décennie, une approche de métagénomique fonctionnelle à haut-débit, basée sur la recherche d'activité, a été développée au sein du groupe «DiscOmics» de l'équipe de Catalyse et d'Ingénierie Moléculaire Enzymatique de l'Institut de Biotechnologie de Toulouse (Toulouse Biotechnology Institute, TBI). Cette approche, appliquée à l'exploration fonctionnelle des microbiomes intestinaux humain et bovin, a permis d'identifier plusieurs centaines de clones d'*Escherichia coli* qui comprennent de grands fragments d'ADN métagénomique (30 à 40 kpb) contenant des clusters entiers de gènes bactériens de type opéron codant pour des CAZymes ciblant de façon synergique les glycosides complexes. Dans ces loci métagénomiques, les gènes codant pour les CAZymes sont très souvent regroupés avec ceux codant pour les transporteurs. En outre, il a été démontré que les transporteurs de glycosides de certaines bactéries spécialistes de l'utilisation de glycosides peuvent être produits sous forme fonctionnelle, à partir d'un PUL cloné dans un fosmide et exprimé dans *E. coli*.

Dans ce contexte, mon travail de thèse, initié en 2017, a été axé sur le développement de nouvelles méthodes très génériques pour permettre l'identification rapide de transporteurs et déterminer leurs spécificités. Nous avons appliqué ces techniques à la découverte de transporteurs d'oligosaccharides impliqués dans l'utilisation des fibres alimentaires par des bactéries non cultivées du tube digestif humain et du rumen bovin, et au déchiffrement de leurs mécanismes de liaison et de translocation. L'identification de transporteurs de glycosides à partir de bactéries non cultivées était l'un des principaux objectifs de ce projet. La première partie de ce travail de thèse a été consacrée à l'évaluation de l'efficacité et de la génécité de la sélection positive pour cribler les banques métagénomiques d'*E. coli* pour les transporteurs fonctionnels. Pour établir la preuve de concept de cette approche, nous avons examiné les séquences de 102 clones métagénomiques issus des microbiomes intestinaux humains et du rumen bovin, qui ont été précédemment identifiés lors du criblage pour des activités CAZymes de quatre banques métagénomiques totalisant 216 320 clones. Nous avons sélectionné 45 clones qui contiennent des loci métagénomiques codant pour au moins un transporteur de glycosides putatif et une CAZyme. Ensuite, 129 tests de croissance ont été effectués dans des microplaques en milieu minimal contenant des oligosaccharides qui pouvaient être hydrolysés par les CAZymes précédemment identifiées dans chaque clone. Enfin, 13 'hit' clones ont montré une capacité de croissance sur les FOS, les XOS, les acides aldouroniques, les cello-oligosaccharides, les  $\beta$ -glucooligosaccharides, le laminariose, les mannooligosaccharides, les galactomannoaligosaccharides et le lactulose, grâce à la co-

expression de gènes codant pour les transporteurs et les CAZymes. Ces tests ont fourni des preuves directes de la fonctionnalité des transporteurs de diverses familles, malgré leurs différences de spécificité d'origine taxonomique et de mécanisme. Grâce à ces résultats, nous avons mis en évidence une nouvelle fonction pour les transporteurs MFS qui est le transport de galacto-mannooligosaccharides. Nous avons également prouvé qu'il est possible de produire des transporteurs fonctionnels en exprimant les gènes des PULs provenant de bactéries Gram-positives dans la bactérie Gram-négative *Escherichia coli*, mettant en évidence la généralité de cette stratégie, bien que l'emplacement précis de chacun des domaines de transport chez *E. coli* reste encore indéterminé. Nous avons calculé que la probabilité de trouver des transporteurs de glycosides fonctionnels à partir de banques métagénomiques fosmidiques en utilisant l'hôte *E. coli* n'est que de 0,7 %, et la probabilité de produire à la fois une glycosidase fonctionnelle et un transporteur fonctionnel à partir d'un PUL (méta)génomique cloné dans un fosmide n'est que de 1,8%. Cela met en évidence l'importance du criblage automatisé pour identifier les transporteurs fonctionnels des métagénomiques, et plus généralement, des génomes bactériens, en utilisant le clonage de PULs dans les fosmidiques et l'expression dans *E. coli*. L'utilisation d'autres hôtes bactériens et d'autres vecteurs pour contrôler l'expression de chacun des gènes des PULs pourrait nous permettre d'augmenter ces rendements. Le débit et la sensibilité de cette méthode de criblage pourraient également être augmentés en utilisant la microfluidique, avec laquelle chaque essai pourrait être miniaturisé en gouttelettes d'un volume de l'ordre du pico-litre. Une telle miniaturisation de plusieurs ordres de grandeur permettra de cribler la métabolisation de substrats très coûteux, comme les glycosides humains par exemple. Enfin, en utilisant l'approche présentée ici, nous ne pouvons pas prouver définitivement qu'un transporteur est responsable de l'absorption d'un glycoside cible ni décrire son mécanisme. Cela nécessite en effet une troncature rationnelle du PUL, afin d'éliminer tout biais potentiel dû à la présence dans le fragment d'ADN métagénomique de plusieurs gènes codant pour des transporteurs. Parce que cela ne peut pas être fait à haut débit, nous avons dû sélectionner les cibles pour une caractérisation biochimique approfondie. Nous avons ainsi choisi certains des systèmes de transport de glycosides les plus originaux, en particulier ceux dont la fonction pourrait être importante pour la santé humaine.

C'est la raison pour laquelle, dans le cadre de cette thèse, nous nous sommes particulièrement intéressés à deux loci d'utilisation des fructooligosaccharides, identifiés dans les clones I9 et F1, tous deux issus de souches non cultivées de Clostridiales. En effet, les FOS sont de plus

en plus utilisés comme probiotiques dans l'alimentation humaine et l'alimentation du bétail. Les loci impliqués dans l'utilisation des FOS F1 et I9 présentent des synténies presque parfaites avec des parties des génomes de *Subdoligranulum* sp. APC924 / 74 et *Dorea longicatena* DSM 13814, respectivement. Plusieurs isolats d'espèces de *Subdoligranulum* sont des producteurs d'acide gras à chaîne courte et pourraient donc être considérés comme plutôt bénéfiques pour l'hôte humain. Au contraire, *Dorea longicatena* a été précédemment décrite comme non bénéfique, ce qui correspond au fait que nous avons trouvé que le locus I9 était surabondant et plus répandu dans le métagénome des patients souffrant de maladies inflammatoires de l'intestin. Le locus F1 ne contient qu'un seul transporteur de la famille PTS, mais qui n'a pas le domaine IIA par rapport au PTS du locus I9, qui inclue les trois domaines spécifiques aux glycosides (les domaines EIIB, C et A). Le locus I9 contient à la fois un transporteur PTS et un transporteur ABC. Après l'étape de criblage, il a donc été impossible de déterminer lequel des deux transporteurs était responsable de l'internalisation du substrat dans les tests de croissance. Premièrement, la troncature du PUL I9 a révélé que le transporteur ABC, pour lequel il manque la protéine de liaison au substrat, confère à *E. coli* une capacité faible mais significative d'utiliser le saccharose. Ensuite, en combinant l'ingénierie des locus, les analyses des dynamiques de croissance et la spectrométrie de masse, nous avons démontré que les deux PTS sont responsables de la liaison, du transport et de la phosphorylation du kestose. Les PTS des clones F1 et I9 sont les premiers PTS spécifiques de l'utilisation des FOS dont le mécanisme moléculaire précis et la spécificité exacte vis-à-vis des oligosaccharides de longueurs précises ont été déterminés. Pour ce faire, nous avons développé une méthode très générique pour quantifier les taux de liaison et d'absorption des transporteurs, basée sur l'utilisation de débris cellulaires bruts et de cellules intactes, et de substrats non marqués dont les variations de concentration sont très faibles (moins d'une nanomole) et ont été contrôlés par HPAEC-PAD. Cette méthode permet de contourner l'utilisation de substrats radiomarqués, dont l'utilisation nécessite des installations dédiées, de moins en moins courantes dans les laboratoires de biochimie et de microbiologie. De plus, des mutations spécifiques dans le génome d'*E. coli* ont révélé que les PTS recombinants des clones F1 et I9 sont fonctionnels dans *E. coli* grâce à la complémentarité de plusieurs composants non spécifiques et spécifiques des PTS de glucose et de fructose d'*E. coli*, soulignant la promiscuité des composants des PTS des bactéries Gram-positives et négatives. En outre, dans cette étude axée sur les PTS, nous avons utilisé un réseau de similarité de séquence pour analyser la diversité fonctionnelle, structurelle et taxonomique des transporteurs de PTS spécifiques de l'utilisation des glycosides. Nous avons revu la

classification de la TCDB et nous recommandons d'être prudent lors de son utilisation en ce qui concerne le niveau de classification des transporteurs. Les trois derniers composants TCID qui sont, dans la plupart des cas, liés aux substrats, ne doivent en effet pas être utilisés pour prédire la spécificité de substrat des PTS, et plus généralement des transporteurs de glycosides.

Une autre partie de ce projet de thèse a été consacrée à la caractérisation d'une protéine SusD d'un PUL spécifique de l'utilisation des xylooligosaccharides (XOS), identifiée dans le clone métagénomique F5. Ce PUL spécifique de l'utilisation des XOS, issue d'une souche intestinale *Bacteroides* non cultivée, et quasi-identique à celle de *Bacteroides vulgatus* ATCC 8482, a été précédemment caractérisée chez *E. coli* et dans la souche native par le groupe DiscOmics, en collaboration avec l'équipe d'Eric Martens (University of Michigan, USA). Ici, en utilisant une grande variété de méthodes, nous avons révélé que de manière très surprenante, la protéine F5\_SusD ne peut pas se lier aux XOS, qui sont pourtant efficacement internalisés par le système SusC/D. Ces méthodes comprennent celles habituellement utilisées pour caractériser les protéines de liaison aux glycanes, telles que la titration calorimétrique isotherme (Isothermal Titration Calorimetry, ITC) et l'électrophorèse sur gel d'affinité (Affinity Gel Electrophoresis, AGE). De plus, nous avons utilisé la méthode que j'ai développée dans ce travail de doctorat, basée sur la quantification par HPAEC-PAD des oligosaccharides non liés, et également la STD-RMN, dont nous avons prouvé pour les deux, être suffisamment sensible pour détecter une interaction entre la protéine et son ligand de l'ordre du micromolaire. Encore une fois, ces techniques sont génériques, élargissant le panel de techniques qui peuvent désormais être utilisées pour caractériser la spécificité des transporteurs de glycosides. Des expériences de complémentation ont également confirmé que la présence du F5\_SusD est essentielle pour l'absorption des XOS. De plus, la structure cristallographique du F5\_SusD a été résolue à une résolution de 2,6 Å sous forme *apo*, ce qui représente la première structure cristalline d'une protéine SusD impliquée dans un PUL spécifique de l'utilisation des XOS. Comparé à d'autres structures résolues de protéines de type SusD, le F5\_SusD a révélé une surface de liaison dépliée et l'absence ou l'orientation inappropriée de plusieurs résidus clés tels que les résidus de tryptophane connus pour être impliqués dans la liaison. Ces résultats mettent en évidence que la spécificité de liaison de SusD ne reflète pas nécessairement la spécificité de transport du transporteur. Ils constituent également la première preuve que les transporteurs SusCD peuvent absorber les oligosaccharides sans protéine de liaison au glycane fonctionnelle. Étant donné que la plupart des systèmes PUL qui ont été caractérisés jusqu'à présent ciblent des polysaccharides et non

des oligosaccharides, contrairement au PUL spécifique de l'utilisation des XOS caractérisés ici, nous ne savons toujours pas si les protéines de type SusD qui ne lient pas de substrat sont une caractéristique spécifique des transporteurs d'oligosaccharides SusC/D. Nous nous sommes également demandé si cette protéine pouvait se lier à d'autres oligosaccharides ou polysaccharides, d'autant plus que nous avons remarqué une surface chargée positivement qui pourrait, hypothétiquement, accueillir des charges négatives d'oligosaccharides comme les acides glucuroniques. Bien que nous devions encore tester cette hypothèse avec des tests appropriés, elle est très improbable, car ce PUL (avec une paire SusC/D quasi-identique) a déjà été mis en évidence chez *Bacteroidetes vulgatus* et il cible spécifiquement les XOS linéaires, mais beaucoup moins efficacement, ou pas du tout, les oligosaccharides ramifiés et les xylanes complexes. Une ingénierie rationnelle du F5\_SusD sera nécessaire pour d'une part prouver définitivement que la protéine SusD de chez *B. vulgatus* (qui diffère par quatre acides aminés au niveau de sa séquence) ne lie pas les XOS, et d'autre part, identifier les déterminants moléculaires de l'interaction SusD/SusC qui induit la fonctionnalité du transporteur. Enfin, la capacité de liaison des autres SusD identifiés dans ce travail (dans les clones 5 et 41O1), qui sont impliqués dans la capture des oligosaccharides, devra être évaluée.

Pour résumer, dans ce travail de thèse nous avons effectué plusieurs améliorations technologiques pour permettre d'accroître l'identification de transporteurs de glycosides et de perfectionner la caractérisation de leur spécificité vis-à-vis des oligosaccharides de structures définies. L'analyse du mécanisme de certains d'entre eux a élargi notre vision de la façon dont les fibres alimentaires, en particulier les prébiotiques, sont métabolisées de manière sélective par les espèces non cultivées et cultivées des deux phyla dominants dans le microbiome intestinal humain. Néanmoins, le travail requis pour caractériser autant de transporteurs de glycosides que de CAZymes est énorme. Ce domaine peut encore être considéré comme presque inexploré compte tenu de la grande diversité des séquences de transporteurs de glycosides révélées dans les génomes et métagénomes.







## Summary:

Glycoside metabolization is a crucial function for bacteria, in particular those residing in the gut of mammals. To feed on the dietary fibers of plant origin, they have developed very diverse proteic machineries to recognize, transport and degrade glycosides of complex structures. Despite the crucial role of transporters, their identification and characterization remain limited due to the fact that these membrane proteins are difficult to study in recombinant form, and that the majority of them belong to uncultured bacteria. In this thesis work, we screened a library of *Escherichia coli* fosmid clones, containing metagenomic loci involved in the metabolization of oligo- and polysaccharides, for transport functions. Thirteen metagenomic clones issued from the human and the bovine gut microbiomes were identified as being able to utilize oligosaccharides as sole carbon sources, thanks to the co-expression of genes encoding carbohydrate active enzymes and transporters from different families. Then, we developed a generic strategy to analyze, *in vitro* and *in cellulo*, the specificity of recombinant transporters in *Escherichia coli*, combining carbohydrate utilization locus and host genome engineering, screening and quantification of the binding, translocation, and growth rates, based on the use of unlabeled glycosides. Thanks to these technological developments, we characterized two loci identified after the screening step to be involved in the metabolization of fructooligosaccharides by uncultured Clostridiales, of which one constitutes a dietary and pathology biomarker in the human gut microbiome. We deciphered the mechanism of two phosphotransferase systems (PTS) of different structures, that we proved to be specific for kestose, and highlighted the promiscuity of the PTS components of Gram-positive and negative bacteria. Furthermore, integration of biochemical and (meta)genomic data allowed us to revisit the classification and taxonomical distribution of bacterial PTS and to better understand the interaction between prebiotics and human gut bacteria. Finally, the binding ability of a surface glycan binding protein, SusD, involved in xylooligosaccharide uptake by Bacteroidetes, was investigated with a large panel of techniques, and its crystallographic structure solved. These results demonstrate the efficiency of our strategy to rapidly identify glycoside transporters from various families and taxonomical origins, and to characterize the relationships between their structure and their binding and translocation specificities.

## Résumé

La métabolisation des oligosaccharides et polysaccharides est une fonction capitale pour les bactéries, en particulier celles des microbiotes digestifs des mammifères. Pour s'en nourrir, elles ont développé une machinerie protéique très diversifiée pour reconnaître, transporter et dégrader les sucres de structures complexes, notamment ceux composant les fibres alimentaires d'origine végétale. Malgré le rôle crucial des transporteurs, leur identification et leur caractérisation restent limitées du fait de plusieurs verrous techniques, liés au fait que ces protéines membranaires sont difficiles à étudier sous forme recombinante, et que la majorité d'entre elles appartiennent à des bactéries non-cultivées. Dans le cadre de cette thèse, nous avons criblé une banque de clones fosmidiques d'*Escherichia coli*, contenant des loci métagénomiques impliqués dans la métabolisation d'oligo- et polysaccharides, pour des fonctions de transports. Treize clones métagénomiques issus des microbiomes digestifs humain et bovin ont été identifiés comme pouvant utiliser des oligosaccharides comme seules sources de carbone, grâce à la co-expression de gènes codant pour des enzymes actives sur les hydrates de carbone et pour des transporteurs de différentes familles. Nous avons ensuite développé une stratégie générique pour analyser, *in vitro* et *in cellulo*, la spécificité de transporteurs recombinants chez *Escherichia coli*, combinant ingénierie de loci et du génome de l'hôte, criblage et quantification de l'efficacité de liaison, de translocation et de croissance, à partir d'oligosaccharides non marqués. Grâce à ces développements technologiques, nous avons caractérisé deux loci identifiés après l'étape de criblage comme étant impliqués dans la métabolisation des fructooligosaccharides par des Clostridiales non cultivées, dont l'un constitue un biomarqueur du régime alimentaire et de pathologies. Nous avons décrypté le mécanisme de deux systèmes phosphotransférases (PTS) de compositions différentes, spécifiques du kestose, et avons mis en évidence la promiscuité des constituants des PTS des bactéries Gram-positives et négatives. De plus, l'intégration de données biochimiques et (méta)génomiques nous a permis de revisiter la classification et la distribution taxonomique des PTS bactériens, et de mieux comprendre l'interaction entre les prébiotiques et les bactéries intestinales. Enfin, la capacité de liaison d'une protéine de liaison aux glycanes, SusD, impliquée dans le transport de xylooligosaccharides par des bactéries du phylum Bacteroidetes, a été étudiée avec un large panel de techniques, et sa structure cristallographique résolue. L'ensemble de ces résultats démontre l'efficacité de notre stratégie pour identifier rapidement des transporteurs d'oligosaccharides de familles et d'origines taxonomiques variées, et pour caractériser les relations entre leur structure et leurs spécificités de liaison et de translocation.