Functional and structural insights into Glycoside Hydrolase family 130 enzymes : implications in carbohydrate foraging by human gut bacteria

Simon Ladevèze

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)
Discipline ou spécialité : Ingénieries Moléculaires et Enzymatiques

---

Présentée et soutenue par Simon LADEVÈZE
Le mardi 28 avril 2015

Titre :
Functional and Structural insights into glycoside hydrolase family 130 enzymes: implications in carbohydrate foraging by human gut bacteria

---

JURY

Gideon John DAVIES, Professor (York University), rapporteur
Jean-Guy BERRIN, Chargé de Recherche INRA, rapporteur
Magali REMAUD-SIMÉON, Professeur INSA, présidente du jury
Bernard HENRISSAT, Directeur de Recherche CNRS, examinateur
Marion LECLERC, Chargée de Recherche INRA, examinatrice
Samuel TRANIER, Maître de Conférences (UPS), examinateur
Gabrielle POTOCKI-VÉRONÈSE, Directrice de Recherche INRA, directrice de thèse

---

Ecole doctorale : Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)
Unité de recherche : *LISBP, UMR CNRS 5504, UMR INRA 792, INSA Toulouse*
Directeur(s) de Thèse : *Gabrielle POTOCKI-VÉRONÈSE, Directrice de Recherche INRA*
Elisabeth LAVILLE, Chargée de Recherche INRA

THÈSE

En vue de l'obtention du

DOCTORAT

de

L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE TOULOUSE

Présentée et soutenue publiquement le mardi 28 avril 2015

Spécialité : Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries

Filière : Ingénieries Microbienne et Enzymatique

Par **Simon LADEVÈZE**

Functional and Structural insights into glycoside hydrolase family 130 enzymes: implications in carbohydrate foraging by human gut bacteria

Directrices de thèse:

Mme Gabrielle POTOCKI-VÉRONÈSE, Directrice de Recherche INRA
Mme Elisabeth LAVILLE, Chargée de Recherche INRA

NOM: LADEVÈZE **PRÉNOM**: Simon

TITRE: Apports Fonctionnels et Structuraux à la famille des glycoside hydrolase 130: implications dans la dégradation des glycanes par les bactéries de l'intestin humain.

SPÉCIALITÉ: Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries

FILIÈRE: Ingénieries Moléculaires et Enzymatiques

ANNÉE: 2015 **LIEU**: INSA, TOULOUSE

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

RÉSUMÉ:

Les relations entre bactéries intestinales, aliments et hôte jouent un rôle crucial dans le maintien de la santé humaine. La caractérisation fonctionnelle d'Uhgb_MP, une enzyme de la famille 130 des glycoside hydrolases découverte par métagénomique fonctionnelle, a révélé une nouvelle fonction de dégradation par phosphorolyse des polysaccharides de la paroi végétale et des glycanes de l'hôte tapissant l'épithélium intestinal. Les déterminants moléculaires de la spécificité d'Uhgb_MP vis-à-vis des mannosides ont été identifiés grâce à la résolution de sa structure cristallographique, sous forme *apo* et en complexe avec ses ligands. Un nouveau procédé de synthèse par phosphorolyse inverse d'oligosaccharides mannosylés à haute valeur ajoutée, a aussi été développé. Enfin, la caractérisation fonctionnelle de la protéine BACOVA_03624 issue de *Bacteroides ovatus* ATCC 8483, une bactérie intestinale hautement prévalente, a révélé que la famille GH130 comprend à la fois des glycoside-hydrolases et des glycoside-phosphorylases capables de dégrader les mannosides et les galactosides, et de les synthétiser par phosphorolyse inverse et/ou transglycosylation. L'ensemble de ces résultats, ainsi que l'identification d'inhibiteurs des enzymes de la famille GH130, ouvrent de nouvelles perspectives pour l'étude et le contrôle des interactions microbiote-hôte.

MOTS-CLES: Microbiote intestinal, Mannanes, N-Glycans, Glycoside-phosphorylases, GH130.

ECOLE DOCTORALE:
Sciences Ecologiques, Vétérinaires, Agronomiques, et Bioingénieries (SEVAB)

LABORATOIRE: Laboratoire d'Ingénierie des Systèmes Biologiques et Procédés (LISBP). UMR CNRS 5504, UMR INRA 792, INSA Toulouse.

SURNAME: LADEVÈZE **FIRST NAME**: Simon

TITLE: Functional and structural insights into glycoside hydrolase family 130 enzymes: implications in carbohydrate foraging by human gut bacteria.

SPECIALTY: Ecological, Veterinary, Agronomic Sciences and Bioingeneering

FIELD: Enzymatic and Microbial Engeneering

YEAR: 2015 **PLACE**: INSA, TOULOUSE

THESIS DIRECTORS: Mrs Gabrielle POTOCKI-VÉRONÈSE and Elisabeth LAVILLE

SUMMARY:

The interplay between gut bacteria, food and host play a key role in human health. The functional characterization of Uhgb_MP, an enzyme belonging to the family 130 of glycoside hydrolases, discovered by functional metagenomics, revealed novel functions of plant cell wall polysaccharide and host glycan degradation by phosphorolysis. The molecular determinants of Uhgb_MP specificity towards mannosides were identified by solving its crystal structure, in *apo* form and in complex with its ligands. A new process of high added value mannosylated oligosaccharide synthesis by reverse-phosphorolysis was also developed. Finally, the functional characterization of the BACOVA_03624 protein from *Bacteroides ovatus* ATCC 8483, a highly prevalent gut bacterium, revealed that GH130 family both contains glycoside phosphorylases and glycoside hydrolases, which are able to degrade mannosides and galactosides, and to synthesize them by reverse-phosphorolysis and/or transglycosylation. All these results, together with the identification of GH130 enzyme inhibitors, open new perspectives for studying, and potentially also for controlling, interactions between host and gut microbes.

KEYWORDS: Gut microbiota, Mannans, N-Glycans, Glycoside-phosphorylases, GH130.

DOCTORATE SCHOOL:
Ecological, Veterinary, Agronomic Sciences and Bioingeneering (SEVAB)

LABORATORY: Laboratory of Biosystems and Chemical Engeneering (LISBP). UMR CNRS 5504, UMR INRA 792, INSA Toulouse.

Remerciements

Ce manuscript vient conclure trois années et demie de thèse. Trois années et demie au cours desquelles beaucoup de travail a été fait. Pour autant, même si la bonne fortune à contribué à ce résultat, rien n'aurait été possible sans le concours de nombreuses personnes. Car une thèse, ce sont aussi des gens, des rencontres, une riche expérience de la vie. Que tous soient ici remerciés pour leur participation, matérielle, morale, de circonstance, ou plus régulière.

Je tiens en premier lieu à remercier les membres de mon jury, *Jean-Guy Berrin*, *Gideon Davies*, *Bernard Henrissat*, *Marion Leclerc*, *Magali Remaud-Siméon* et *Samuel Tranier* pour me faire l'honneur de juger mon travail.

Gabrielle, je souhaite te dire un immense merci. Immense, tout d'abord, pour avoir fait confiance, il y à près de quatre ans, à ce petit jeune, ce débarqué de la fac, cet inconnu au bataillon, à l'enthousiasme débordant et qui ne jurait à l'époque que par des cristaux en d'obscurs termes cristallographiques. Pour ton encadrement, ensuite, lors de ces années merveilleuses passées à travailler ensemble sur ces coquines enzymes, pour ta grande patience, mais aussi et surtout pour ta présence et ta grande honnêteté, aux antipodes de la langue de bois, qualités essentielles qui n'auraient pu me faire rêver de meilleures conditions pour l'accomplissement de cette thèse. Ton savoir, ta pertinence et ta rigueur ont été pour moi tout au long de ces années à la fois source d'admiration et d'inspiration. Pour la qualité de notre relation enfin, pour la joie partagée lors des bons moments, ainsi que pour le soutien lors des moins bons. Et finalement pour ton endurance, pour ce record inégalé d'avoir supporté deux thésards Gersois, venant du même petit village qui plus est…

Elisabeth, merci également pour ta présence tout au long de ces années, pour tes conseils, et pour ton aide dans les analyses génomiques et métagénomiques en particulier. Ta gentillesse et ton humour m'ont toujours ravi, tout comme tes expressions cultes qui sont venues enrichir mon répertoire.

Magali, merci de m'avoir si bien accueilli au sein de cette équipe, vivante, dynamique et chaleureuse. Ta pertinence, ta rigueur, ainsi que l'honnêteté et l'humanisme dont tu fais preuve afin de la faire vivre ont toujours ont toujours suscité en moi le respect.

Je tiens ensuite à remercier l'ensemble de l'équipe Catalyse et Ingénierie Moléculaires et Enzymatiques, pour le dynamisme qui y règne, pour cette ambiance, décontractée et libre qui fut un excellent facteur d'intégration dès mon arrivée. Beaucoup de monde est présent dans cette équipe, ce qui participe à sa vivacité ; néanmoins il est fort dommage que la situation actuelle du monde nous interdise de partager davantage de moments ensembles. Tout d'abord merci à mes collègues de bureau, à ma voisine *Marion* qui a supporté mes bavardages, même s'il est vrai que mon silence peut s'acheter avec du

chocolat. Les grands délires que nous avons eus avec *Yannick*, *Marlène*, *Etienne* et *Florent* vont beaucoup me manquer, tout comme les « craquages » systématiques du vendredi après-midi, ou un peu plus tôt. *Yannick*, éternel fêtard et mélomane, ton soutien lors des soirées de manip fut une aide précieuse, occasion de riches découvertes musicales et d'intenses discussions philosophico-politiques. *Thomas*, heureusement que tu étais là afin de parfois arriver à trouver la motivation pour aller transpirer à Movida. Merci beaucoup à *Morgane* pour tes éclats de rire, et merci *Marina*, pour tous ces moments que nous avons partagés. Muchas gracias a *Alvaro*, a los *Pablos* y a *Agustina*, con quién pasé muchos buenos momentos, fue para mí una única oportunidad de practicar el idioma de Cervantes. Sin embargo, todavía no he decidido si prefiero el Real o el Barça, sobre todo cuando estamos probando vinos franceses. No olvidaré tampoco este concierto de massive attack en el castillo de Carcassona o la música trance del bikini. *Alvaro*, mi amigo, es así que entiendo la Unión Europea. Enfin merci à *Marc* pour ta gentillesse et ton humour, à *Florence* pour ta bonté, ainsi qu'à la douce *Emilie* qui cuisine divinement bien.

Merci *Laurence* et merci *Angeline*, qui m'avez assisté à la paillasse et démultiplié mon travail, sans vous un tel projet n'aurait pu aboutir à une si belle conclusion et n'aurait jamais amassé des résultats dans de telles proportions.

Merci à *Nelly* et *Laurence* pour leur aide sur le plateau HPLC, vous m'avez tiré de pas mal de galères et m'avez tiré d'un grand nombre de prises de tête avec chroméléon.

Gianluca, grazie mille per tutto, pour le soutien et l'assistance que tu m'as apporté depuis ton arrivée dans l'équipe, à la fois sur le plateau Akta, mais aussi lors de la grosse prise de tête sur le SAXS et l'immonde assignation des quelques milliers de molécules d'eau. Merci aussi pour ta gentillesse, ton humour et ta disponibilité ainsi que pour les riches échanges que nous avons eu. Je suis sur que lorsque ce magnifique accent italien s'accommodera avec le Tolosenc, une petite Pietra en ta compagnie aura un goût unique…

Je tiens aussi à remercier l'ensemble des membres du Groupe de Biophysique Structurale de l'IPBS, ma deuxième maison, où naquit de façon concrète mon profond attachement à ce métier, à son idéal, à ces gens qui jour après jour, bon an, mal an, y mettent leurs tripes et tout leur cœur.

Samuel, merci pour ton enseignement, tout d'abord, comme professeur, sur les bancs des amphis de Paul Sab', tu m'as permis de découvrir le monde merveilleux des cristaux et de la Biochimie. Merci ensuite pour tes conseils et ton assistance tout au long de cette thèse, sans toi cet hexamère ne se serait jamais révélé. Merci enfin à titre personnel, pour ta gentillesse ainsi que pour les nombreux et agréables moments de partage que nous avons eu, autour d'un café, le temps de patienter pendant les cycles d'affinement, et en dehors. Je ne doute pas que la vue régulière de paniers de « produits de nos régions » ait un effet sur tes exotiques poissons et qu'ils prennent un jour un peu l'accent Aveyronnais.

Merci ensuite à *Lionel*, pour ta gentillesse, pour m'avoir accueilli au sein de l'équipe, pour m'avoir permis d'apprécier les shifts de nuit sur les lignes de lumière, et les découvertes musicales que tu m'a permis de faire. *Valérie*, merci pour la confiance que tu m'as accordée lors du mon stage de master et pour ton précieux enseignement sur l'art complexe de la cristallisation des FAD32, pour ton humour, ton sourire rayonnant, et pour la fabuleuse recette des greenies… Merci ensuite pêle-mêle à *Laurent*, *Jean-Denis*, *Sabine*, *Cécile*, *Sylvianne*, *Eric*, *Suzana*, *Virginie*, *Monique* et les anciens : *Fabien* (encore un Gersois, décidément…), *Karine*, *Fred*, *Romain*, *Yohann*, *Luis*, *Alex*, et la p'tite *Cécile*, au grand cœur et aux rires si communicatifs, pour tous les bons moments que nous avons partagés ensemble, autour d'un nombre astronomique de gâteaux, galettes, crêpes, barbecues et autres découvertes culinaires et œnologiques en tous genres, telles les spécialités houblonnées Alsaciennes et Belfortaines. Enfin, merci à *Ania* et à la p'tite murène *Sandrine*.

Un très grand merci à *Bernard Henrissat* pour ton accueil à l'AFMB, ainsi qu'à toute la CAZy team, en particulier à *Vincent*, *Pedro* et *Elodie*. Merci de m'avoir fait profiter de ton immense savoir et pour m'avoir fait découvrir les entrailles de CAZy. Merci aussi pour ta gentillesse et pour m'avoir fait découvrir ce petit resto si sympathique aux abords de la cité phocéenne. Le soleil et la mer sont si agréables sous les pins et le regard bienveillant de la *Bonne mère*.

Je voudrais remercier également *Sébastien Gouin* et *Yoan Brissonnet* et le groupe des Nantais pour leur excellent travail sur la synthèse de ces iminosucres au comportement si particulier, ainsi que *Pascale Mosoni*, *Jordane Despres* et les Clermontois pour leur maîtrise de la culture des bactéries anaerobies, ainsi que leur excellent travail de transcriptomique.

Merci encore à *Pierre Roblin* (un ancien de l'IPBS, comme le monde est petit…), pour son assistance sur SWING, pour l'acquisition des données SAXS et leur traitement.

Je tiens aussi à remercier l'ensemble du service pédagogique de biochimie de l'université, à commencer par *Anne* et *Karine*, qui m'ont permis d'exercer le difficile mais non moins stimulant travail de transmission du savoir auprès des futures recrues de nos équipes dans des conditions optimales. La chaleureuse ambiance du service, à la fois drôle et stimulante, tant par les nombreux débats suscités par les incohérences et lourdeurs du système universitaire que par les « TP raclette », option singulière affectionnées par de nombreux anciens étudiants passés côté craie, participe de sa vivacité. Merci encore à *Samuel* pour tes conseils, ainsi qu'à *Laurent*, *Eric*, *Fabienne*, *Suzana*, *Cécile*, *Anne*, *Maryelle*, *Chantal*, *Marie-Pierre*, *Manue*, *Alain*, *Catherine*, *Rémi* et tous les autres pour tous les échanges que nous avons eu. Merci à *Jean-Charles* et *Fabien*, qui après m'avoir enseigné le métabolisme, m'ont associé à sa diffusion auprès des étudiants. Merci à tous les membres de leur équipe, *Stéphanie*, *Lindsay*, *Lara*, *Sergeï*, *Pierre* et *Edern* pour ne citer qu'eux, pour leur gentillesse, et chez qui j'ai mis les pieds pour la toute première fois dans un laboratoire.

Dans cette même ligne, merci au BTS du lycée Salièges, de m'avoir permis d'encadrer deux stagiaires au cours de cette thèse, *Kévin* et *Brenda*. Ils m'ont permis de développer une autre approche dans la transmission du savoir, plus directe et plus concrète.

Puis je tiens à remercier mes amis, pour m'accompagner depuis toutes ces années. Amis d'enfance, amis Gersois, planant, des pistes de ski au fond des mers, *Arnaud* et *Cindy*, *David*, *Thomas*, *Romain* et le gang des *Célines*, le temps passe mais l'affection que j'ai pour vous reste intacte, et vous revoir m'est toujours aussi agréable. Amis de fac, amis de bringue, amis des quatre coins de France, *Quentin* et *Gaëlle*, *Joseph* et *Hélène*, *Charlène* et *Lou*, *Célia*, *Pierre*, la distance nous contraint à nous voir moins souvent, mais c'est avec toujours autant de plaisir, même au tour d'un simpe verre. Mais, j'ai ouï dire qu'un joyeux évènement devrait nous réunir cet été pour à nouveau faire la fête comme avant... *Greg*, mon camarade route au sein du monde biochimique depuis toutes ces années, comme il est bon de se prendre un petit café en refaisant le monde, ou de goûter toutes ces bières en planifiant le prochain ciné ou en parlant de choses plus légères. Je serai là moi aussi lorsque ce sera ton tour de passer des longues heures à rédiger. Dubliners, Mulligan's, London Town, Ancienne Belgique, Carbet d'Oc, La Tireuse, Connexion Café, La Maison, The Frog & Rosbif, Café-théatre les 3T, Le Bikini, Les Cinémas Gaumont, vous nous avez sauvé la vie. Pour finir, mention spéciale pour celle qui a toujours été là, ma très chère *Stéphanie*, un très grand merci pour tout. Pendant toutes ces années, tu as été de toutes les histoires, de toutes les soirées, de tous les bons moments et aussi de tous ceux qui l'étaient moins. Tu me connais et me comprends comme nulle autre, et les difficultés réciproques qui nous ont tant rapprochés font que la relation si particulière qui nous unit m'est très précieuse.

Enfin je tiens à remercier ma famille, et en premier lieu mes parents *Angeline* et *Jacques*, pour leur affection, permanente, depuis toujours ; pour leur soutien, de tous les instants, ainsi que pour les efforts qu'ils ont tant développé afin comprendre ce qu'il se passait dans toutes ces pipettes. Merci à mon frère, *Nicolas*, pour être là, et pour les efforts démesurés qu'il déploie afin de comprendre la subtile nuance épistémologique qui règne entre sciences exactes et sciences expérimentales. Ma cousine et mes cousins, ensuite, du plus jeune au plus grand, pour la joie qu'ils m'apportent, et pour tous les moments de rire que nous passons ensemble. Mes oncle et tantes, pour tous les bons moments, pour tous les supers-apéro-tariquet-tapas-foie-gras-grillades-café et les grands délires familiaux en général. Pour la grande connaissance que nous développons, année après année, avec assiduité, dans le domaine fort complexe de l'œnologie, car il est bien connu qu'*in vino veritas*. Mes grands parents enfin, pour être là, et être si uniques et irremplaçables. Vous voir est toujours pour moi source de bonheur, de fierté et de joie.

Je souhaiterais pour finir dédier ce travail à ma grand-mère, qui doit sans nul doute apprécier le spectacle du haut d'un petit nuage.

«Nous sommes comme des nains juchés sur les épaules de géants, de sorte que nous pouvons voir plus de choses qu'eux, et des choses plus éloignées qu'ils ne le pouvaient, non pas que nous jouissions d'une acuité particulière, ou par notre propre taille, mais parce que nous sommes portés vers le haut et élevés par eux à une taille gigantesque.»

Jean de Salisbury, *Metalogicon* (III, 4), 1159.

Table of Contents

List of figures and tables

Abbreviations list

GIT:	Gastro-Intestinal Tract
NGS:	Next Generation Sequencing
SFCA:	Short Chain Fatty Acids
kDa:	Kilo Dalton
MDa:	Mega Dalton
CAZyme:	Carbohydrate Active enZyme
LPS:	LipoPolySaccharide
IBD:	Inflammatory Bowel Diseases
IBS:	Irritable Bowel Syndrome
CD:	Crohn's Disease
UC:	Ulcerative Colitis
AIEC:	Adherent-Invasive *E. coli*
Kbp:	Kilo base pairs
Gbp:	Giga base pairs
Sus:	Starch utilization system
PUL:	Polysaccharide Utilization Locus
PDB:	Protein Data Bank
BT:	*Bacteroides thetaiotaomicron*
DP:	Degree of Polymerization
GH:	Glycoside hydrolase

GT: GlycosylTransferase

GP: Glycoside phosphorylase

HMNG: High Mannose *N*-Glycan

CNG: Comlex *N*-Glycan

HNG: Hybrid *N*-Glycan

PLM: PhosphoLipoMannan

GalXM: GalactoXyloMannan

GXM: GlucoXyloMannan

LPG: LipoPhosphoGlycan

P_i: Inorganic Phosphate

CBM: Carbohydrate Binding Module

Introduction générale

Le microbiote intestinal assure de nombreuses fonctions (protection contre les bactéries pathogènes, immuno-modulation, régulation de l'inflammation de la muqueuse intestinale, métabolisation des constituants alimentaires…) qui jouent un rôle fondamental dans le maintien de la santé humaine. Pour assurer leur croissance dans cet écosystème complexe, les bactéries intestinales produisent une impressionnante machinerie d'enzymes capables de dégrader les glucides de structures complexes, les glycanes, qui constituent leur principale source de carbone. Les glycanes sont issus principalement des fibres provenant des fruits, légumes et céréales composant notre régime alimentaire, mais aussi des glycoprotéines humaines, qui tapissent l'épithélium intestinal, ou microbiennes. Les enzymes bactériennes capables de les dégrader en sucres simples, principaux acteurs du métabolisme des glucides, sont donc cruciales pour l'équilibre du microbiote intestinal humain. De plus, la diversité et l'efficacité naturelle de ces biocatalyseurs peuvent être exploitées *in vitro* pour développer des procédés propres de conversion de la biomasse végétale en synthons, nouveaux matériaux non persistants, et autres molécules à haute valeur ajoutée pour les biotechnologies appliquées aux secteurs de l'alimentaire, de la santé et de l'environnement. Cependant, l'exploitation de cet extraordinaire réservoir d'enzymes se heurte à une limite majeure. En effet, seulement 20 à 30 % des espèces microbiennes composant le microbiote intestinal humain ont été cultivées à ce jour. Dans ce contexte, le potentiel de la metagénomique est immense. Cette technologie, aussi appelée « génomique des communautés », est l'analyse des génomes de tous les organismes d'une niche écologique donnée, sans étape de culture.

Récemment, le groupe 'DiscOmics' de l'équipe de Catalyse et Ingenierie Moléculaire Enzymatiques du LISBP, a développé une approche d'exploration à haut-débit de métagenomes, basée sur le criblage multi-étapes de banques metagénomiques. Cette technologie permet d'explorer à une cadence de 100 000 tests par jour le potentiel fonctionnel codé par plusieurs Gigabases d'ADN microbien. Elle permet de construire des collections de clones « hits » produisant les activités d'intérêt, et de focaliser les efforts de séquençage sur les seuls fragments d'ADN codant les enzymes cibles.

Cette approche, appliquée à l'exploration fonctionnelle du metagénome intestinal humain, a permis d'isoler plusieurs centaines de clones particulièrement efficaces pour la dégradation des glucides complexes. Grâce au séquençage en profondeur de moins de 2 Megabases d'ADN metagénomique (l'équivalent d'un génome bactérien), plus d'une centaine de nouvelles enzymes actives sur les hydrates de carbone ont été découvertes, dont certaines appartiennent à des familles qui n'ont jamais été caractérisées sur le plan fonctionnel ni structural. Cette étude a aussi permis de mettre en évidence les mécanismes

d'adaptation des bactéries intestinales à la dégradation de substrats complexes, par le biais de fréquents transferts horizontaux de gènes codants des systèmes muti-enzymatiques.

Une telle stratégie, bien qu'employée à un débit plus restreint, est couramment utilisée à travers le monde afin d'identifier de nouvelles activités biologiques d'intérêt au sein des écosystèmes. Son efficacité est directement dépendante des capacités de criblage et d'analyse déployées au sein des centres de recherche.

Néanmoins, la caractérisation des relations entre structure et fonction des nouvelles enzymes découvertes par le biais de ces approches haut-débit reste l'étape limitante de l'ensemble du processus. Cette situation est illustrée par l'écart sans cesse grandissant entre le nombre de séquences génomiques et metagénomiques déposées dans les bases de données et le nombre d'enzymes caractérisées, tant sur le plan fonctionnel que structural.

Pourtant, la caractérisation biochimique fine de ces biocatalyseurs constitue une étape clef pour étendre le catalogue de fonctions catalytiques disponibles, et ainsi évaluer leur potentiel applicatif pour les biotechnologies industrielles. Intégrée aux données génomiques et metagénomiques, elle est aussi fondamentale pour mieux comprendre, voire contrôler, le fonctionnement des écosystèmes microbiens.

L'acquisition de données structurales, quant à elle, est indispensable pour comprendre la catalyse enzymatique dans ce qu'elle a de plus intime. C'est aussi la base de l'ingénierie du vivant, en commençant par ses éléments les plus basiques, les enzymes, pour aller jusqu'à la cellule entière. Ciblés en particulier sur les enzymes actives sur les sucres, ces travaux permettent, *in fine*, de réinventer les modes de production énergétique dans un contexte de raréfaction des ressources fossiles, mais aussi de convertir la biomasse végétale en nouveaux matériaux biosourcés et molécules glucidiques à haute valeur ajoutée pour les secteurs de l'alimentation et de la santé.

C'est dans ce contexte qu'a été initiée ma thèse, qui a porté sur la caractérisation fonctionnelle et structurale d'une des cibles les plus prometteuses identifiée par metagénomique fonctionnelle du microbiote intestinal humain. Il s'agit d'une enzyme que nous appellerons Uhgb_MP (Unknown human gut bacterium mannoside-phosphorylase), en lien avec les fonctions que nous avons découvertes dans le cadre de cette thèse. Uhgb_MP est codée par un gène très prévalent du metagénome intestinal, suggérant qu'elle jouerait un rôle clef pour le métabolisme des glucides dans l'écosystème digestif. Cette cible, qui n'appartenait à aucune famille connue jusqu'en 2011, a été clonée et produite sous forme recombinante chez *E. coli* avant le début de mon projet, grâce à Juliette Bercovici et Davide Cecchini, qui ont réalisé les travaux préliminaires de caractérisation fonctionelle. Uhgb_MP a

par la suite été classée dans la famille GH130 de la classification CAZy durant la première année de ma thèse.

Ce manuscrit est organisé en sept chapitres, les trois premiers constituant l'introduction bibliographique:

Le **chapitre un** est centré sur le microbiote intestinal humain, sa description, son implication dans certaines pathologies humaines, et sur le rôle primordial joué par les bactéries du genre *Bacteroides* en son sein.

Le **chapitre deux** est focalisé sur les enzymes impliquées dans la dégradation des glycosides, et décrit leur système de classification. C'est dans ce chapitre que nous porterons un regard détaillé sur l'état de l'art concernant le groupe spécifique des glycoside-phosphorylases, notamment celles appartenant à la famille GH130.

Le **chapitre trois** constitue une revue bibliographique destinée à la publication, décrivant les mécanismes de dégradation des mannosides par les bactéries. Il fournit une description de leur diversité structurale, souligne les interactions cellulaires qu'ils peuvent médier, ainsi que les stratégies développées par les bactéries pour les métaboliser.

Les quatre chapitres suivants sont relatifs aux résultats obtenus lors de cette thèse, lesquels sont présentés sous la forme de quatre articles:

Le **chapitre quatre** est un article publié en 2013 dans la revue *Journal of Biological Chemistry*, traitant de la caractérisation biochimique de la mannoside phosphorylase Uhgb_MP. Dans cet article nous révélons la spécificité fonctionnelle de cette enzyme et son mécanisme cinétique, et présentons une analyse intégrant données biochimiques, génomiques et metagénomiques à l'échelle de l'écosystème intestinal humain, permettant de discuter le rôle *in vivo* des enzymes de la famille GH130, structurée en deux sous-familles. Ce chapitre fait aussi le point sur le potentiel d'Uhgb_MP pour la synthèse de mannosides à haute valeur ajoutée, application brevetée en 2013 (Annexe).

Le **chapitre cinq** est un article publié en 2015 dans la revue *Acta Crystallographia section D*, portant sur l'analyse structurale d'Uhgb_MP. Il décrit la structure cristallographique de cette enzyme sous sa forme *apo* ainsi qu'en complexe avec ses substrats, ce qui permet de valider son mécanisme catalytique. Nous présentons également dans cet article une analyse structurale de la famille des glycoside hydrolases 130, en établissant les relations structure-activité de ces enzymes, au regard des données fonctionnelles et structurales disponibles des différent représentants de chacune des sous-familles.

Le **chapitre six** est un article publié en 2015 dans la revue *Bioconjugate Chemistry* décrivant la capacité d'imminosucres multivalents à inhiber des enzymes actives sur les sucres (dont Uhgb_MP) probablement impliquées dans plusieurs pathologies humaines. Il met en lumière un effet jamais décrit auparavant pour cette famille de molécules, qui peuvent se comporter à la fois comme inhibiteurs ou comme activateurs de glycosidases et glycoside-phosphorylases. Enfin, il apporte des données biophysiques permettant de suggérer le mode d'action de ces molécules.

Le **chapitre sept** est un article en préparation portant sur la caractérisation fonctionnelle de la protéine BACOVA_03624, un autre membre de la famille GH130 provenant d'une bactérie intestinale très prévalente, *Bacteroides ovatus* ATCC 8483. Dans ce chapitre, nous démontrons pour la première fois que certaines GH130 sont capables d'agir à la fois sur les galactosides et mannosides, par hydrolyse, phosphorolyse, transglycosylation et phosphorolyse-inverse. Ce chapitre fournit également une analyse transcriptomique préliminaire de *B. ovatus* ATCC 8483, visant à démontrer l'implication des enzymes de la famille GH130 dans la dégradation des β-mannanes et N-glycanes eucaryotes.

Un résumé général conclut ce manuscrit, en soulignant les avancées majeures réalisées au cours de ce travail de thèse, et propose des perspectives quant à la caractérisation des enzymes de la famille GH130 et de leurs applications.

Scientific production

Research papers

- Ladevèze S., Tarquis L., Cecchini D.A., Bercovici J., André I., Topham C.M., Morel S., Laville E., Monsan P., Lombard V., Henrissat B., and Potocki-Véronèse G. 2013. Role of glycoside-phopshorylases in mannose foraging by human gut bacteria. J. Biol. Chem., 288: 32370-32383. doi: 10.1074/jbc.M113.483628.

-Ladevèze, S., Cioci, G., Roblin, P., Mourey, L., Tranier, S. & Potocki-Véronèse, G. Acta Crystallogr D Biol Crystallogr. 2015 Jun;71(Pt 6):1335-46, doi:10.1107/S1399004715006604.

- Brissonnet, Yoan; ladevèze, Simon; teze, David; Fabre, Emeline; Deniaud, David; Daligault, Franck; Tellier, Charles; Sestak, Sergej; Remaud-Siméon, Magali; Potocki-Véronèse, Gabrielle; Gouin, Sébastien. Polymeric iminosugars improve the activity of carbohydrate-processing enzymes. Bioconjug Chem. 2015 Apr 15;26(4):766-72. DOI: 10.1021/acs.bioconjchem.5b00081.

- Simon Ladevèze, Angeline Rizzo, Elisabeth Laville, Jordane Despres, Pascale Mosoni, Sandrine Morel and Gabrielle Potocki-Véronèse. The human gut symbiont Bacteroides ovatus possesses a highly polyvalent GH130 enzyme acting on galactosides and mannosides. In preparation.

Review paper

- Simon Ladevèze and Gabrielle Potocki-Véronèse. Mannoside recognition and degradation by bacteria. To be submitted.

Patent

Ladevèze S., Tarquis L., Henrissat B., Monsan P., Laville E., Potocki-Véronèse G. Use of specific glycoside phosphorylases for the implementation of phosphorolysis or reverse phosphorolysis reactions. 2013. European patent deposited by Toulouse Tech Transfert (N° EP 13306108.5). International extension in 2014 (N° PCT/EP2014/066565).

Oral Communications

- Ladevèze S., Cecchini D., Laville E., Tarquis L., Robe P., Henrissat B., Leclerc M., Doré J., Remaud-Siméon M., Monsan P., and Potocki-Véronèse G. Functional metagenomics reveals

novel pathways of carbohydrate foraging by human gut bacteria, JIB 2013, Hammamet, Tunisia, December 20-24 2013. (Keynote Lecture).

- Ladevèze S., Laville E., Cecchini D., Tarquis L., Remaud-Siméon M., Potocki-Véronèse G. Functional metagenomics reveals novel pathways of carbohydrate foraging by human gut bacteria. Gut microbiology: from sequence to function. June 16-19, 2014, Aberdeen, Scotland.

- Ladevèze S., Laville E., Tarquis L., Tranier S., Potocki-Véronèse G. Functional metagenomics reveals novel pathways of carbohydrate foraging by human gut bacteria. Biocat congress, September 2014, Hamburg, Germany.

- Ladevèze S., Cecchini D., Laville E., Robe P., Henrissat B.,Leclerc M., Doré J., Remaud-Siméon M., Monsan P., Potocki-Véronèse G. Functional metagenomics to boost enzyme discovery. Enzynov, October 2014, Romanville (Plenary invited conference).

- Ladevèze S., Laville E., Cecchini D., Tarquis L. Potocki-Véronèse G. Functional metagenomics reveals novel pathways of carbohydrate foraging by human gut bacteria, GFG 2014, May 12-15, 2014, Paris, France (Plenary invited conference).

- Ladevèze S., Laville E., Tarquis L., Tranier S., Potocki-Véronèse G. Functional metagenomics reveals novel pathways of carbohydrate foraging by human gut bacteria, Colloque CBSO, 3-6 june 2014, Carry-le-Rouet, France.

Introduction

1. The human gut microbiota

The human body is colonized by a huge number of microorganisms, collectively termed as the human microbiota. From a certain point of view, one could say that human beings are more microbial than human. Indeed, being composed of around a hundred trillion (10^{14}) microbial cells, a human body is made of ten times more bacterial cells than human ones. Even if these microorganisms only represent a reduced part of the total body weight (around 1.5 kilograms of alive bacteria), they are found in all parts of the human body (Xu and Gordon 2003). From the surface and deep layers of the skin, to the saliva and all mucus (oral, conjunctiva, lungs or genitourinary), the human body is actually forming in itself a dense and complex ecosystem. But including 10^{11} microorganisms per gram of wet weight of feces, the largest part of the human microbiota is certainly located in the gastro-intestinal tract (GIT) (Wexler 2007). Recent studies have shown that this very high number of cells only accounts for a thousand species, but whose genomes, put together, largely exceeds the human genome by two orders of magnitude, making us, genetically speaking, only one percent human (Qin et al. 2010). As a permanent companion from birth to death, the microbiota plays a key role in a large set of physiological functions in the human body, and has a tremendous impact on our lives, being now considered as an additional organ. Approaching and understanding *Homo sapiens* only on the basis of its own eukaryotic genome, as it has been during the genomic era, is no longer sufficient. It is now necessary to consider human as a supra-organism, because we need to take into account this large amount of DNA information, referred to as the human gut metagenome, and all the interactions and co-evolution occurring between the human host and its microbial guests.

1.1. The human gut microbiota: environment and composition

The GIT is a heterogeneous environment which can be divided into several distinct regions showing important differences due to their physiological functions. Along with the human evolution, the miscalled “gut flora”, relic of a static Linnaeus *Scala naturæ* view of life, has also evolved to adapt to these particular environmental constraints, and some species developed specific features, to allow them to colonize efficiently these environments. One of the most important differences between the different areas in the GIT is pH, which progressively increases along the GIT, starting with a very acidic pH of 1 to 2 in the stomach, to a value comprised between 5 to 7 in the intestine (Prakash et al. 2011). After the stomach, the GIT is composed of two distinct elements, the small intestine (duodenum, jejunum and ileum) where most of the host intake process takes part, and the colon (or large intestine) (fig. 1a). Together with this progressive pH increase, the number of micro-organisms increases along the GIT, starting with 10^1 to 10^2 cells per gram of wet weight of stomach content to no less than 10^{12} cells per gram of intestinal content in the

colon (fig. 1a). Different elements may explain such a high longitudinal variation in the amount of micro-organisms. First, the very acidic pH of the stomach inhibits the growing of most of micro-organisms, and only very specialized and extremophile ones can grow there. In the duodenum, bile and pancreatic juice which contains digestive enzymes (mainly proteases, but also pancreatic lipases and amylases), form a harsh environment for micro-organisms, preventing their growth. In the small intestine, even if pH is much more compatible with bacterial growth (between 6 and 7), the host digestive process involves mechanical movements called 'peristalsis', which are periodic contractions of smooth muscles surrounding the gut, which negatively impact micro-organism growth (Dukowicz et al. 2007). Finally, the best conditions for microbial growth are found in the distal part of the GIT, the colon. There, the pH is suitable for life, and there are no digestive enzymes which would inhibit bacterial growth. In addition, this organ is dedicated to water absorption, and to nutrient absorption. For that purpose, peristaltic movements are much slower, allowing optimal microbial growth, mainly by fermenting dietary constituents that were not digested in the upper parts of the GIT.

Most of the gut microbiota is composed of strict anaerobes, as their environment is not in contact with the exterior medium. However, presence in lower amounts from two to three orders of magnitude of facultative aerobes and aerobes has been reported (Sekirov et al. 2010). Historically, advances in microbial composition studies can be divided into two majors eras, related to technological developments. Culture based and microscopic descriptions were the first methods used to describe microbial content in ecosystems. After 1995, development of next generation sequencing (NGS) technologies allowed to enter the metagenomic era, providing new data and expanding our knowledge in microbial diversity. Even if most of changes in bacterial diversity and amounts are correlated to diet, pH, local oxygenation levels and balance with the immune system, microbial diversity and amounts can be described according to three main parameters. This three dimensional diversity can be described first longitudinally, as the stomach mainly contains *Helicobacter*, *Streptoccoci* or *Lactobacilli* genus, the small intestine mainly contains *Streptococaceae*, Actinobacteria (high G+C organisms), while *Firmicutes* and *Bacteroidetes* are forming the main part of the colon microbiota (fig. 1a) (Lagier et al. 2012; Zoetendal et al. 2012). Secondly, we can note that diversity and specialization are found transversally, as oxygenation, pH and kind of nutrient are varying from the lumen to the epithelium. Phyla such as *Clostridium*, *Lactobacillus* and *Enterococcus* are the most common ones found in the mucus layer and epithelial crypts in the small intestine, while *Bacteroides* or *Firmicutes* are mainly found in feces (Lagier et al. 2012) (fig. 1b). Finally, temporal differences are found from birth to death with changes correlated to origin, environment, nutrition habits, prebiotics and antibiotics consumption or diseases (the last one will be further detailed). Initial establishment of gut microbiota begins at birth and its composition is first affected by the way babies are delivered. Babies born by cesarean show different microbial composition compared to those

born by vaginal delivery, the later showing similarities with the mother's vaginal composition (Donovan et al. 2012). During the first year of life, microbial composition and diversity are rapidly increasing, only stabilizing after teenage, the gut microbiota being so far not similar to the ones of adults (Agans et al. 2011) (fig. 1c). Contradictory results tend to show that gut composition may differ depending on the origin of the individuals (Lay et al. 2005; Grześkowiak et al. 2012), but these differences seems to be mostly artefactual, and primarily due to the wealth of the studied population, largely impacting diet habits which seems to be the most impacting factor. In the same way, prebiotics, a category of functional foods composed by oligosaccharidic and polysaccharidic fibers, can affect bacterial composition, as they feed specific bacterial genus (Quigley 2010). Surprisingly, probiotics (ingested living microorganisms which beneficially affects human health and digestive comfort) consumption has been shown to weakly impact bacterial composition, while significantly changing the functional gene content (Gerritsen et al. 2011; McNulty et al. 2011). Antibiotic consumption is also an important factor shaping the bacterial composition and diversity, as antibiotic treatments often used in western countries induce important changes in bacterial composition (Sommer et al. 2010). These changes occur rapidly, and bacterial equilibrium recovery is dependent on the treatment duration (Lagier et al. 2012; David et al. 2014). However, it has been shown that most of the commensal bacteria possess one or more genes responsible for resistance to antibiotics, and antibiotic treatments affect only some phyla, depending on the antibiotic used. But considering the very high number of microorganisms living in the gut, it can be considered as a reservoir of bacterial species sharing and exchanging antibiotic resistance genes. Actually, since the beginning of antibiotic use at a large scale, some commensal bacteria have developed strong antibiotic resistance capacity, which may be correlated to antibiotic resistance gene spreading from pathogens.

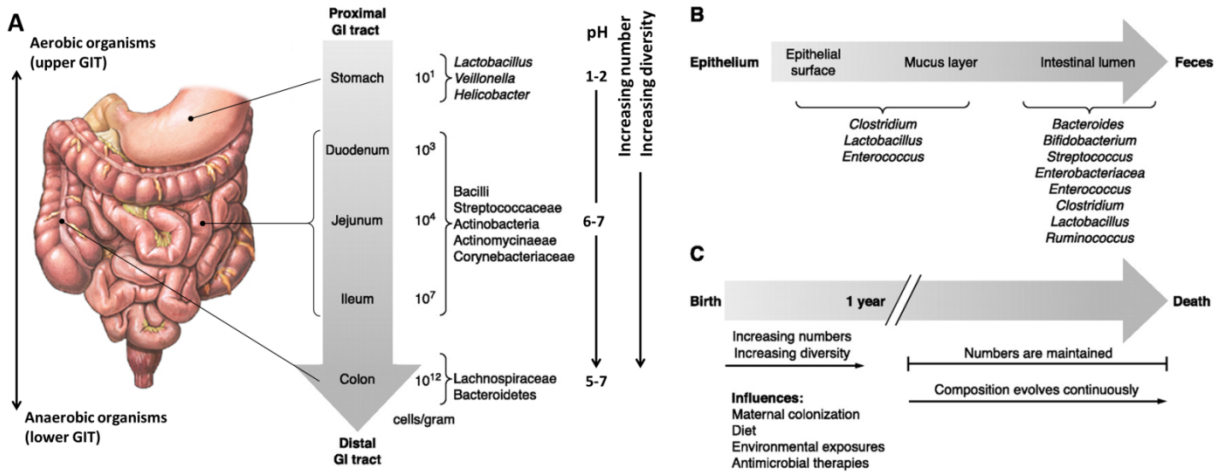


Figure 1: Overview of bacterial composition of the human gut.

A. Overview of bacterial distribution along the GIT, showing the pH and bacterial variations found at the phylum level from stomach to colon. B. Bacterial distribution across the gut. This pane shows the encountered variations at the phylum level between the lumen and the mucosa-associated region in the small intestine. C. Temporal variation in bacterial composition. This pane shows the global evolution of gut microbiota from birth to death and emphasizes the dynamic aspect (adaped from Sekirov et al, Physiological Reviews, 2010 and S. Prakash et al, J. Biomed.Biotechnol, 2011).

Despite a great number of factors which can affect microbial diversity and amounts, it is surprising to note that at the phylum level, the human gut microbiota is stable, and its composition is conserved between individuals whatever their origin. Three major phyla are found in everybody's gut microbiota, which only differ on the predominant group. Arumungam *et al* claimed the existence of three robust clusters of specified bacterial composition and abundances, termed enterotypes (Arumugam et al. 2011) (fig. 2). Enterotype 1 has been showed to be enriched in *Bacteroides* and *Parabacteroides*, while enterotype 2 is enriched in *Prevotella*. Enterotype 3 is the most frequent, and is enriched in *Ruminoccocus*. The authors also hypothesized that these enterotypes are correlated to a specialization for the route gut microbiota use to get energy from diet. However, it might be reductionist to consider that some bacterial genera are dedicated to certain substrate metabolization and not to others. Rather, it would be more precise to consider that all bacteria can grow on many dietary compounds, but some have developed better harvesting capabilities on some specific ones. This gives them a competitive advantage when growing on them compared to other bacteria, and thus increases their number in the total bacterial pool. Enterotype 1 is thus expected to be more efficient in carbohydrates and proteins use, while enterotype 2 is hypothesized to be more efficient in mucin glycoprotein degradation by the synergetic activities of *Prevotella* and *Desulfovivrio*. *Ruminococcus*, the major group of enterotype 3, is also known for its ability to degrade mucins, and showed to be enriched in sugar membrane transporters providing advantageous for simple sugar resulting from mucin degradation. Metagenomic analyses also concluded that aside the three enterotypes, a core microbiota exists in humans, comprising 57 species found in 90% of people. This 'core 57' contains bacteria from all phyla, and the enterotypes only refers to the dominant phylum in the total microbiota (Qin et al. 2010).

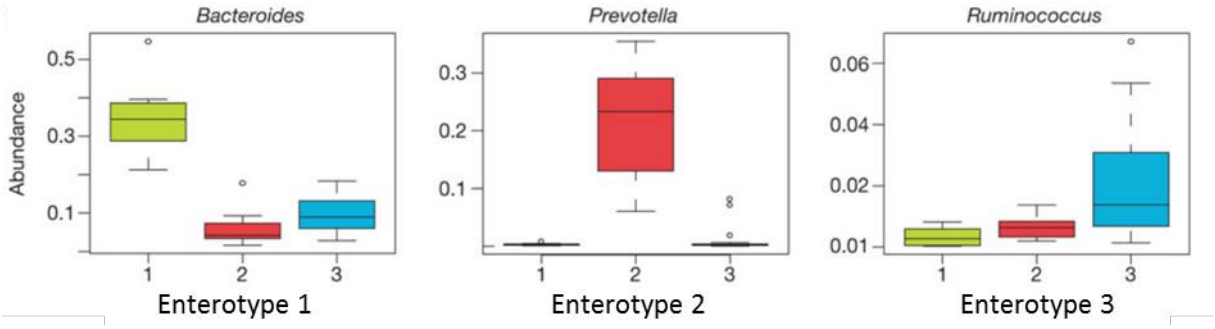


Figure 2: Human gut enterotypes description.
This figure shows the abundance of the three dominant phyla *Bacteroides*, Prevotella and *Ruminococcus* for the three enterotypes (adapted from Arumugam et al, Nature, 2011).

Among the microorganisms found in the gut, bacteria are the far most studied and known due to their overwhelming proportion (around 99 %), but others are present. In particular, eukaryotes such as *Saccharomyces spp* or *Candida* can be found, as well as fungi such as *Galactomyces* or *Paecilomyces*. Recently, archaeal species were discovered such as *Methanomassiliicoccus luminyensis* or even giant viruses such as the Senegal virus known for

infecting amoeba (*Acanthamoeba polyphaga*) (Lagier et al. 2012). Metagenomic studies recently identified viruses and phages in our microbiota, but to date little is known about these organisms due to their recent discovery and the difficulties to cultivate them, but they probably take part in the human gut microbiota roles (Reyes et al. 2010).

1.2. Physiological roles

Importance of gut microbiota in human health has long been recognized, as testifies the well-known Hippocrates' statement 'Bad digestion is the root of all evil', but even half a century ago little was known about the precise mechanisms involved in the interactions between the human host and its inhabitants. As our gut microbiota is accompanying us since the beginning of mankind, a co-evolution phenomenon occurred, which led it to actually play a key role in numerous physiological functions. Obviously, the most evident roles played by our gut microbiota concerns our metabolism, as they are living in the organ dedicated to nutrient absorption in the human body, but we will see that it is also involved in a wide variety of other functions such as immunity, primarily by competing with pathogens for the same ecological niche, but also inflammation, and even brain functions.

1.2.1. Food - gut microbiota interactions

Our own human enzymatic toolbox is not sufficient to degrade all the dietary compounds, especially the vast diversity of plant carbohydrates which are daily ingested everywhere in the world. Indeed, the human genome only encodes enzymes dedicated to the catabolism of some oligosaccharides and polysaccharides containing glucosyl residues, such as lactose (β-D-Gal*p*-1,4-β-D-Glc), sucrose (α-D-Glc*p*-1,2-β-D-Fru*f*) and starch ([β-1,4-D-Glc*p*]$_n$) by lactases, amylases or invertases. On the contrary, the human gut microbiome, which encodes a much larger panel of enzymatic activities, is able to degrade most of carbohydrate linkages found in nature. Therefore, due to the paucity in their genome to encode complex polysaccharide-degrading enzymes (the so called Carbohydrate Active enZymes, or CAZymes), mammals depend on their symbiotic microorganisms within their digestive tract to breakdown the glycans that remain undigested in the upper part of the GIT, termed 'dietary fibers'. They are mainly composed of oligo- and polysaccharides of vegetal origin, whether coming from storage tissues, such as resistant starch (the starch fraction which is resistant to the action of human enzymes), and inulin (Glc-α(1,2)-[β(1,2)Fru]$_n$), a fructan found in the roots of certain plants, such as chicory or artichoke, or structuring elements composing the plant cell wall, such as cellulose [β(1,4)-Glc]$_n$, the most abundant polymer on earth), hemicelluloses (a collection of linear and branched polysaccharides of various composition, mainly containing xylosyl, arabinosyl, glucosyl, galactosyl, mannosyl residues and uronic acids) and pectins [α(1,4)-GalUA]$_n$, substituted by xylosyl, galactosyl rhamnosyl and arabinosyl residues, sometimes associated in disaccharidic

branches. Dietary fibres have been identified as a strong positive dietary factor in the prevention of obesity, diabetes, colorectal cancer and cardiovascular diseases (World Health Organization 2003). Even if a single bacterium is not able to digest the whole panel of carbohydrate linkages present in nature, the addition of sugar degrading capabilities of all human gut bacteria allow degradation of a wide variety of plant oligosaccharides and polysaccharides. The functions and the evolutionary relationships of CAZyme-encoding genes of the human gut microbiome are being extensively studied through functional and structural genomics investigations (Flint et al. 2008; Lozupone et al. 2008; Martens et al. 2008a; Mahowald et al. 2009; Drissi et al. 2014), which are nevertheless restricted to cultivated bacterial species. CAZyme diversity has also been described in the pioneering metagenomics studies focused on this microbiome (Gill et al. 2006; Turnbaugh and Gordon 2009; Qin et al. 2010). These pioneer works carried out in particular by the Human Microbiome Project (HMP) and MetaHit consortia, revealed the presence of at least 96 families of glycoside-hydrolases (El Kaoutari et al. 2013), making the human gut metagenome one of the richest source of CAZymes (Li et al. 2009). Thanks to its performance to breakdown polysaccharides, the colon has often been considered as a complex bioreactor, and one of the effects of human gut bacteria is the release of short chain fatty acids (SFCA), mainly acetate, propionate and butyrate as carbohydrate fermentation products, which are then absorbed by the host to provide to it at least 10% of energy intake (Scott et al. 2013). In addition to its key role in fermenting dietary carbohydrates, the gut microbiota is capable of metabolizing other dietary constituents. For example, it has been shown to modulate host intake of dietary lipids by promoting host's lipases activity (Hooper et al. 2001; Bäckhed et al. 2004). In addition, even if most of the proteins are efficiently degraded and absorbed by the host in the upper part of the GIT, gut microbiota secretes proteases, which are key elements of secreted nitrogen recycling (Hooper et al. 2002). More importantly, the gut microbiota is involved in vitamin metabolism, as some gut bacteria possesses the pathways for their production (vitamin K, biotin, panthotenic and folic acids for example) and are actually the first vitamin B12 provider for their host, as bacterial production is the only source in nature. The only exception concerns vitamin C, found in fruits and vegetables, which is not produced by primates' gut microbiota (Resta 2009).

1.2.2. Host - gut microbiota interactions

Gut bacteria play a key role in protecting host against pathogens and in participating in the control of immunity (fig. 3). With about 400 m², the gastro-intestinal mucosa is far the largest body's contact surface with the environment, compared to the 1-2 m² for the skin. It is therefore the main entrance route for pathogenic microorganisms and their immuno-active epitopes (Fukata and Abreu 2008). Pathogen invasion is limited by nutrient absorption competition, and for maintaining in the gut. To invade the host, pathogens first need to

adhere to epithelial cells, but protective effect of commensals has been demonstrated by specific adherence or degradation of heavily secreted or anchored (glycocalyx) proteins composing the mucus produced by goblet cells (Juge 2012).

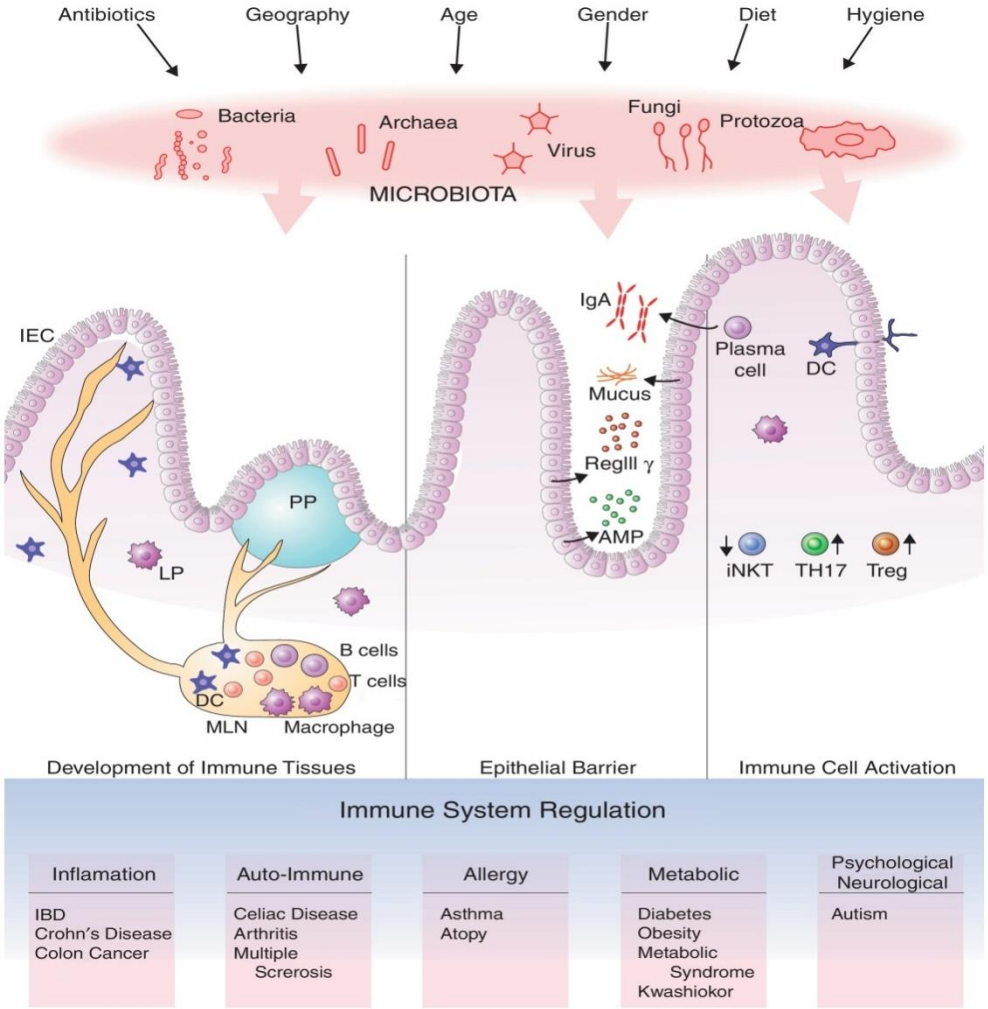


Figure 3: Global relationships between gut microbiota and the immune system.
This figure shows the main elements involved in GI immune homeostasis. IEC Intestinal Epithelial Cells, LP: Lamina Propria, PP: Payer's Patches, DC: Dendritic Cell, MLN: Mesenteric Lymph Nodes, AMP: AntiMicrobial Peptides. From D. Erturk-Hasdemir & Dennis L Kasper, Current Opinion in Immunology, 2013.

The mucus is formed by two layers of secreted glycoproteins, which vary in nature and thickness upon the localization in the GIT. The human genome contains twenty mucin encoding genes (Muc 1 to Muc 21), whose expression is varying upon the considered epithelium. In the colon the sole expressed secreted mucins are Muc2 and Muc19, while anchored mucins encoded by Muc1, 3A /3B, 11, 12, 13, 15, 17, 20 and Muc21 are found (Dharmani et al. 2009). The major feature common to these proteins is that they are found heavily glycosylated, providing them their gel-like visco-elastic, rheological and chemical properties. These large glycoproteins, whose sizes vary from 0.5 to 20 MDa are composed of distinct regions, with the presence of central repetitive domains holding O-glycosylation sites (fig.4). These O-linked oligosaccharides can make up to 80% of the total mass of the

mucin, while the rest is mainly formed by the proteic part. The central, heavily glycosylated region of mucin is flanked on each side by cysteine rich domains and von Willebrand factor (vWf) C and D domains, holding additional O- and N-glycosylation sites, involved in mucin oligomerization to form gels (Bansil and Turner 2006).

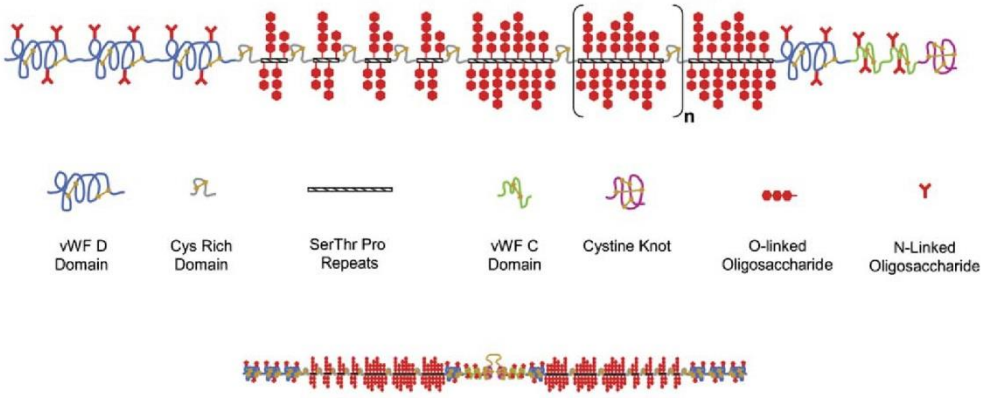


Figure 4: Schematic view of a mucin glycoprotein.
The upper part details the basic features of a secreted mucin. The central zone is composed of a variable number of tamdem repeats (Serine-Threonine-Proline repeats), containing the attachment sites for O-glycans. On each side are present cysteine-rich and vWf C and D domains, holding N-glycans, and required for oligomerization of mucin proteins, as shown below (adapted from Rama Bansil et al, 2006).

O- and N-glycosylation (fig. 5) are co-translational modifications of proteins, which consists in addition of sugars by host glycosyl-transferases on specific residues of the protein being translated in the endoplasmic reticulum. Mucin O-glycans are composed of a *N*-acetylgalactosamine (GalNAc) residue α-linked to the hydroxyl moiety of serine and threonine on the central repeated region. The GalNac residue may be extended by addition of other sugars, such as galactose, *N*-acetylglucosamine, fucose or sialic acid, forming several types of core O-glycans, termed core 1 through 8. These diverse core structures can then be subjected to further extension, yielding extended core structure 1 through 4. N-glycans are found on the vWf domains, and exist as different types, all composed of a core pentasaccharide $Man_3GlcNAc_2$ that is covalently linked to asparagine, at the sequence consensus N-X-S/T. Other sugars comprising mannose, glucose, *N*-acetylglucosamine, fucose, galactose or sialic acid are added to this core structure to form three main types of N-glycans, of which the structure and degradation processes will be detailed in chapter 3. The several glycosidic linkages found on these host glycans constitute specific epitopes that are used for cell recognition and bacterial adherence (Bolam and Koropatkin 2012; Cameron et al. 2014).

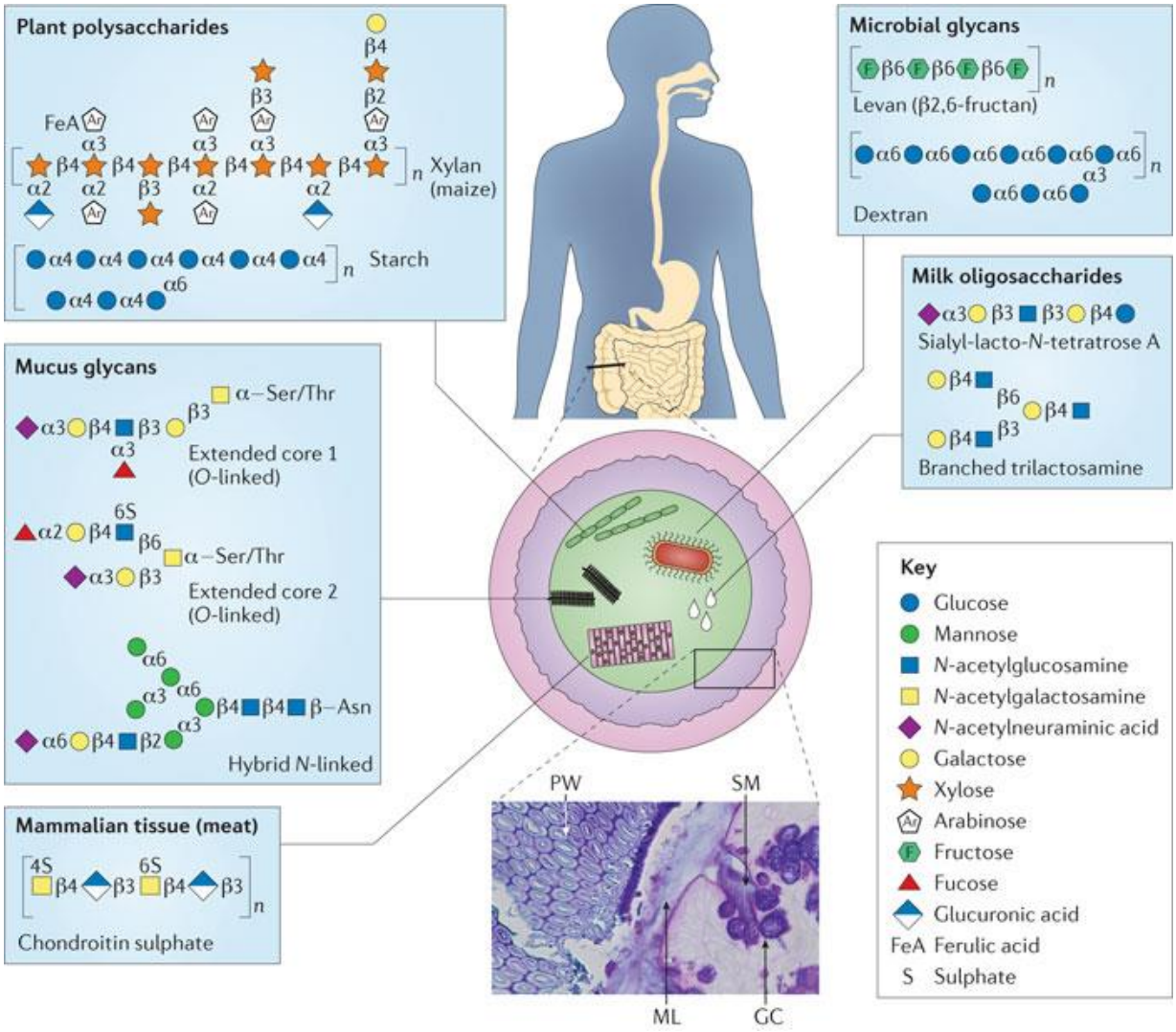


Figure 5: Overview of glycans accessible for gut bacteria.

This figure gives only a brief overview of the diversity of glycans available for gut bacteria, as a large part is diet-dependent, such as plant or meat polysaccharides which arbor great variations both in structures and amounts. For instance, a single representative of hemicelluloses is shown here, while other are encountered, such as xyloglucan, galactans or mannans. The lower panel is a cross section of mouse colon showing mucus-producing goblet cells (GC), secreted mucus (SM), the mucus layer (ML) and fragments of plant cell wall (PW) coming from diet. From Koropatkin et al, 2012.

Once the mucus barrier is crossed, crosstalk between gut bacteria and the host's immune system can be established, as evidenced by the following examples, implicating various bacterial genera. The polysaccharide A (PSA) produced by the ubiquitous human commensal *Bacteroides fragilis* strains induces CD4+ T-cell proliferation (Mazmanian et al. 2005), while butyrate (part of ubiquitously produced SFCAs) induces production of LL-27 cathelicidin, an antimicrobial peptide produced by colon epithelial cells (Termén et al. 2008). *Lactobacillus species* such as *Lactobacillus reuteri* DSM 12246 or *Lactobacillus fermentum* Lb20 were found to regulate dendritic cells, which play a key role in T-cell maturation into Th1/Th2 or Th17 (cells of the adaptive immune system) (Christensen et al. 2002), while *Streptococcus salivarius* strains CIP102503^{T} and JIM8772 Sn were shown to exhibit an anti-

inflammatory effect by down regulating the production of IL-8 by intestinal epithelial cells (Kaci et al. 2011). Peptidoglycan of Gram-negative bacteria were also found to induce lymphoid tissue genesis, a tissue part of the lymphatic system, found in the basal part of the gut mucosa and producing B lymphocytes (Bouskra et al. 2008). This tissue can be attacked by *Helicobacter pylori* strains in the stomach, causing lymphomas.

One other element illustrating the mutualistic nature of the host-microbes relationship is the ability of gut bacteria to increase host's absorptive capability by promoting formation of new blood capillaries in the small intestine during the initial colonization process of the gut. It has been demonstrated that some gut bacteria such as *Bacteroides thetaiotaomicron* VPI-5482 can enhance angiogenesis through activation of Paneth cells, located at the bottom of Lieberkühn crypts (Stappenbeck et al. 2002). Activation of human intestinal microvascular endothelial cells by commensal bacteria's lipopolysaccharides (LPS) through Toll-like and Nod-like receptors induced a pro-angiogenic signal comparable to the one of vascular endothelial growth factor (VEGF), one of the most powerful angiogenic factor known (Schirbel et al. 2013).

In addition to a deep and tight communication between gut microbiota and host's immune system, recent studies revealed an important bidirectional communication with the central nervous system. In response to a restrained stress, germ free mice were shown to exhibit an increased stress reactivity through higher corticosterone and adrenocorticotrophin responses compared to standard animals, an effect that could be corrected using probiotics (*Bifidobacterium infantis*) (Sudo et al. 2004). As stress was known to increase intestinal permeability, the authors hypothesized that it could have given gut bacteria an easier way to access to neurons belonging to the enteric nervous system, and directly influence nervous stress circuits in newborns. This was confirmed by studies showing an induction of the neuronal activation marker cFOS of the vagus nerve sensory neurons by oral administration of pathogens (*Campylobacter jejuni* ATCC 29428 and *Citrobacterrodentium* ATCC 51459, and even directly in the brain by *Bifidobacterium infantis* 35624 in germ-free mice (Sudo et al. 2004; Goehler et al. 2008). This communication between gut and host through the peripheral nervous system (PNS) has started to be characterized at the molecular level. Indeed, in culture, *Lactobacillus brevis* DPC 6108 and *Bifidobacterium dentium* were shown to produce γ-aminobutyric acid (GABA) from glutamate (Barrett et al. 2012). In addition, *L. rhamnosus* was shown to alter expression of GABA receptors in brain regions related to stress, *in vivo* (Bravo et al. 2011). These elements are very important since glutamate is a stimulating neurotransmitter, while GABA is one of the main inhibitory neurotransmitters of the central nervous system in mammals, and dysfunctions in GABA signaling are implicated in anxiety and depression, while an imbalance between GABA and glutamate is linked to epilepsy (Cryan and Kaupmann 2005). The modulation of neurons' activity at the molecular level conducted the authors to even

consider the human gut microbiota as an epigenetic element in the supra-organism *Homo sapiens* (Stilling et al. 2014). Thus, if one of the elements of the supra-organisms is altered, it may lead to a plethora of physiological dysfunctions or diseases.

1.3. Health issues

To ensure proper functioning of the GIT, a fine equilibrium must be found and maintained between the host, its commensal microbes, and diet. Under normal conditions, this is achieved by a tight control of pro- and anti-inflammatory responses of the host's immune system which allows for a correct discrimination between commensals and pathogens, both the host and the microbes being effectors of this equilibrium (fig. 6). Therefore, several elements may be involved in dysfunctions or diseases due to or causing a disturbance of the microbiota, a state termed 'dysbiosis', which can also manifest itself as a lowered amount of beneficial bacteria (such as bifidobacteria). However, in the vast majority of cases, diseases affecting the GIT are directly or indirectly associated to an unbalance of these pro- and anti-inflammatory responses at the gastrointestinal epithelium level. These deregulations of the immune system often generate excessive pro-inflammatory responses leading to lesions in the intestinal epithelium, such as reduced mucus production or loss of barrier integrity. Symptomatically, it appears as a dysbiosis, associated to intestinal disorders ranging from diarrhea to much more serious illnesses such as inflammatory bowel diseases (IBD) or cancers affecting all parts of the GIT. In addition, this dysbiosis also alters the microbiota-associated normal functions and could lead to several metabolic diseases or deficiencies such as diabetes or obesity. These can also generate affections spreading to the whole body, which is obvious for infecting pathogens, bacteria or viruses, but the deregulation of the immune system can also generate hyper sensibility, leading to allergies, or auto-immune diseases.

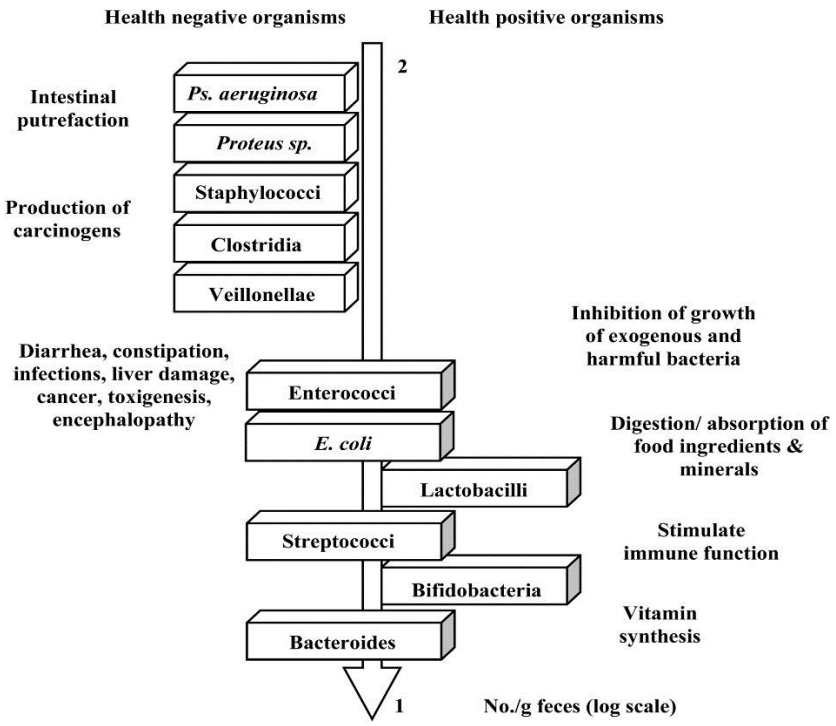


Figure 6: Overview of the effects of human gut bacteria on human health.
This figure lists the major bacterial phyla found in the GIT and the known associated behavior of each in regard of their positive, negative, or neutral effect on human health (from Rastall RA, J Nutr. 2004).

1.3.1. Inflammatory Bowel diseases

Among the intestinal diseases, one major concern are the inflammatory bowel diseases (IBD). They form a group of diseases characterized by a thinner and discontinuous mucus layer leading to chronic inflammatory conditions in the small intestine and the colon (Strugala et al. 2008). They are represented by two main diseases, the Crohn's disease (CD) and the ulcerative colitis (UC) manifested by symptoms such as diarrhea, severe abdominal pains or weight loss. Crohn's disease manifests all along the GIT, while UC is restrained to the distal part of the colon. In all cases, patients bear reduced microbiota diversity, but, as for the chicken or the egg, whether this altered microbiota is causal or secondary is still in debate (Walker et al. 2011). There is currently no etiologic agent identified for IBD, but theories have emerged to explain these affections. A first hypothesis for IBD would be that they would be consequent to an inadequate genetic predisposed host's immune response to the microbiota, in which the host does not tolerate its own microbiota any longer. This would lead to a disturbance of the signals that maintain barrier function, associated with an inflammation that can favor the selection of aggressive symbionts (Juste et al. 2014). Indeed, opportunistic pathogens may trigger this pre-inflammatory state. Even if no single species has been found constantly associated with IBD, adherent-invasive *Escherichia coli* (AIEC), *Listeria monocytogenes*, *Chlamydia trachomatis*, *Pseudomonas maltophilia*, *Bacteroides fragilis*, *Mycobacterium kansasii*, and *Mycobacterium avium paratuberculosis* appear to be regularly associated with IBD, while levels of butyrate producers such as *Roseburia* and

Feacalibacterium appear to be lowered (Strober 2011; Ouwerkerk et al. 2013a; Naser et al. 2014).

An increasing number of studies also tend to show that IBD would be initiated by an over-controlled auto-immune reaction, in which a pathogenic bacterial epitope would mimic a commensal's one (Sartor 2008; Principato and Qian 2014). One could think that such an event appears to be highly probable, considering the *quasi* unlimited number of epitopes presented by the commensal microbiota that would lead to CD4+ memory T cells number increase. In fact, under normal conditions, the immune system is finely balanced to discriminate between commensals and pathogens, avoiding an uncontrolled growing of these particular adaptive immune memory cells. First, to avoid a chronic stimulation of the immune system, a physical separation between the microbiota and immune cells is strongly established, through a mucosal firewall combining the mucus layers, antimicrobial peptides and secreted immunoglobulin A (IgA) (fig. 3). Two mucus layers are actually found at the surface of the GIT epithelium. The inner one, firmly adherent, contains a lot of antimicrobial peptides and secreted IgA associated with a dense network of mucins and is assumed to be devoid of bacteria. On the contrary, the outer layer only contains diluted antimicrobial peptides and IgA which allow bacteria to penetrate it. While the inner mucus layer thickness remain constant along the GIT (100µm), the outer one is varying upon the location in the GIT (150-70 µm), providing it with a variable viscoelasticity, giving the host a way to control the bacterial charge associated to its epithelium upon the location, and fatalistically saying, finely tune the sensitization of immune cells (Ouwerkerk et al. 2013b). Colonization of the inner mucus layer by commensals has been associated with respectively 94% and 98% of case of UC and CD, 78% of cases of colitis, while only 11% in healthy patients, showing the importance of inner mucus bacterial colonization in gastro-intestinal malignancies (Dejea et al. 2013). In normal cases, if bacteria penetrate this defense line, immune system is activated, involving Th1 lymphocytes. After clearance of intruders and restoration of the mucosal firewall, regulation of this adaptive response normally occurs, involving regulatory T lymphocytes (Treg), a subgroup of T cells involved in T lymphocyte regulation, especially in regard with self-specific and commensal-specific ones (Belkaid et al. 2013). The key point in IBD appears to be in the incorrect commensal-specific T cell shut down by Treg lymphocytes, which leads to their *quasi* permanent activation. As a consequence, a massive amount of pro-inflammatory cytokines are produced such as HIP/PAP (homolog of the antimicrobial peptide RegIII-γ produced by TLR activation) which has been shown to be specifically increased in inflamed parts of the colon and in enteropancreatic tumors (Ogawa et al. 2003). Genetic susceptibility has been shown to be also involved in the pathology. For instance, mutations in the DLG5 gene, involved in epithelial protein scaffolding, or MDR1, an efflux transporter for drugs, have been associated respectively with CD and UC, underlining the fact that these diseases remains largely multifactorial (Sartor 2006). Finally, diet-related causes may be involved in IBD development. Indeed, existing data suggest that decreased

dietary plant polysaccharides shifts microbiota's nutritional support to host mucus glycans, leading to an increased stress over this barrier (Mahowald et al. 2009). Logically, the idea of using pre- or probiotics for IBD treatment emerged, but to date, studies are mostly restricted to *in vitro* and animal model analyses, with the exception of very few analyses on small patient cohorts (Scaldaferri et al. 2013). The results remain preliminary and partially unconsistant even if these studies tend to show a decrease in inflammation markers when using prebiotics (inulin in combination with fructo-oligosaccharides) or some probiotics such as *Lactobacillus GG* for UC.

1.3.2. Colorectal cancer

The third most common cancer worldwide in 2012, colorectal cancer, is of particular interest in public health (World Health Organization 2003). Even if carcinogenesis remains primarily due to human DNA damage, environmental factors are the main source of elements leading to DNA instability. Also, in colorectal cancers, enrichment in specific gut bacteria have been observed, particularly for *Clostridium* or *Bacteroides spp*, but again, no evidence tend to show that this dysbiosis is the cause or the consequence of the disease (Scanlan et al. 2008). Statistics show that diet is a major cause of colorectal cancers and it has been proposed that microbiota metabolic by-products could be involved in the tumorigenesis process (O'Keefe 2008). Of particular interest is hydrogen sulfide, a major toxic compound, which is produced by sulfate reducing bacteria through amino acids metabolism, and hence proteins degradation (Kim et al. 2013). Epidemiological surveys suggests that the western high-meat diet is positively correlated with an increase in colorectal cancer risk, but a clear link between high-protein diet and increased amounts in sulfate reducing bacteria remains to be established. However, even without involving diet specific compounds, it has been demonstrated that metabolization of common nutrients by commensals such as *Enterococcus faecalis* ATCC 29212 may produce reactive oxygen species (ROS), which are well known to induce DNA damage, hence promoting tumorigenesis (Wink et al. 2011).

1.3.3. Metabolic diseases

Obesity is a multifactorial disease of major concern in western societies resulting from the interaction between the environment, genetic and life-style factors. Murine models clearly show that gut microbiota play a key role in its development, since transplantation of gut microbiota from obese mice to lean germ-free animals lead to a 60% increase in body fat over two weeks without any change in diet (Bäckhed et al. 2004). Contradictory results seem to involve the decreased *Bacteroides*/*Firmicutes* ratio observed between obese compared to lean persons, but the presence in microbiota of species from *Oscillibacter* and *Clostridium* clusters XIVa and IV seem to be a strong obesity predisposing element (Turnbaugh et al.

2006; Schwiertz et al. 2010; Duca et al. 2014). Moreover, a recent report analyzing the richness of gut microbiota and total gene count of 292 individuals concluded that obesity is clearly associated with a decreased diversity in the microbiota, a reduced diversity in functional genes, which leads to a metabolic profile change (Le Chatelier et al. 2013), particularly with an over-representation of bacterial genes related to energy harvest (Turnbaugh et al. 2009). Gut microbiota assistance in host's energy intake thus seems to be the most significant contribution to obesity development, even if other effects have been implicated. Indeed, Bäckhed *et al* showed that gut microbiota transplant was also associated with a 2.3 fold increased triglyceride production in the mouse liver (Bäckhed et al. 2004). SFCAs production was also shown to be increased in obese rather than lean patients, with an associated shift towards propionate production instead of butyrate (Schwiertz et al. 2010).

It has long been known that type-1 diabetes is primary an auto-immune disease, due to the presence of self-reacting T lymphocytes, destructing insulin-producing β-cells in the pancreas. Changes in the gut microbiota composition were found to promote this pathology. Indeed, in rats, dietary changes inducing a decrease in gut bacteria amounts yielded a protective effect against type-1 diabete development (Brugman et al. 2006). One can imagine that development of such an auto-immune pathology may follow the same process as for IBD, but how the gut immunity affects distal parts of the body remains to be elucidated, even if it is becoming clear that a great part of human auto-immune diseases may be linked to an impairment of the gut immune system regulation (Sorini and Falcone 2013). In addition, such a process is strongly hypothesized to also occur in numerous allergies or asthma cases (Nikulina et al. 2004).

Even if it is obvious that pathogenic bacteria may induce GIT malignancies (like the well-known stomach inhabitant *Helicobacter pylori* which may lead to gastric carcinoma (Bartsch and Nair 2006) and other diseases, several commensals, mostly belonging to the *Bacteroides* phyla, are opportunistic pathogens, and can switch from friends to foes depending on the condition they encounter (fig. 6).

1.4. Bacteroides: colleagues or traitors?

Bacteroides species counts from 5% up to 30% of the total colonic bacteria depending on the enterotypes considered, making them the most abundant anaerobes in the gut (Yang et al. 2009) (fig. 2). They are strict anaerobes, non-spore forming, Gram negative bacilli living in the GIT of humans, and many other ecosystems (mammal and insect guts, wastewaters, soils or composts). When associated to animal hosts, they usually maintain a mutualistic relationship with him, but can cause severe pathologies such as abscesses or septicemia in case of gastrointestinal epithelium rupture. For instance, *Bacteroides fragilis* is actually involved in more than 80% of anaerobic infections whatever the location in the body

(Wexler 2007). Most of them were discovered in human feces, and *Bacteroides* taxonomy has been subjected to major revisions in the past decades, mainly to be indicative for clinicians about the pathogenic or antibiotic resistant behavior of strains (Nagy et al. 2009). If around 70 % of gut bacteria are uncultured, *Bacteroides* strains have been amongst the firsts isolated (mainly from human feces) and are the most studied gut commensals due to their large proportion in the total bacterial count, but also because of their involvement in gut homeostasis and development. Among *Bacteroides* species, two are of prime interest: *Bacteroides fragilis,* which is involved in several pathologies and *Bacteroides thetaiotaomicron,* which constitutes a model organism for functional genomics studies targeting *Bacteroides* species.

1.4.1. Beneficial effects

To maintain in the gut, a bacterium need to evade host's immune response. In link to that purpose, *Bacteroides* species have the ability to finely modulate the antigens they present to the immune system, by flipping promoters of genes involved in capsular polysaccharides synthesis to appear as a ghost for the immune system. Even if genomic inversions are ubiquitous in bacteria, it has been estimated that half of *Bacteroides thetaiotaomicron* VPI-5482 genome is subjected to such inversions which occur at very high frequencies (up to one each four generations) which underlines the importance of this mechanism for this strain survival (Cui et al. 2012). On the other hand, many aspects of the intercourse between *Bacteroides* and the host are useful for this last one. As described before, *Bacteroides* species are also involved in vitamin production. Biotin, thiamin, folates, riboflavin, panthotenic acid and menaquinones commensally produced are absorbed by epithelial cells in the colon (the habitat of most of *Bacteroides* cells), contrarily to dietary vitamins which are absorbed in the small intestine (Rossi et al. 2011). More, germ free mice colonized with *Bacteroides thetaiotaomicron* VPI-5482 showed an increase in the expression of glucose transporters, an upregulation of genes involved in lipid absorption and metabolism, as well as upregulation of many other genes involved in copper transport or drugs detoxification (Zocco et al. 2007). In addition, colonization of germ-free mice with *Bacteroides thetaiotaomicron* VPI-5482 induced an upregulation of proteins involved in desmosomes cross-bridging, thus increasing intestinal barrier fortification (Hooper et al. 2001).

In immunity, interactions between *Bacteroides* and the immune system are beneficial for the host. *Bacteroides fragilis'* zwitterionic polysaccharide is shown to activate CD4+ T cells and B cells, increasing immunoglobulins production. In mice, *Bacteroides thetaiotaomicron* VPI-5482 was shown to stimulate Paneth cells for Ang4 and RegIIIγ production, antimicrobials peptides directed against *Listeria monocytogenes* and other Gram + bacteria (Hooper et al. 2003). In addition, *Bacteroides* reduces inflammation by decreasing

expression of NF-κB controlled genes through the export of RelA (the NF-κB transcription factor p65 subunit) out of the nucleus (Kelly et al. 2004).

1.4.2. Deleterious effects

However, *Bacteroides* species can turn to foes, as some of them are clearly involved in a variety of pathologies in case of epithelial breach. *Bacteroides* can be a cause of illness, due to their role as reservoir of antibiotic resistance genes, which can pass to other aggressive bacteria. For example, a study analyzing 275 metagenomes showed that 97% of the gut metagenomes harbored the same resistance genes to tetracycline (Ghosh et al. 2013), as an identity percentage >96% between *tetQ* genes of different *Bacteroides* species is observed (Huddleston 2014). One of the major features of pathogenic bacteria is that they possess massive destruction weapons. In that sense, the *B. fragilis* toxin (fragilysin) makes it clearly a pathogen. It is a secreted metalloprotease which induces physical damage to the intestinal epithelium by cleavage of zonula adherens cadherins, the proteins involved in cell tethering (Hwang et al. 2013).

1.4.3. Carbohydrate foraging by *Bacteroides* species

Bacteroides thetaiotaomicron (BT), first identified from human feces in 1912 by A. Distaso, is an important member of the *Bacteroides* genus, accounting for up to 12% and 6% of the *Bacteroides* and total gut microbes (in terms of bacterial cells), respectively. BT VPI-5482 genome, which was the first *Bacteroides* genome to be sequenced, consists in one circular chromosome sizing 6.26 Mbp, and of the 33 kbp p5482 plasmid. They code for 4864 and 38 predicted proteins, respectively (Xu et al. 2003). Among them, an impressive proportion is dedicated to polysaccharide sensing, adhesion, synthesis or degradation processes, which is only exceeded by *Bacteroides cellulosilyticus* WH2 (McNulty et al. 2013) in term of CAZyme content. *B. thetaiotaomicron* is able to cleave most of the major glycosidic linkages found in Nature (Xu et al. 2004). No less than 261 catalytic CAZyme modules were found to be encoded by the BT VPI-5482 genome, together with 208 sugar transporters and sensors (Martens et al. 2009b). As we saw above, among the overall carbohydrate content supplied to the gut through diet, only a small fraction is made of simple sugars, while the rest is composed of dietary fibers. The success displayed by *Bacteroides* species in their colonization of the colon is therefore associated to an efficient breakdown of a wide array of carbohydrates whatever they come from plant, mucus glycans and even meat, milk or bacterial glycans, which is a very useful feature when living in an omnivorous host's intestines (Martens et al. 2011)(fig. 4).

In addition to their extraordinary capability to breakdown plant polysaccharides, gut *Bacteroides* are indeed also able to efficiently metabolize host glycans, as highlighted by the

pioneer transcriptomic studies of Martens et al, targeting *Bacteroides thetaiotaomicron* VPI-5482 and *Bacteroides ovatus* ATCC 8483 (Martens et al. 2008a; Martens et al. 2011). The *Bacteroides* capability to grow on host glycans is a great advantageous in the ecosystem, because it gives to them the access to a permanently renewed carbon source. For instance, it has been shown that BT VPI-5482 grows on host glycan only when dietary fibers are absent in the gut, for example in infants before weaning (Sonnenburg et al. 2005). Prioritization in host glycan use is observed, glucosaminoglycans being used first, followed by O-glycans and finally N-glycans (Martens et al. 2008b). In the same way, the other well-characterized *Bacteroides* species, *Bacteroides fragilis*, is able to efficiently use mucins as carbon source, particularly in case of infections, when it escapes from the gut and no more accesses to any other carbohydrate source of dietary origin (Cao et al. 2014).

As seen in this chapter, gut bacteria play a critical and sometimes dual role in maintaining or negatively affecting human health. To maintain in this competitive environment, they have developed a highly complex enzymatic machinery to harvest carbohydrates, their main carbon source, from both dietary and host glycans. The massive sequencing of gut bacterial genomes and metagenomes highlighted these last years the vast diversity of the CAZymes produced by gut bacteria, notably from the *Bacteroides* genus. To understand how these enzymes act, alone and synergistically, to breakdown complex carbohydrates, Chapter 2 will focus on their classification, mode of action, and genomic organization.

2. Glycoside-phosphorylases, these fascinating Carbohydrate Active enzymes

Carbohydrates are one of the four main building blocks of biological macromolecules. They can build-up to form either polysaccharides or associate with proteins and/or lipids to form glycoconjugates. Carbohydrates play various and significant roles in a multitude of cellular processes, including cell structuration, signaling and differentiation, modulation of immune response, inflammation, and mediation of microbial and host-microbe interactions. Because of their highly versatile structures and biological functions, and because they are forming a substantive part of biomass on earth, carbohydrate based molecules are used for a broad range of applications in food and feed industries (functional foods, prebiotics or as additives), health (as drugs, drug carriers or vaccine intermediates), as well as for 2nd generation biofuel production and for the synthesis of biosourced materials as alternative to fossil carbon-based compounds. Understanding the processes by which carbohydrates are produced, modified, recycled or degraded by enzymes is therefore of prime interest. As briefly introduced in Chapter 1, these enzymes are called Carbohydrate Active Enzymes (CAZymes). They are central elements of numerous microbial ecosystems, as they are required to transform complex glycans into simple sugars to support cell growth.

Chemically speaking, carbohydrates are poly-hydroxyl aldehydes or ketones, having a general formula $C_n(H_2O)_n$, n being between 3 and 9. One carbon is holding a carbonyl moiety, which can be held at an extremity of the molecule (aldose) or in the carbon chain (ketose). The other carbons forming the molecule are holding hydroxyls, which means that various asymmetric centres exist on these molecules. The configuration of the asymmetric carbon atom furthest from the carbonyl group gives the D or L configuration. Sugars containing more than 5 carbons can undergo spontaneous cyclization in solution, involving the carbonyl moiety and another hydroxyl of the molecule, leading to the formation of a hemiacetal (aldoses) or hemiketal (ketoses), and a 5-atoms ring (furanose) or 6-atoms ring (pyranose). Thus, the carbon holding the carbonyl becomes an additional asymmetric centre, from which the two anomers α and β are deduced.

Glycosides, or glycans, are composed of repetitive units of carbohydrate monomers, linked by an O-glycosidic bond between the anomeric carbon and a hydroxyl of another monomer (also called glycosyl residue). They include disaccharides, oligosaccharides, of which the degree of polymerization (DP) ranges between 3 and 10, and polysaccharides of which the DP is higher than 10 (Cummings and Stephen 2007). The combination of the nature of the monomer residues, their absolute configuration, the way they cyclisize, their anomery and the hydroxyl involved in the linkage, give a tremendous variety of glycoside

structures, which largely exceeds protein diversity for the same residue number (Laine 1994).

2.1. The CAZy classification

To handle this huge diversity, life has evolved a panel of enzymes to recognize, modify, polymerize or degrade glycans. The Enzyme Classification system (EC) was created in 1955 to classify enzymes (CAZymes and others), and help to handle this great diversity. This classification is a functional classification which was designed mainly to standardize the name of enzymes and give a first hierarchized nomenclature system based on their catalytic activities. Basically, it associates each enzyme to a four number code indicating the type of reaction, the substrate and product. However, this classification addresses to all enzymes, which makes it inefficient to handle the CAZyme diversity because it is only based on functional properties of enzymes, which abolishes any activity prediction power for the constantly growing number of sequences housed in genomic and metagenomics databases. In order to address some of these problems, the CAZy classification was created. The CAZy classification is sequence-based, and was created in 1991. It allows to address several problems present in the EC classification, such as:

- Incompleteness of numbering in the EC codes
- Not adapted for enzymes that have promiscuous activities
- Not adapted to enzymes that have multiple domains with various catalytic activities
- Not adapted to enzymes which contains elements with no catalytic function

CAZymes contain one or several catalytic domains, and, often, one or several other elements distributed along the polypeptidic chain. These elements are involved in i) addressing the protein to the correct cell compartment or outside the cell (signal peptides), ii) specific binding to carbohydrates (carbohydrate binding modules), or iii) protein-protein interaction (dockerins). Catalytic domains and carbohydrate binding modules are described in the CAZy classification as ‘modules’ and constitute the base of CAZy classification (Lombard et al. 2014a). Therefore, this classification precisely allows classifying and describing a protein, even if it contains multiple modules, which is particularly relevant to interpret bio-information from genomic data. Each module is part of a ‘class’, based on its function, which can be catalytic or non-catalytic:

- Glycoside Hydrolases (GHs)
- GlycosylTransferases (GTs)
- Polysaccharide Lyases (PLs)
- Carbohydrate Esterases (CEs)

- Auxiliary Activities (AAs)
- Carbohydrate-Binding Modules (CBMs)

Glycoside hydrolases families were the first described in the classification, and gather catalytic modules able to perform hydrolysis of glycosidic bonds or transglycosylation reactions. It can be either between two carbohydrate monomers, or between a carbohydrate and another molecule, protein or lipid (Henrissat and Davies 1997).

GlycosylTransferases are enzymes involved in the synthesis of glycosidic linkages, using a phospho-activated sugar donor and an acceptor, which can be a carbohydrate or another class of molecule harbouring an hydroxyl (Coutinho et al. 2003).

Polysaccharide lyases are also enzymes which cleave glycosidic bonds in uronic acid-containing polysaccharides. However, their mechanism is not hydrolysis, but β-elimination, which does not require a water molecule. It conducts to the formation of an unsaturated hexenuronic acid residue, and a new reducing end (Lombard et al. 2010).

Carbohydrate esterases perform O- or N- deacetylation reaction on esterified sugars. Two classes of CE exist, as the sugar moiety can play the role of the acid (uronic acids) or the alcohol (neutral sugars esterified with short chain fatty acids) in the ester formation (Jayani et al. 2005).

Auxiliary activities are modules performing synergetic redox reactions with other CAZymes. It is the last class created in the CAZy classification, and contains two main kinds of enzymes: lignolytic enzymes, which are not acting on polysaccharides, but help other CAZymes to gain access to the carbohydrates encrusted in the plant cell wall by oxidising the surrounding lignin, and lytic polysaccharide mono-oxygenases (LPMOs), which degrade polysaccharides by an oxidative mechanism (Levasseur et al. 2013). These can be found associated on the same polypeptidic chain with other CAZy modules to deconstruct recalcitrant substrates, such as chitin (Vaaje-Kolstad et al. 2010).

Carbohydrate-binding-modules are non-catalytic modules often found associated with catalytic modules on a single polypeptidic chain. They enhance the associated module's activity by facilitating the association with its substrate through specific binding (Boraston et al. 2004).

Inherently to the reaction they catalyse, GHs, as well as GTs, can be divided into two main groups, basing on the catalytic mechanism they use. Indeed, through the reaction, the asymmetric carbon C_1 can undergo a conformational change, yielding a carbohydrate of identical anomery of the substrate one, or reversed (Coutinho and Henrissat 1999; Coutinho

et al. 2003). These two groups of enzymes are therefore classified as 'retaining' or 'inverting' GHs or GTs, respectively (fig. 7).

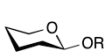
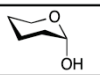
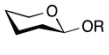
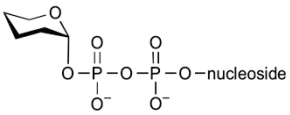
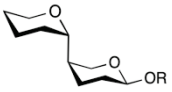
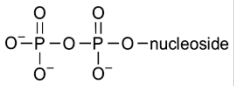
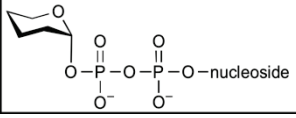
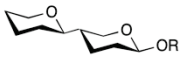
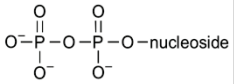
	Substrate 1	Substrate 2	Product 1	Product 2
Retaining Glycoside Hydrolase		H_2O		HOR
Inverting Glycoside Hydrolase		H_2O		HOR
Retaining Glycosyltransferase				
Inverting Glycosyltransferase				

Figure 7: Catalytic mechanisms of retaining and inverting GHs and GTs.

The CAZy classification is a sequence-based classification, which describes structural similarities in the fold and in the catalytic machinery between related sequences (Henrissat et al. 1995). Consequently, the more two sequences are similar; the more probable is the catalytic mechanism to be identical. In addition, the data integrated in CAZy are manually-curated which limits automatic assignation problems that are encountered in the other databases dedicated to CAZyme annotation, namely the CAT (Park et al. 2010) and dbCan (Yin et al. 2012) databases. In CAZy, new families are only created when at least one member has been functionally characterised. This makes it a powerful tool for inferring the activity of uncharacterised enzymes, only using their sequences. However, the identity of two sequences classified in the same CAZy family may vary considerably. Depending on it, structural identity has been split into different hierarchical levels, which brings accuracy and confidence in the description of the catalytic mechanisms and strengthens its predictive power, as we go further in the sub-categories (fig. 8).

The **superfamily** level groups together families that share a common fold, and only indicates the basic feature of the reactions catalysed by its members, for instance creating or cutting α or β glycosidic bonds.

The **clan** level groups families that share sufficient common secondary structure elements to fold in the same way, but also hold the catalytic machinery on the same secondary structure elements.

The **family** level gathers all enzyme sequences which share the same catalytic mechanism. The catalytic mechanism is identical, because sequences are sufficient identical to harbour conserved catalytic amino acid(s) and functional residues. These residues are also

positioned in the same space location in the catalytic or binding site, which means that the fold is also conserved.

The **subfamily** level is used to address the problem of poly-specific families. Indeed, several members of a family can have different substrate specificities, which makes the prediction of activity difficult based only on the family membership. The sub-division of CAZyme families into different subfamilies has not been performed for all CAZy families; it is rather family specific, based on the amount of available biochemical data. An example of sub-family creation has been done in this thesis for the GH130 family.

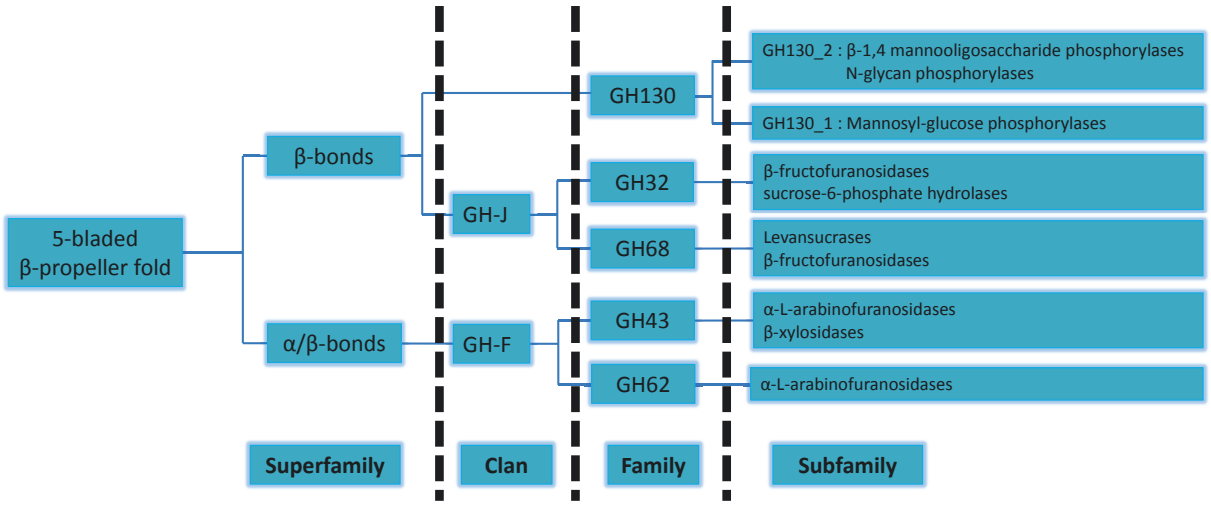


Figure 8: Hierachical organisation of the CAZy classification.
This figure is an example of the different levels of classification (superfamily, clan, family and sub-family) of modules presenting a 5-bladded β-propeller fold.

Another important additional feature of the CAZy database compared to the previous EC system is that it allows a better handling and interpreting of the large amount of genomic and metagenomic data available in regard to the ecology of organisms. For instance, identifying proteases in the genome of an unknown organism only indicates that this organism is able to degrade proteins. On the contrary, as the assimilation of a specific glycan requires the appropriate set of CAZymes, the analysis of the CAZyme content (also called the CAZome) encoded in a genome brings additional information about the organism lifestyle. The listing of an organism's CAZyme content can therefore underline its metabolization abilities, specialized to a certain substrate or rather an 'omnivorous' behaviour. More, combining CAZome analysis with transcriptomics become a strong tool to detect changes in an organism's behaviour during different stages of its life (Veneault-Fourrey et al. 2014).

Today, CAZy lists no less than 133 GH families (209972 entries), 97 GT families (164117 entries), 23 PL families (4959 entries), 16 CE families (23064 entries), 13 AA families (8340 entries) and 71 CBM families (45732 entries). This large number of families has been continuously increasing over time, with a strong speed up these lasts years. Indeed, since 2010, no less than 13 GH families were added, together with 12 CBMs families. However, only 5 GT families were added, 2 PL and no CE. But the biggest increase was observed with the creation of the 13 AA families (André et al. 2014).

2.2. Bacterial genomic organisation of CAZy encoding genes: the Bacteroidetes Polysaccharide Utilisation Loci (PULs)

To efficiently perform polysaccharide sequestration, depolymerisation and metabolization, many bacteria evolved a particular strategy, involving specific carbohydrate transporters and the synergetic action of different CAZymes, encoded by specifically regulated multigenic clusters, named polysaccharide utilization loci (PULs). As these loci have been particularly studied in *Bacteroidetes* species (Bjursell et al. 2006), the PUL denomination is specific to these bacteria (Terrapon and Henrissat 2014a). However, other PUL-like systems have been predicted in other genomes, in particular from gut bacteria, and were also evidenced in few metagenomic studies which generated sufficiently large contigs to identify these kinds of multigenic clusters (Tasse et al. 2010a; Cecchini et al. 2013). However, these metagenomic CAZy gene clusters still remain putative PULs, which are not experimentally validated.

Polysaccharide utilization by *Bacteroidetes* species has mostly been studied through BT starch utilization system (Sus), which is the prototype of all other PULs (also called Sus-like systems). The Sus-like systems are multiple cell-envelope associated protein complexes used for sensing, binding, and subsequent depolymerization of complex carbohydrates (fig. 9). For instance, the BT Sus system is composed of eight proteins (SusR, A, B, C, D, E, F, and G) expressed when growing on starch. Only several of them are necessary, since some have redundant or complementary functions. Indeed, recognition of starch is mainly performed by the membrane associated SusC and SusD proteins, providing 60% of the total starch binding affinity, the rest being provided by additional recognition mediated by SusE and SusF (Shipman et al. 2000a). After binding to the membrane, the substrate is roughly deconstructed by action of SusG, an endo α-amylase, yielding malto-oligosaccharides that are translocated to the periplasmic space *via* SusC. Final deconstruction of the malto-oligosaccharides is performed in the periplasm by the action of SusA and SusB to yield glucose units that finally are imported in the cytoplasm. The correct expression of the proteins of the Sus system is finely controlled by SusR, a regulatory protein, which senses the periplasmic space for malto-oligosaccharides, leading to the transcription of the required genes only when starch is present. It has been shown that *Bacteroidetes* species, and BT in

particular, possess multiple PULs dedicated to the assimilation of various oligo- and polysaccharides (Martens et al. 2009b). Hence, the BT CAZyme repertoire is grouped in 88 PULs, encompassing 18% of its genome (Martens et al. 2008b).

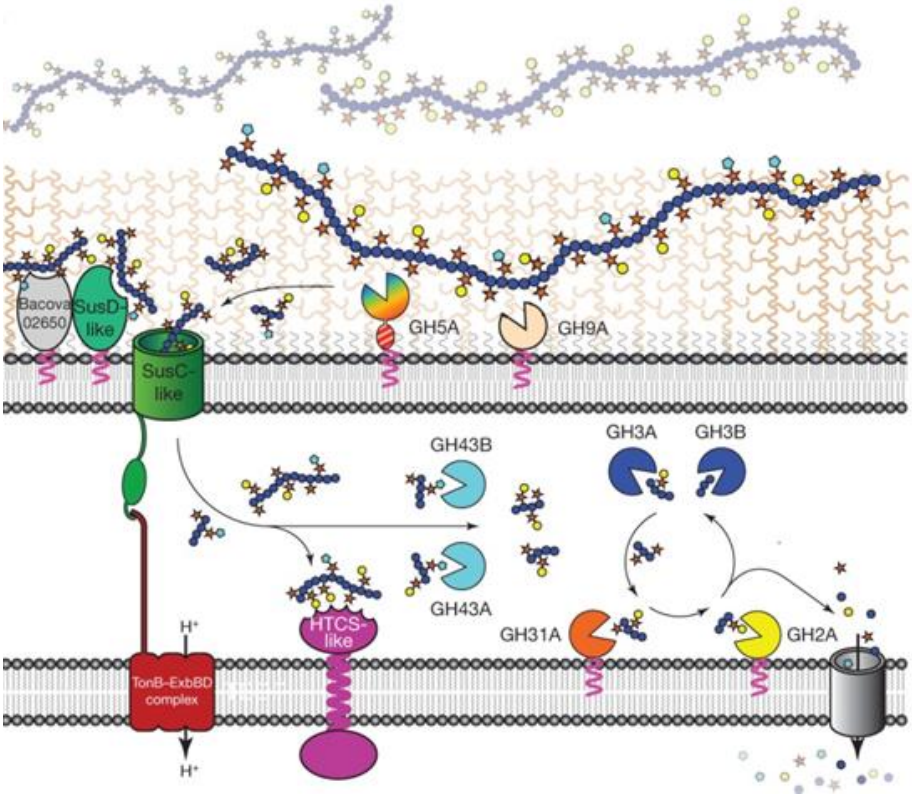


Figure 9: *Bacteroidetes* polysaccharide utilisation system.

Exemple of the xyloglucan degradation pathway of *Bacteroides ovatus* ATCC 8483. The xyloglucan oligosaccharide is made of a main chain composed of β-1,4 linked glucose units (blue spheres), extensively ramified with α-1,6-linked xylosyl units (red stars), β-1,2 galactosyl (yellow spheres) and α-1,2 L-fucosyl units (pale blue pentagons). Binding of the raw substrate and initial degradation steps are ensured by the membrane anchored Bacova_02650, GH5A, GH9A, SusC and SusD-like proteins, allowing translocation into the periplasmic space of shorter xylo-oligosaccharides. The final deconstruction is ensured by the synergetic action of additional GH43A/B, GH31, GH2A and GH3A/B. Monosaccharides are then translocated to the cytosol through a final transporter. Regulation of the expression on the enzymes involved in this complex multi-enzymatic system is ensured by a hybrid two component system (HTCS)-like regulatory transmembrane protein. From J. Larsbrink et al, Nature, 2014.

Therefore, based on functional genomic data targeting of 8 sequenced *Bacteroidetes* genomes, a general mechanism for polysaccharide metabolization, involving the five main steps i) recognition, ii) cleavage initiation, iii) oligosaccharide sequestration in the periplasm, iv) PUL expression regulation and v) final depolymerization has been proposed (fig. 10). A fully-automated approach for PUL prediction in *Bacteroidetes* or genomes has been recently developed and released (Terrapon and Henrissat 2014a) (fig. 11). However, as mentioned above, this automatic-method is still restricted to *Bacteroidetes*, as specific PUL signatures elements have been experimentally demonstrated for *Bacteroides thetaiotaomicron* VPI-

5482 and *Bacteroides ovatus* ATCC 8483 (Martens et al. 2011; Cameron et al. 2012a; Larsbrink et al. 2014; Cuskin et al. 2015a).

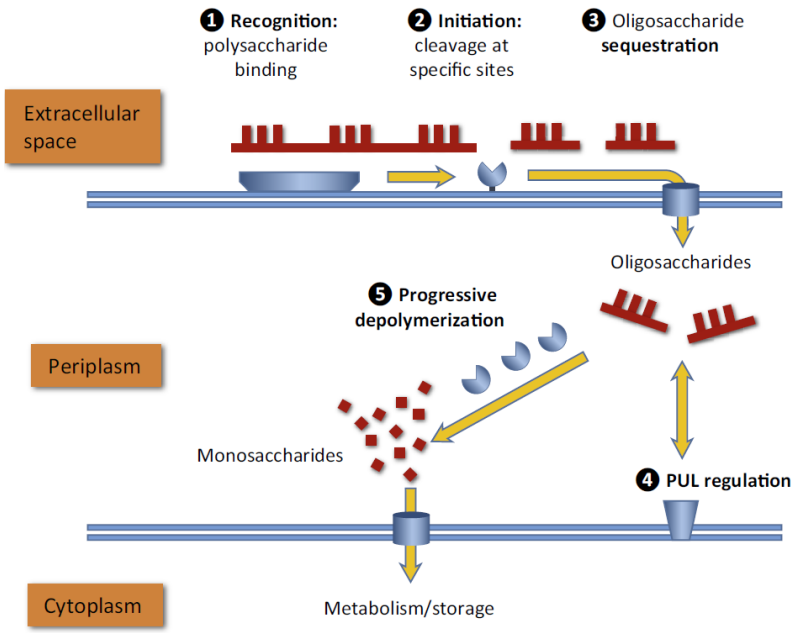


Figure 10: General *Bacteroidetes* Sus-like degradation system.
This figure shows the general mechanism used by *Bacteroidetes* for polysaccharide harvesting. They share the five steps of regulation, recognition, cleavage initiation, translocation and final deconstruction found in the starch utilization system of *Bacteroides thetaiotaomicron* VPI-5482 (from Terrapon N. and Henrissat B., Trends in Biochemical Sciences, 2014).

Results of search for PULs containing GH130 in all species

Experimentally validated PULs

Species	N°	Modularity	Overlap
Bacteroides ovatus ATCC 8483	Experimentally validated 42	HTCS SusC SusD unk GH26 GH26 GH36 **GH130** unk unk unk	Reverse Pred 44
Bacteroides ovatus ATCC 8483	Experimentally validated 44	unk HTCS GH43 unk unk unk unk unk unk SusC SusD GH115 unk GH105 unk GH43 unk SusC SusD unk unk unk GH2 **GH130** GH2 GH105 GH105 unk unk	Reverse Pred 45
Bacteroides ovatus ATCC 8483	Experimentally validated 46	GH5_2 SusC SusD unk GH5_5 GH2 GH5_5 unk **GH130** unk GH78 unk unk unk	Reverse Pred 47
Bacteroides ovatus ATCC 8483	Experimentally validated 58	**GH130** unk unk SusC SusD unk HTCS	Reverse Pred 57
Bacteroides ovatus ATCC 8483	Experimentally validated 82	GH92 **GH130** unk GH125 GH76 unk GH92 HTCS unk unk GH76 SusC SusD unk	Reverse Pred 79
Bacteroides ovatus ATCC 8483	Experimentally validated 85	ECF-σ Anti-σ SusC unk unk unk GH18 SusD SusC Int unk unk SusC SusD GH18 unk unk unk unk **GH130** GH92	Reverse Pred 82
Bacteroides thetaiotaomicron VPI-5482	Experimentally validated 14	GH92 **GH130** unk CBM32 unk unk GH18 SusD SusC Int SusC SusD GH18 unk SusC SusD GH18 unk unk GH20 Anti-σ ECF-σ	Reverse Pred 13
Bacteroides thetaiotaomicron VPI-5482	Experimentally validated 68	GH92 GH38 GT32 GT32 unk unk unk **GH130** GH125 GH76 GH92 HTCS unk SusC SusD unk unk GH76	Reverse Pred 64
Bacteroides thetaiotaomicron VPI-5482	Experimentally validated 74	SusR ECF-σ unk GH38 GH92 CBM32 GH2 unk GH78 CE1 unk unk unk SusC SusD unk unk SusD unk unk SusC SusD SusC unk GH92 GH92 **GH130** GH43 unk	Reverse Pred 69

Figure 11: Screenshot of a search from web interface of the CAZy PUL browser.
The query was the CAZy GH130 family, and the results presented here are only partial, as other putative PULs containing GH130 encoding genes have been detected in other genomes.

This tool specifically relies on the central roles of the SusC and SusD proteins in the Sus system. Identification of SusC/D homologs allows identifying possible PULs. However, this is only a starting point, as PULs also encode for catalytic enzymes and regulatory proteins whose automatic annotation can be imprecise. The method used is also based on the identification of putative operons, which means identifying genetic markers and features of co-expressed adjacent protein-encoding genes. To determine which gene is considered as 'adjacent' to another one, using the intergenic distance concomitantly with other genetic markers has been demonstrated to be the most powerful tool for operon prediction in *E.coli* and *B. Tethaiotaomicron* (Westover et al. 2005). Therefore PUL annotation is relying on the identification of CAZymes within a specific maximum distance from the predicted SusC/D homologs, on the same DNA strand, with no interruption by any gene on the other strand. *Stricto sensus*, this structure is called a directon (Salgado et al. 2000). However, sometimes, one or more genes on the opposite strand can insert into the PUL structure (which can also appear as another PUL inserting in a previous one). Such complex structures may interfere with the software and cause prediction problems, which appear as overlapping regions between two or more predicted PULs. In addition, PUL genetic signatures may vary from strain to strain. In order to allow a correct prediction in all cases, a confidence level colour system has been integrated in the JBrowse engine to guide the user.

Today, the PUL database (http://www.cazy.org/PULDB/) describes experimentally characterized PULs in the literature and also 3,833 predicted PULs in 69 *Bacteroidetes* species. Among them, not less than 43 and 47 PULs have been identified and experimentaly assigned to the catabolism of plant and host glycans, respectively, in three major *Bacteroidetes* species:

- *B. thetaiotaomicron* VPI-5482 contains 88 experimentaly identified PULs, together with 85 predicted ones (Martens et al. 2008b; Cuskin et al. 2015a). Among them, 43 were demonstrated to be dedicated to metabolisation of host glycans and 16 to plant glycans (Martens et al. 2011), while 3 were specifically active on yeast α-mannan (Cuskin et al. 2015b). It was in addition demonstrated to be deficient in PULs targeting hemicelluloses (Martens et al. 2011).
- *B. ovatus* ATCC 8483 contains 112 experimentaly identified PULs, along 106 predicted ones (Martens et al. 2011). Among them, 17 were demonstrated to be up-regulated when cultivated on plant glycans. In addition, the xyloglucan PUL has been completey characterized (Larsbrink et al. 2014).
- *B. cellulosilyticus* WH2 contains 111 experimentaly identified PULs, even if this organism is not referenced in the CAZy PULdb (McNulty et al. 2013). Twenty were demonstrated to be involved in cell wall degradation (8 for hemicelluloses, 6 for starches/fructans, and 6 for pectins), and 4 are dedicated to host glycans degradation (Sonnenburg et al. 2010a).

Therefore, many PULs belonging to these organisms have been assigned to specific functions. However, many were only revealed recently by transcriptomics, and much work remain to be conducted in order to finely understand their functioning, since characterization and *in vitro* enzymatic studies require much more effort than sequencing.

A manual detection of putative PULs has been used in this thesis in collaboration with Bernard Henrissat and his AFMB group prior to the release of the automated tool, in order to identify conserved genetic clusters which could be involved in mannoside harvesting by gut bacteria. These new pathways, described in the 'Results' section (chapter 4), involve real glycoside hydrolases, and also glycoside phosphorylases (GPs) classified in the GH130 family targeted in this PhD project.

2.3. Glycoside Phosphorylases

Glycoside Phosphorylases (GPs) are a fascinating class of CAZymes, because they efficiently combine both advantages of GHs and GTs. Similarly to GHs, they are involved in the degradation of glycosidic linkages, but instead of transferring the non-reducing end carbohydrate onto water (hydrolysis), they transfer it onto inorganic phosphate, yielding a glycosyl-phosphate (phosphorolysis) and a glycoside of reduced chain length (fig. 12). Also, similarly to GTs, they are able to use the glycosyl-phosphate as donor to transfer the glycosyl residue onto an hydroxylated acceptor (Kim et al. 2002). However, contrarily to both GHs and GTs, whose reaction equilibrium is shifted towards the glycosidic bond hydrolysis or synthesis, the essential characteristic of GPs is the reversibility of the reaction, since the free energy of glycosidic linkage between carbohydrate monomers is approximately the same as that of ester linkage in glycosyl phosphates. The ability of GPs to reuse the previously produced glycosyl-phosphate into a synthetic reaction is energy saving, and makes them valuable tools for glycoside synthesis. Indeed, the major drawback of GTs is that they use expensive activated nucleotide-sugars as donors to perform the synthesis reaction. Therefore, GPs combine the efficiency of the strict regio- and stereo-specificity of GTs to produce high added-value oligosaccharides, and the ability of GHs to degrade cheap plant biomass derived glycans (Nakai et al. 2010).

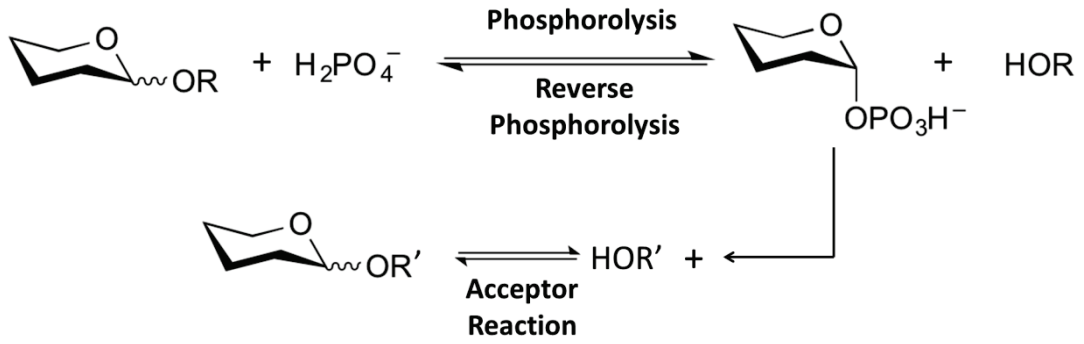


Figure 12: Overview of Glycoside phosphorylase reactions.
Similarly to GHs, GPs catalyze the degradation of glycoside linkages, but do not require a water molecule. Instead, they transfer the carbohydrate moiety onto inorganic phosphate (phosphorolysis) and yield a glycosyl-phosphate plus a glycoside of reduced chain length. The other important feature of GPs catalysis is the reversibility of the reaction, due to an equilibrium constant of same magnitude in both directions, the reverse reaction being called 'reverse phosphorolysis' or 'synthetic reaction'. The generated glycosyl-phosphate can be used to perform an 'acceptor reaction' to produce other compounds.

2.3.1. GP catalytic mechanisms

This unique ability to perform phosphorolysis (degradation) and reverse-phosphorolysis (synthesis) reactions relies on shared structural and mechanistic features with GHs and GTs. Also, as for GHs and GTs, GPs display two reaction mechanisms, with either retention or inversion of configuration at the anomeric centre. The main catalytic mechanism used by inverting GPs is very close to the one of inverting GHs, while the one used by retaining GPs is close to the one of retaining GTs (even if to date no retaining GPs have been found to act on β-linked glycosides) (fig. 13).

The catalytic mechanism used by most of the characterised inverting GPs only requires one catalytic amino acid and proceeds through a S_n2-like (i.e. a bimolecular nucleophilic substitution), one step mechanism. This catalytic amino acid is playing the role of proton donor or acceptor if one is considering phosphorolysis or reverse-phosphorolysis, respectively, while the nucleophile role is played by the inorganic phosphate itself. Indeed, the phosphorolysis is initiated by a nucleophilic attack of the inorganic phosphate to the C_1 of the sugar at the non-reductive end, concomitantly to the proton capture by the inter-osidic oxygen on the proton donor. This leads to the formation of an oxocarbenium-ion transition state, stabilized by the formation of a partial double bond between the C_1 of the sugar at the non-reductive end and its ring oxygen by delocalisation of an electronic lone pair of the oxygen. Then, bond breaks between the two glycosyl residues with the inter-osidic oxygen being held on the monomer at the reductive-end, which yields a glycosyl-phosphate of inverted configuration and an oligosaccharide of reduced chain length (fig. 13a). The existence of a unique non-canonical catalytic mechanism for an inverting GP

belonging to the family GH130 has been proposed in 2013, as mentioned in the section 2.3.4 dedicated to GH130 enzymes.

The catalytic mechanism for retaining glycoside phosphorylases is involving, such as for retaining GHs, the formation of a glycosyl-enzyme intermediate, and proceeds through a two steps mechanism. First, the nucleophile attacks the C_1 of the sugar at the non-reductive end, concomitantly to the proton capture by the inter-osidic oxygen on the proton donor, just as for inverting GPs. However in that case the nucleophile is not the phosphate itself, it is a catalytic amino acid. The same transition state is observed, which leads to the break of the inter-osidic oxygen-C_1 bond, and yields in that case a glycosyl-enzyme intermediate plus a glycoside of reduced chain length that may leave the catalytic site. Then, the proton donor recovers its proton by activating the inorganic phosphate which is involved in a concomitant nucleophilic attack onto the C_1 of the glycosyl-enzyme intermediate. The formed transition state is closely related to the one formed in the first step, stabilized by the delocalisation of an electronic lone pair of the ring oxygen. The final step is the break of the glycosyl-enzyme bond, yielding a glycoside-phosphate of retained configuration (fig. 13b).

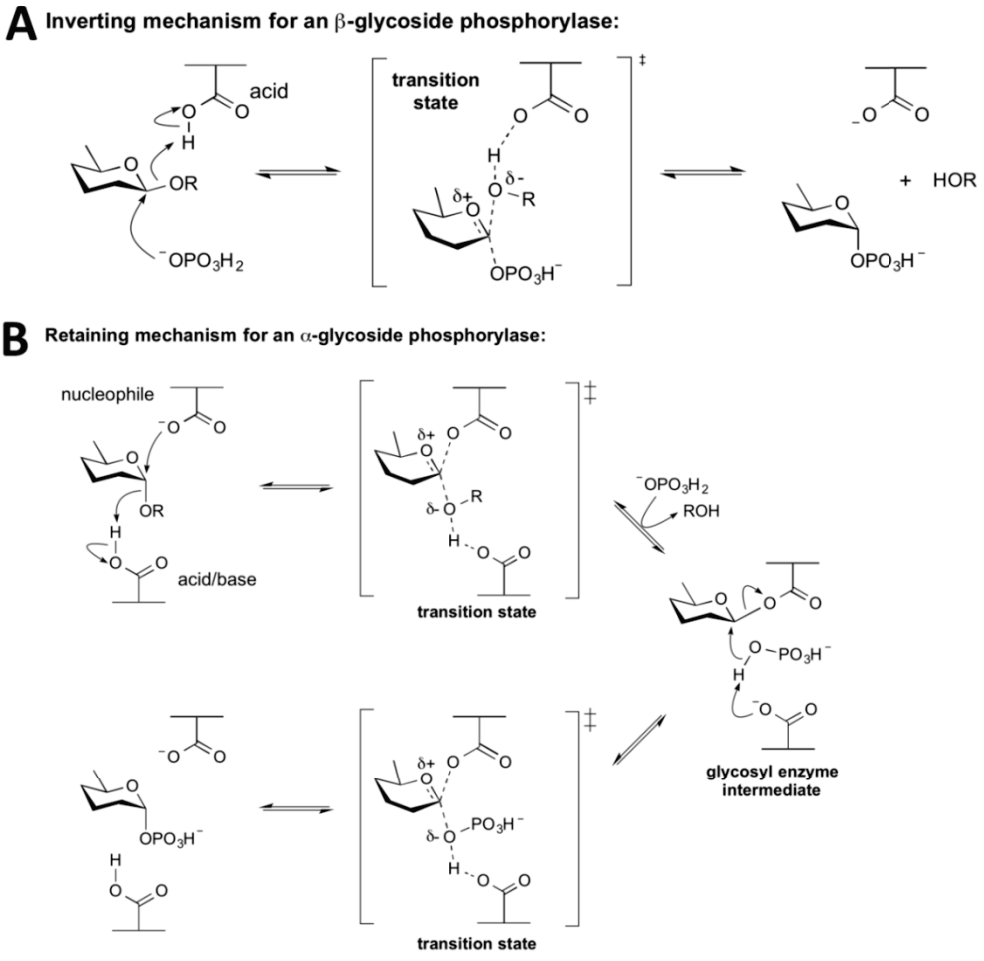


Figure 13: Detailed catalytic mechanisms proposed for inverting and retaining GPs.

Most inverting GPs are acting through a S_n2-like, one step mechanism, involving only one catalytic amino acid, playing the role of proton donor/acceptor. Most retaining GPs on the contrary, are using a two steps mechanism, requiring the two catalytic amino acids, to allow the formation of a glycosyl-enzyme intermediate, just as for the classical mechanism of retaining GHs.

Some evidences tend to show that a second mechanism, called 'direct front side nucleophilic displacement', or 'internal return', exists for retaining GPs, in which only one catalytic amino acid is required (Goedl and Nidetzky 2009). Indeed, in that case, only the nucleophile is required since the role of proton donor is played by the inorganic phosphate itself (fig. 14). First, one phosphate hydroxyl is providing the proton caught by the glycosidic oxygen, which is accompanied by a nucleophilic attack on the reductive-end sugar C_1 by the same oxygen providing the proton. It is important to note that both the protonation and the nucleophilic attack occur on the same side of the sugar ring to ensure the retention of configuration. The oxocarbenium-ion transition state, in which the four atoms C_4-O_5-C_1-C_2 are planar and the positive charge delocalized on the partial double bond between the O_5 and the C_1, is stabilized by the nucleophile which is in that case a glutamine or an asparagine (Goedl and Nidetzky 2009).

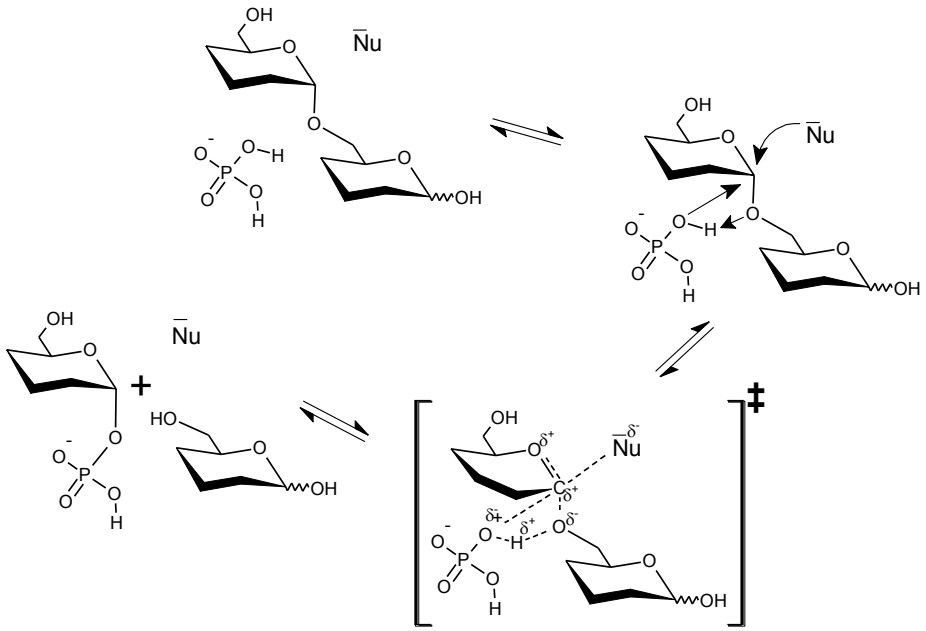


Figure 14: Alternative catalytic mechanism adopted by some retaining GPs.
This mechanism called direct front side nucleophilic displacement is unique in that it only requires one catalytic amino acid, the nucleophile, as the proton donor role is played by the inorganic phosphate itself.

Regarding the kinetics of the reaction, here also, two main mechanisms are harboured by GPs. Retaining GPs for which the formation of a glycosyl-enzyme intermediate is required proceed through an ordered substrate binding order mechanism. Following a ping-pong mechanism, the oligosaccharide binding and the nucleophilic attack are the first steps for phosphorolysis, while the glycosyl-phosphate binding is the first step for reverse-phosphorolysis. Retaining enzymes using a direct front side nucleophilic replacement

constitute an interesting case. Indeed, even if the catalytic mechanism is involving a single step, ordered substrate binding is sometimes observed, as for the *Schizophyllum commune* trehalose phosphorylases (Goedl et al. 2006).

On the contrary, the one step, S_n2-like mechanism used by inverting GPs requires the simultaneous interaction of the two substrates with the enzyme, and leads to the subsequent release of the two products after the reaction. In that case, the order of substrate binding and product release in the catalytic site has no impact on the reaction kinetics, which explain why inverting phosphorylases use a mechanism with formation of a ternary complex which is formed randomly, or use an ordered substrate binding order, or a mixed fashion in which the substrates binding is done at random while the products release is ordered, or the contrary (Nidetzky et al. 2000; Hamura et al. 2012; Nihira et al. 2012; Kawahara et al. 2012a).

2.3.2. GP classification and substrate specificity

As described above, it appears that GPs are no more different from GHs than GTs from a mechanical as well as structural point of view, which explains that no specific sequence signature allows for GP prediction from their sequence, neither allows to group them into a specific family. Logically, identifying a GP appears to be a matter of luck, which may explain why so few data on these particularly interesting enzymes have been gathered so far. Therefore, GPs are listed in both GHs and GTs families, namely GH3, GH13, GH65, GH94, GH112, GT4, GT35 and GH130 (Table 1).

All characterised GPs belonging to GT families are acting with a retaining catalytic mechanism and are part of the GT35 and GT4 families. Representatives of both families are acting on α-glucosides, namely glycogen/starch or trehalose, respectively. It is to note that only two GH families contains retaining GPs. GH13 GPs also act on α-glucans and sucrose, yielding glucose-α-1-phosphate. The sole characterized GP of the GH3 family, the *N*-acetylglucosamine phosphorylase Nag3 from *Cellulomonas fimi*, was recently identified as a phosphorylase because it is able to perform both phosphorolysis and hydrolysis (Macdonald et al. 2014). This enzyme is involved in peptidoglycan recycling, degrading the core β-D-GlcNAc-1,4-MurNAc disaccharide to yield β-D-Glc*p*-1-phosphate. This enzyme, part of a subset of GH3 sequences, is specific in that it is using as conserved His-Asp catalytic diad as acid-base residue which favours anionic phosphate binding rather than water.

All the other GPs belonging to GH families harbour a catalytic mechanism yielding an inversion of configuration on the released glycosyl-phosphate. Those acting on glucosides, α- and β-linked, share a common GH-L fold but are separated in the two GH65 and GH94 families, respectively. All are disaccharide phosphorylases, the GH65 acting on maltose,

trehalose and other isomers of α-linked di-glucosyl units, while those belonging to the GH94 family act on β-linked di-glucosyl units, namely cellobiose, laminaribiose and cellodextrins. A few exceptions concern the enzymes acting on modified glycosyl units, namely the GH94 chitobiose phosphorylase (Hidaka et al. 2004) and the cellobionic acid phosphorylases (Nihira et al. 2013a), the β-1,2-oligo-D-glucan phosphorylase which is the sole GH94 acting on a polymer (Nakajima et al. 2014), and finally the GH65 glucopyranosyl glycerol phosphorylase which is the unique example of natural phosphorylase which does not process a disaccharide (Nihira et al. 2014). Recently, two other GPs have been characterized, the 1,3-β-D-glucan phosphorylase from *Ochromonas danica* and the GH94 laminaribiose phosphorylase, two enzymes acting on 1,3-linked β-D-glucosyl units which differ only on their ability to degrade long substrates or not, respectively (Nihira et al. 2012; Yamamoto et al. 2013). The other characterized inverting GPs are the GH112 phosphorylases acting on β-galactosides and the GH130 β-mannoside phosphorylases. This thesis focuses on this latter family of enzymes involved in metabolization of β-1,4-Manno-oligosaccharides, β-1,4 mannan, β-D-Mannopyranosyl-1,4-D-glucose, β-D-Mannopyranosyl-1,4-*N*-acetyl-D-glucosamine or β-D-Mannopyranosyl-1,4-*N*,*N*'-diacetylchitobiose, other associated β-linked mannosides, and of β-linked galactosides. More about this specific topic will be discussed further in this bibliographic introduction and in the 'Results' section.

Table1: Knowledge on glycoside-phosphorylases, released in January 2015.

The creation of the GH130 subfamilies was carried out in the frame of this PhD. The activities of the GH130_2 β-D-Mannopyranosyl-1,4-N-glycan and mannan-phosphorylase that were determined in the frame of the present thesis is more precisely described in chapter 4. ND: not determined.

EC number	Enzyme	Substrate	Product	Mechanism	Fold	Clan	Family	PDB
2.4.1.231	Trehalose phosphorylase	α-D-Glc*p*-1,1-α-D-Glc*p*	α-D-Glc*p*-1-P + D-Glc	Retaining	GT-B		GT4	0
2.4.1.1	Glycogen/starch phosphorylase	(α-D-Glc*p*-1,4-)$_n$-D-Glc*p*	α-D-Glc*p*-1-P + (α-D-Glc*p*-1,4-)$_n$-1-D-Glc*p*	Retaining	GT-B		GT35	8
2.4.1.-	β-*N*-acetylglucosamine phosphorylase	β-D-GlcNAc-1,4-MurNAc	β-D-GlcNac-1-P + MurNAc	Retaining			GH3	0
2.4.1.7	Sucrose phosphorylase	β-D-Fru*f*-2,1-α-D-Glc*p*	α-D-Glc*p*-1-P + D-Fru	Retaining	(β/α)8	GH-H	GH13_18	1
2.4.1.-	Sucrose 6'-phosphate phosphorylase	6-P-β-D-Fru*f*-2,1-α-D-Glc*p*	α-D-Glc*p*-1-P + 6-P-D-Fru	Retaining	(β/α)8	GH-H	GH13	0
2.4.99.16	α-Maltosyl phosphate:α-1,4-D-glucan 4-α-D-maltosyltransferase	(α-D-Glc*p*-1,4-)$_n$-D-Glc*p*	α-D-Glc*p*-1,4-α-D-Glc*p*-1-P + (α-D-Glc*p*-1,4)$_{n-2}$-D-Glc*p*	Retaining	(β/α)8	GH-H	GH13_3	2
2.4.1.8	Maltose phosphorylase	α-D-Glc*p*-1,4-D-Glc*p*	β-D-Glc*p*-1-P + D-Glc	Inverting	(α/α)6	GH-L	GH65	1
2.4.1.64	Trehalose phosphorylase	α-D-Glc*p*-1,1-α-D-Glc*p*	β-D-Glc*p*-1-P + D-Glc	Inverting	(α/α)6	GH-L	GH65	0
2.4.1.216	Trehalose 6-phosphate phosphorylase	6-P-α-D-Glc*p*-1,1-α-D-Glc*p*	β-D-Glc*p*-1-P + 6-P-D-Glc	Inverting	(α/α)6	GH-L	GH65	0
2.4.1.230	Kojibiose phosphorylase	α-D-Glc*p*-1,2-D-Glc*p*	β-D-Glc*p*-1-P + D-Glc	Inverting	(α/α)6	GH-L	GH65	1
2.4.1.279	Nigerose phosphorylase	α-D-Glc*p*-1,3-D-Glc*p*	β-D-Glc*p*-1-P + D-Glc	Inverting	(α/α)6	GH-L	GH65	0
2.4.1.282	α-D-Glucopyranosyl-1,3-L-rhamnose phosphorylase	α-D-Glc*p*-1,3-L-Rha*p*	β-D-Glcp-1-P + L-Rha	Inverting	(α/α)6	GH-L	GH65	0
2.4.1.-	2-O-α-D-Glucopyranosylglycerol phosphorylase	2-O-(α-D-Glc*p*)-glycerol	β-D-Glc*p*-1-P + glycerol	Inverting	(α/α)6	GH-L	GH65	1
2.4.1.20	Cellobiose phosphorylase	β-D-Glc*p*-1,4-D-Glc*p*	α-D-Glc*p*-1-P + D-Glc	Inverting	(α/α)6	GH-L	GH94	3
2.4.1.31	Laminaribiose phosphorylase	β-D-Glc*p*-1,3-D-Glc*p*	α-D-Glc*p*-1-P + D-Glc	Inverting	(α/α)6	GH-L	GH94	0
2.4.1.49	Cellodextrin phosphorylase	(β-D-Glc*p*-1,4-)$_n$-D-Glc*p*	α-D-Glc*p*-1-P + (β-D-	Inverting	(α/α)6	GH-L	GH94	0

			Glc*p*-1,4-)$_{n-1}$-D-Glc*p*					
2.4.1.280	*N*,*N*'-Diacetylchitobiose phosphorylase	β-D-GlcNAc-1,4-D-GlcNAc	α-D-GlcNAc-1-P + D-GlcNAc	Inverting	(α/α)6	GH-L	GH94	1
2.4.1.321	Cellobionic acid phosphorylase	β-D-Glc*p*-1,4-D-Glc*p*A	α-D-Glc*p*-1-P + D-Glc*p*A	Inverting	(α/α)6	GH-L	GH94	0
2.4.1.-	β-1,2-Oligo-D-glucan phosphorylase	(β-D-Glc*p*-1,2-)$_n$-D-Glc*p*	α-D-Glc*p*-1-P + (β-D-Glc*p*-1,2-)$_{n-1}$-D-Glc*p*	Inverting	(α/α)6	GH-L	GH94	0
2.4.1.211	β-D-Galactopyranosyl-1,3-*N*-acetylhesoamine (galacto-*N*-biose/lacto-*N*-biose) phosphorylase	β-D-Gal*p*-1,3-D-Gal*p*/Glc*p*	α-D-Glcp-1-P + D-Gal/Glc	Inverting	(β/α)8		GH112	1
2.4.1.247	β-D-galactopyranosyl-1,4-L-rhamnose phosphorylase	β-D-Gal*p*-1,4-L-Rha	α-D-Gal*p*-1-P + L-Rha	Inverting	(β/α)8		GH112	0
2.4.1.281	β-D-Mannopyranosyl-1,4-D-glucose phosphorylase	β-D-Man*p*-1,4-D-Glc*p*	α-D-Man*p*-1-P + D-Glc	Inverting	5-bladded β-propeller		GH130_1	1
2.4.1.-	β-1,2-oligomannan phosphorylase	(β-D-Man*p*-1,2-)$_n$-D-Man*p*	α-D-Man*p*-1-P + (β-D-Man*p*-1,2-)$_{n-1}$-D-Man*p*	Inverting	5-bladded β-propeller		GH130_2	0
2.4.1.- 2.4.1.319 2.4.1.320	β-D-Mannopyranosyl-1,4-*N*-glycan phosphorylase and β-D-1,4-Mannan phosphorylase	β-D-Man*p*-1,4-GlcNAc-β-1,4-GlcNAc and (β-D-Man*p*-1,4-)$_n$-D-Man*p*	α-D-Man*p*-1-P + β-D-GlcNAc-1,4-D-GlcNAc and α-D-Man*p*-1-P + (β-D-Man*p*-1,4-)$_{n-1}$-D-Man*p*	Inverting	5-bladded β-propeller		GH130_2	0
2.4.1.97	β-1,3-D-Glucan phosphorylase	(β-D-Glc*p*-1,3-)$_n$-D-Glc*p*	α-D-Glc*p*-1-P + (β-D-Glc*p*-1,3-)$_{n-1}$-D-Glc*p*	Inverting			ND	0
2.4.1.30	β-1,3-Oligo-D-glucan (laminaridextrin) phosphorylase	(β-D-Glc*p*-1,3-)$_n$-D-Glc*p*	α-D-Glc*p*-1-P + (β-D-Glc*p*-1,3-)$_{n-1}$-D-Glc*p*	Inverting			ND	0

2.3.3. GP structures

Crystallographic structure of at least one representative of each family has been determined, with the exception of GT4 retaining threalose phosphorylase and GH3 β-*N*-acetylglucosamine phosphorylase for which no structure is yet available. Interestingly, most of crystallized forms of GPs are dimers. This is consistent with the state in solution of nearly all crystallized GPs and others for which only functional data is available. The single exception concerns GH130 GPs, which present monomeric crystallized forms and others with a higher degree of oligomery. Indeed, members of both subfamilies GH130_1 and GH130_2 crystallize as tetramers, pentamers or hexamers, these states being conserved in solution.

GPs crystallization in complex with their natural substrates allowed highlighting their catalytic mechanism, since both phosphorolysis and reverse phosphorolysis were observed *in cristallo* (Mirza et al. 2006; Nakae et al. 2013). The analysis of the available structures revealed in some cases that specific domains may be directly implicated in oligomerization. For instance, the case of the GH130 *Bf*MGP shows that its hexameric assembly is directly related to N- and C-terminal α-helices joining together the different protomers. The oligomeric state of GPs may also directly impact their substrate specificity. Examples include the *Bifidobacterium adolescensis* sucrose phosphorylase, for which the oligomeric assembly restrains the different protomers to fold in a specific shape. Another more drastic impact of oligomery on substrate specificity is depicted in the present work and will be described in detail in chapter 5.

Substrate recognition is generally mediated by stacking interactions due to the presence of residues which forms aromatic platforms, and also involves a dense network of hydrogen bonds and van der Walls interactions, providing a strong specificity on the selected carbohydrates. Phosphate recognition seems to be mediated by the most important interactions, since generally several arginines or others positively charged residues are providing strong ionic/polar interactions with the negative charges of oxygen, or with the non-charged oxygens. A very strong specificity is observed for the non-reducing end of the glycosyl donors (or for the glycosyl phosphate substrate), both in phosphorolysis and reverse-phosphorolysis reactions. This is due to a very specific recognition of the positioning of at least three of the hydroxyls held by the glycosyl moiety in the subsite -1, whether directly by specific residues or through a water molecule. Except for some enzymes like the GT35 maltodexin phosphorylase (Campagnolo et al. 2008) or the GH130_1 β-D-Mannopyranosyl-1,4-D-glucose phosphorylases (Senoura et al. 2011; Kawahara et al. 2012a; Jaito et al. 2014), generally a more relaxed specificity towards the reducing end of glycosides (in the phosphorolysis reaction) and towards acceptors (in the synthetic reaction) is observed, due to a more flexible accommodation in subsite +1 and the following subsites. In

these cases, flexibility is tolerated because substrate variations imply replacement of one interaction by another one, or because accommodation occurs on a hydroxyl not mediating the binding.

This is well exemplified by the GH130_2 mannoside-phosphorylases, as further described in the 'Results' section of this manuscript.

2.3.4. Focus onto the Glycoside Hydrolase 130 family

The GH130 family was created in 2011, consecutively to the functional characterization of its first member, the β-D-Mannopyranosyl-1,4-D-glucose (mannosyl-glucose) phosphorylase from *Bacteroides fragilis* NCTC 9343 (*Bf*MGP) (Senoura et al. 2011). This is the first example of CAZy family creation based on the characterisation of a phosphorylase, which are not as frequent as real GHs or GTs. In addition, surprisingly, this family contained more 3D structures available than functionally characterized representatives when created. Indeed, four 3D structures of enzymes belonging to this family were deposited in the protein data bank (PDB) before its creation. These 3D structures were all determined by a structural genomic consortium, the Joint Center for Structural Genomics (JCSG) as early as 2004. These are the 3D structure of the *Tm*1225 protein from *Thermotoga maritima* MSB8 (PDB code 1VKD), the BACOVA_03624 protein from *Bacteroides ovatus* ATCC 8483 (3QC2), the BT4094 protein from *B. thetaiotaomicron* VPI-5482 (3R67), and the BDI_3141 protein from *Parabacteroides distasonis* ATCC 8503 (3TAW). All these proteins crystallized as monomers except for *Tm*1225 which crystallized as a dimer. However, none of these structures was joined to either a publication or any functional characterization. That was the picture of the family when this thesis started, in october 2011: only one functionally characterized member and four 3D structures with no associated function.

In 2012, a Japanese group characterized two new GH130 enzymes (Kawahara et al. 2012a). These two enzymes, *Ra*MP1 and *Ra*MP2, originate from a ruminal bacterium, *Ruminococcus albus* 7. Similarly to *Bf*MGP, *Ra*MP1 catalyses the conversion of β-D-Man*p*-1,4-D-Glc*p* and inorganic phosphate (P_i) into α-D-Man*p*-1-phosphate and D-Glucose. Reverse phosphorolysis reaction attempts also showed that both *Bf*MGP and *Ra*MP1 share a very narrow flexibility since only α-D-mannose-1-phosphate is tolerated in the -1 subsite, while only glucose is accepted in the +1 (a very low activity on laminaribiose was however observed for *Ra*MP1 exclusively, without any activity on cellobiose). The authors deduced from the genomic context of *Ra*MP1 and *Ra*MP2 encoding genes that they are involved in a mannan/glucomannan degradation pathway, as further described in chapter 3.

In 2013, the GH130 family was already containing more than 300 sequences (369 late January), and two additional members were concomitantly functionally characterized: the β-D-mannopyranosyl-1,4-*N*-acetyl-D-glucosamine phosphorylases *Bt*1033 from *B. thetaiotaomicron* VPI-5482 (Nihira et al. 2013b), published during the 2nd review of the first paper presented in the 'Results' section of this manuscript), and the β-D-mannopyranosyl-1,4-*N*-glycan phosphorylase Uhgb_MP from an uncultured gut bacterium (Ladevèze et al. 2013), which constitutes the main target of this thesis project. The functional characterization of Uhgb_MP is detailed in the first chapter of the 'results' section of this manuscript. Both enzymes appear to be very similar to one another. Indeed, as the other GH130 characterized enzymes, the -1 subsite is highly specific for α-D-mannose-linked mannosyl residue. Phosphorolysis reaction assays showed that the preferred substrate of *Bt*1033 and Uhgb_MP was not β-D-Man*p*-1,4-D-Glc*p* but β-D-Man*p*-1,4-GlcNAc. Together with the analysis of *Bt*1033 and Uhgb_MP genomic environments, Nihira *et al.* and we hypothesized a functional role *in vivo* different to the one of the others previously characterized GH130 enzymes, namely *Bf*MGP, *Ra*MP1 and *Ra*MP2. Taking also into account the fact that the micro-organism producing this intracellular enzyme originates from the human gut, collectively with Nihira *et al*, we deduced that such a disaccharide is likely to be derived from the degradation of host N-glycans lining the intestinal epithelium. This N-glycan degradation pathway will be described in details in Chapter 3.

Late 2013, the crystal structure of *Bf*MGP was determined in the *apo* form (4KMI) and in complex with its different substrates (3WAT, 3WAS, 3WAU) (Nakae et al. 2013). An unusual hexameric structure was shown, and the analysis of the α-D-mannose-1-phosphate binding mode, as well as the direct *in cristallo* observation of the reverse phosphorolysis reaction allowed the authors to propose the existence of a non-canonical catalytic mechanism for inverting GPs (fig. 15). Indeed, the structure revealed that neither the distance between the catalytic proton donor and the interosidic oxygen was suitable for direct proton transfer, nor a water molecule that could relay it was observed. Consequently, the authors proposed an alternative mechanism, similar to the canonical mechanism for inverting GPs, but in which the proton is relayed through the mannose C_3 hydroxyl, which is at H-bond distance of both the catalytic aspartic acid and the interosidic oxygen atom (fig. 15). Such a carbohydrate hydroxyl mediated proton transfer has been proposed for other enzymes, including ribozyme (Filling et al. 2002; Schmeing et al. 2005), and a closely related mechanism involving a C_2-OH neighbouring group participation has been proposed for the GH99 endo-α-mannosidase from *B. thetaiotaomicron* VPI-5482, yielding a 1,2 anhydro-sugar as transition state (Thompson et al. 2012). However, the mechanism harboured by the latter case is not suitable for *Bf*MGP because it is an exo-acting inverting β-mannoside phosphorylase, for which the reaction occurs in one step, which means a direct attack of the nucleophile onto the anomeric carbon. On the contrary, the involvement of the C_3 hydroxyl for proton relay is supported by the specific position of the mannose hydroxyls in the

adopted boat $B_{2,5}$ transition state conformation. It is less unstable in mannosides compared to other monosaccharides, due to the pseudo-equatorial position of the C_2-OH, in an *anti* configuration to the ring oxygen (Speciale et al. 2014), thus bending the C_3-OH towards the β-glycosidic oxygen, in a *syn*-axial position. This work, published concomitantly with our first study regarding Uhgb_MP and the identification of its most probable catalytic residue (Ladevèze et al. 2013), allowed one to answer the crucial question of the mechanism of the GH130 mannoside-phosphorylases.

Apart from the ones described in this manuscript, in 2014, 4 other GH130 enzymes were characterized. The BACOVA_02161 enzyme from *B. ovatus* ATCC 8483 has been crystalized, and its structure deposited in the PDB (4ONZ) by the JCSG. It is closely related to the other crystalized GH130 from *B. ovatus* ATCC 8483, BACOVA_03624, also crystalized as a monomer. As for BACOVA_03624, no associated paper was published upon release of this structure. *Rm*MGP, originating from *Rhodothermus marinus* (a hot spring bacterium belonging to *Bacteroidetes*) has been functionally characterized and demonstrated to be closely related to *Bf*MGP (Jaito et al. 2014). Sequences alignment showed a 68% and 61% sequence identity with *Bf*MGP and *Ra*MP1, respectively, which lead the authors to hypothesize that *Rm*MGP, *Ra*MP1 and *Bf*MGP are orthologs. The functional characterization of these 4 GH130 enzymes, including Uhgb_MP, highlighted a strong specificity of the -1 subsite for mannosyl residues, but a much more relaxed one for +1 and following subsites, allowing accommodation of a large range of acceptors for glycoside synthesis. A developed description of substrate specificity for phosphorolysis and reverse-phosphorolysis and demonstration that the natural substrate of Uhgb_MP is β-D-Mannopyranosyl-1,4-*N*,*N*'diacetylchitobiose can be found in the first chapter of the 'Results' section and structural elements of the recognition are presented in the second one. Functional characterization of the BACOVA_03624 protein which has its 3D structure solved is presented in the fourth one.

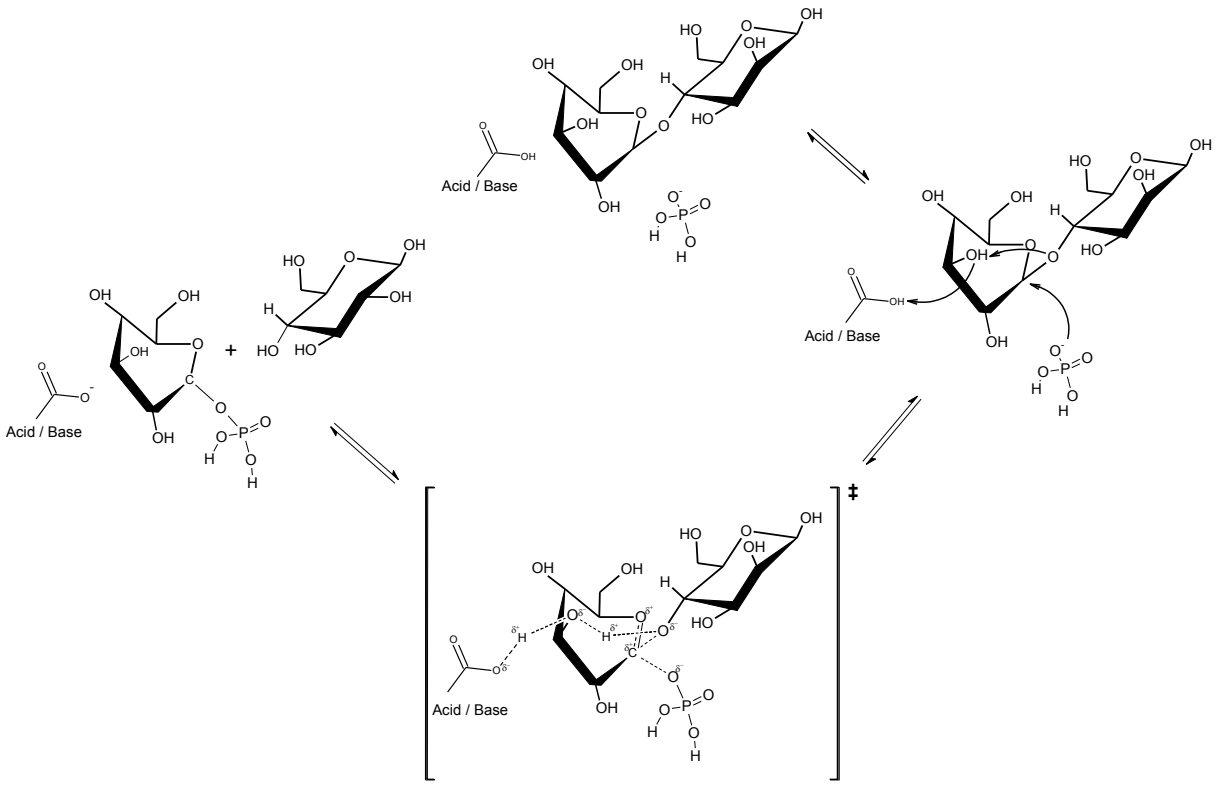


Figure 15: Alternative mechanism for inverting glycoside phosphorylases.
In this mechanism proposed for the GH130 *Bf*MGP mannoside phosphorylase, the proton donor is too far away from the interosidic oxygen atom to be able to directly interact with it. The mannose present in the -1 subsite has been observed in a stressed boat $B_{2,5}$ conformation The proton is relayed through the C_3 hydroxyl, which is located at H-bond distance of both.

Late 2014, two additional GH130 phosphorylases were functionally characterized (Chiku et al. 2014). Teth514_1788 and Teth514_1789 are two enzymes belonging to *Thermoanaerobacter* sp.514, a Firmicute isolated in a deep subsurface sample (-2000m). These enzymes belong to the pool of non-classified sequences (GH130_NC). Interestingly, these two enzymes are unique in that they are specific for β-1,2 linkages. As for the others GH130, their -1 subsite is highly specific for mannosyl residues. The analysis of their genetic environment revealed that they are part of a PUL containing a mannose-1-phosphate guanylyltransferase and a α-mannosyltransferase, which lead the authors to propose a glycan synthetic role for this PUL. Even if *Ra*MP1 is specific for the disaccharide β-D-man*p*-1,4-D-Glc, the similarity between *Ra*MP1/*Ra*MP2 and Teth514_1789/Teth514_1788 is striking. Indeed, Teth514_1788 has been shown to be specific for long manno-oligosaccharides while Teth514_1789's specificity is restricted to β-1,2-mannobiose. According to the authors, these two enzymes would be feeding the glycan synthesis pathway with mannose residues scavenged from β-1,2-mannosides. However, both enzymes bear the same acceptor profile as GH130_1 enzymes, being only restricted to fructose. This is not surprising, as fructose mostly appears under β-D-fructopyranose form, whose hydroxyls' orientations are identical to those of D-mannopyranose.

In total, in January 2015 the functional and structural characterization of 8 and 2 GH130 enzymes have been published, respectively (including that of Uhgb_MP, one of our targets). In addition, 5 *apo* 3D structures of 5 other GH130 enzymes have been released in the PDB. The last advances in the understanding of the *in vivo* role of these fascinating enzymes are described in Chapter 3.

2.3.5. Biotechnological applications of glycoside phosphorylases

Carbohydrate based industries are encompassing a wide array of fields. Biofuels, biosourced materials, functional food and feed, pharmaceuticals, chemical syntons, cosmetics or texturing agents are few examples of carbohydrate derived products. Chemists have developed and mastered a wide array of reactions to ensure an economically viable production of a large panel of structural motives to fit industry's requirements. However, these chemically based processes, even if being often highly powerful, are, for most of them, not environmentally friendly, and not sufficiently regiospecific. To address these problems, enzyme-based catalysis, which uses enzymes in place of chemicals, has progressively been slipping from assisting tools to the true mainstream process. Indeed, fundamental characteristics of biocatalysts make them strong tools to replace, or to complement classical organic chemistry. Their ability to specifically catalyse the desired reaction, both regio- and enantio-selectively, coupled with higher yield compared to those generally observed in classical chemistry, without the use of hazardous solvents, make biocatalysts key and 'green labelled' actors for carbohydrate industry. As sugars are basic molecules of all living organisms, life has evolved a wide diversity of CAZymes in which one can look for the desired catalyst.

Four enzymatic routes can be used for glycan synthesis, implicating GHs, GTs, and GPs. First, as in nature synthesis of complex oligosaccharides is devoted to Leloir-GTs, it is not surprising that they were at the basis of most of the first synthesis strategies, and are still largely used in carbohydrate chemistry (Chang et al. 2011). However, even if they are very specific and efficient catalysts, their major drawback is the requirement of expensive and unstable Leloir donors (sugar mono- or di-phosphonucleotides) as glycosyl donors, thus reducing their commercial use. Moreover, high yield polymer production based on the use of Leloir-GTs cannot be envisaged as these enzymes use too expensive activated substrates, even if smart living factories have been developed for oligosaccharide synthesis by GTs to circumvent this limitation (Bettler et al. 1999; Peng et al. 2012).

Secondly, the ability of some natural or engineered glycoside hydrolases to catalyse transglycosylation reactions (Monsan et al. 2010), i.e transfer of a glycosyl residue from a donor glycoside onto a hydroxylated acceptor carbohydrate (Champion et al. 2012) or hydroxylated chemical (Malbert et al. 2014) has been intensively studied. However, very few

donor specificities have been described for these particular enzymes, mainly sucrose acting enzymes, and rare are the examples of subsite -1 engineering to render them capable to act on other glycoside donors (Daudé et al. 2013).

Thirdly, glycoside synthesis can also be performed by glycoside hydrolases from chemically modified oligosaccharides, like glycoside-fluorides, through the glycosynthase-based technology (Jahn et al. 2003; Kittl and Withers 2010). Glycosynthases are mutated retaining glycoside hydrolases in which the catalytic nucleophile has been replaced by a smaller and inert residue, to render them capable of catalyzing glycoside bond formation with glycoside-fluoride donors (Cobucci-Ponzano et al. 2012). But this elegant approach cannot be directly used to transform low cost plant biomass derived resources, as it requires chemically modified glycosyl donors.

The fourth route, involving GPs, does not require expensive or chemically modified substrates, and is undoubtedly underexploited, due to the far reduced amount of characterized GP members compared to real GHs and GTs. Yet, some of them have been successfully used as powerful biocatalysts (O'Neill and Field 2014), directly using glycosyl-phosphates, which are more stable than the activated sugars used by Leloir GTs. Moreover, as the catalysed reaction is reversible, it is possible to use cheap oligosaccharides or polysaccharides to generate glycosyl-phosphates, using GPs of the same or different specificities. Presently, there is only a reduced direct application of these glycosyl phosphates. One can mention, the use of α-D-glucose-1-phoshate in parenteral nutrition mixtures, as replacement of phosphate source due to the precipitation of calcium with inorganic phosphate (Ronchera-Oms et al. 1995), or its use as dietary complement providing a vitamin A-like activity, in the particular case of age-related depression of intestinal function (Fujinaka et al. 2007).

On the contrary, the generated glycosyl-phosphates have a lot of applications as substrates for glycoside synthesis using GPs (Nakai et al. 2013; O'Neill and Field 2014). Several examples include synthesis of 2-*O*-α-D-glucosyl-glycerol, a promising moisturizing agent in cosmetics or lacto-*N*-biose, for which production achieved the kilogram scale (Nishimoto and Kitaoka 2007; Goedl et al. 2008). This is of particular importance since lacto-*N*-biose (β-D-Gal*p*-1,3-D-Glc*p*NAc) is a central motif of important oligosaccharides such as lacto-*N*-tetraose and lacto-*N*-fucopentaose, which are key constituents of milk oligosaccharides (Fushinobu 2010). Different GPs sharing the same specificity towards glycosyl-phosphates can be mixed to expand the array of synthesizable oligosaccharides (Suzuki et al. 2009). That way, the strict stereospecificity of GPs can be craftily exploited to reverse the linkages of a polymer for example, as showed in the work of Ohdan K. and colleagues who produced amylose from cellobiose (Ohdan et al. 2007), thus opening the way for valorisation of waste wood in mass production of such a valuable polysaccharide.

In all cases, the major drawback of glycan or glycosyl-phosphate synthesis is the reversibility of the catalysed reaction, which means that the production yield is limited by the equilibrium of the two opposite reactions. In order to address this problem, different strategies have been developed, which *in fine* conducts to a change in the equilibrium of the reaction.

The first strategy consists in depleting the reaction media from one of the products, which by the law of mass action, shifts the reaction towards the desired direction. It can be achieved easily if by any chance, one of the products is insoluble, such as cellobiose or cellulose (Nidetzky et al. 2004). If removal of one product cannot be achieved this way, another possibility is the use of a coupling reaction, which is, most of the time, involving another GP sharing the same specificity towards the glycosyl-phosphate, enabling the co-production of a high added value compound. A striking example is the production of galacto-*N*-biose (β-D-galactopyranosyl-1,3-D-*N*-acetylgalactosamine) from sucrose and *N*-acetylgalactosamine (Nishimoto and Kitaoka 2009).

In order to increase the yield of synthetic reactions catalysed by GPs, other strategies have been used, like replacing glycosyl phosphates by glycosyl fluorides. These last compounds resemble glycosyl phosphates, both not requiring catalytic acid assistance for a departure of these good leaving groups, phosphate or fluoride. The use of glycosyl-fluorides instead of glycosyl-phosphates conducted to excellent oligosaccharide production yields, since the reaction becomes irreversible, no phosphate being available to attack the glycoside products released in the reaction media (Nakai et al. 2010).

GPs can also be used *in vivo* to ferment oligosaccharides and polysaccharides more efficiently than by using GHs. Indeed, as demonstrated in the cellulolytic bacterium *Clostridium thermocellum*, (Zhang and Lynd 2005), during phosphorolysis, the chemical energy of glycosidic bonds is kept and not annihilated as in the case of hydrolysis, of which the reaction products have to be phosphorylated by kinases to enter the central metabolism as glycosyl phosphates. This might explain why so many identified phosphorylases belong to anaerobes.

Finally, as summarized in table 1, known GPs allows the generation of α- and β-D-glucose-1-phosphate, α-D-GlcNAc-1-phosphate, α-D-maltobiose-1-phosphate, 6-phospho-D-fructose, α-D-galactose-1-phosphate and α-D-mannose-1-phosphate. In order to produce them in high amounts, several feedstocks can be employed. It specifically concerns starch/amylose, cellulose/cellobiose, sucrose, and mannan/galacto/glucomannan. Moreover, all these glycosyl-phosphates can be readily used to produce other specific oligo- or polysaccharides of higher added value, thanks to an expanding array of available GPs. In order to increase the catalog of GP-mediated reactions, two ways are available. The first one

would consist in creating mutants of already existing GPs to teach them new tricks. This has been done several times to expand the panel of donors and acceptors that can be used by cellobiose-phosphorylases (Desmet et al. 2012; Hamura et al. 2013) and trehalose phosphorylases (Chen et al. 2014). The second way rather consists in mining what Nature already created to face the structural complexity of glycosides. This aspect will be developed in the next chapter, focusing on bacterial mannoside metabolization.

3. Mannoside recognition and degradation by bacteria

Mannosides are forming a vast group of glycans widely distributed in nature. Produced by almost all organisms living in various habitats, these glycans serve numerous roles in cells, such as protein maturation and signalling, immunity, mediating protein-protein interactions or participating as structural elements. This ubiquitous presence of mannosides in the environment is a great source of energy and carbon for bacteria, which developed complex strategies to harvest them. This review focuses on the different mannosides that can be found in nature, their involvement in cell-cell interactions and finally describes the catalytic machinery and metabolic pathways that bacteria developed to metabolize them.

3.1. Introduction

Mannose is one of the simplest and common hexose found in nature. As a monomer, it can be used by most of living organisms to support their growth. When integrated into glycans, it can also be used either as energy source, signalling molecule, or cell structuring elements, especially in plants. Mannosides thus play a key role in metabolism and cell recognition, and are involved in many diseases, often linked to protein glycosylation disorders (Sharma et al. 2014). Life evolved many kinds of mannosides of different functions, and the appropriate processes to build them up. Besides, to ensure their survival in the microbial ecosystem jungle, many bacteria developed different or complementary strategies to detect and degrade these mannosides. In addition, as mannosides are also involved into cell signalling, mannosylated proteins take also part in the detection of intruders in order to protect one against them. A general overview of the occurrence of mannosides produced by living organisms, their involvement in cell-cell interactions, and the most up-to-date studies on bacterial mannosides assimilation in different environments are reviewed here.

3.2. Diversity of mannosides structures

3.2.1. Eukaryotic mannosides

In mammals, plants, yeasts and fungi, a wide array of mannosides is encountered, either in the form of pure glycans or in glycoconjugates. This later one refers to mannosyl residues carried by proteins as post-translational modifications, to those which are linked to lipids, or to the hybrid structures containing proteic, lipid and glycan parts. In glycoproteins, mannosyl residues are mostly found in N-linked glycans, even if O-linked glycans and glycosylphosphatidylinositol (GPI) anchors also contain mannosyl residues. With the exception of hemicelluloses which are specific to plants, other mannosides are produced by

all eukaryotes. The mechanisms used for their biosynthesis are similar between the different reigns, but inter species variations exist both in their structures and synthesis pathways.

3.2.1.1. Mammalian mannosides

N-glycans are the most common forms of mannosides found in mammals (Apweiler et al. 1999). These glycans are attached to Asn residues of the majority of proteins, shaping their properties (Skropeta 2009). N-glycan biosynthesis has been largely detailed in several reviews (Helenius and Aebi 2004; Aebi et al. 2010; Larkin and Imperiali 2011), and will not be detailed here. The N-glycan maturation processes yield different N-glycan structures, depending on the fate of the protein, but all mature N-glycans share a common $Man_3GlcNAc_2$ pentasaccharide core. An addition of N-acetylglucosaminyl, galactosyl, fucosyl or sialyl residues yields a wide array of structures that can be grouped under 3 different classes, namely High mannose (HMNG), Complex (CNG) and Hybrid N-glycans (HNG) (fig. 16A). Mammalian N-glycans found on mature proteins are rarely of high mannose type, but rather hybrid or complex, the high mannose being mostly restricted to immature proteins (Nagae and Yamaguchi 2012).

O-mannosyl glycans are the second kind of mannosides bound to mammalian proteins (Lommel and Strahl 2009). Believed for a long time to be restricted to fungi, where they are highly abundant (De Groot et al. 2005), they have also been identified in metazoans and particularly in humans, mainly in nerve tissues or chondroitin sulfate proteoglycans (Praissman and Wells 2014). As for N-glycans, O-mannosyl glycans display a relatively broad structural diversity. All of them, however, share a common β-D-GlcNAc-1,2-D-mannose core structure which can be extended by additional sugars (galactosyl, sialyl, glucuronyl, *N*-acetylglucosaminyl or fucosyl resiudes), species specific (Fig. 17A).

GPI anchors are post-translational modifications of the C-terminal extremity of many proteins, allowing them to bind to the outer layer of the cell membrane (Paulick and Bertozzi 2008). A great number of GPI-anchored proteins have been identified in eukaryotes, ranging from protozoa and fungi to humans. GPI anchor structure is formed by 3 domains: a phospholipid tail, a conserved glycan part and a phosphoethanolamine moiety linked to the bound protein (Ikezawa 2002). The glycan part is formed of a highly conserved α-D-man*p*-1,2-α-D-man*p*-1,6-α-D-man*p*-1,4-α-D-GlcN-1,6-*myo*-inositol motif (fig. 17C). This modification is uncommon in that proteins always have their C-termini linked to the phosphoethanolamine group on the non-reductive mannosyl end. This core can be extensively modified by attachment of side chains containing phosphoethanolaminyl, mannosyl, galactosyl, sialyl, *N*-acetylglucosaminyl and *N*-acetylgalactosaminyl residues (Fujita and Kinoshita 2010). GPI anchor affects a wide array of proteins, which are involved in

signal transduction (Mukasa et al. 1995), immunity, interaction with trypanosomal parasites (Ferguson et al. 1988), or even prion pathogenesis (Chesebro et al. 2005).

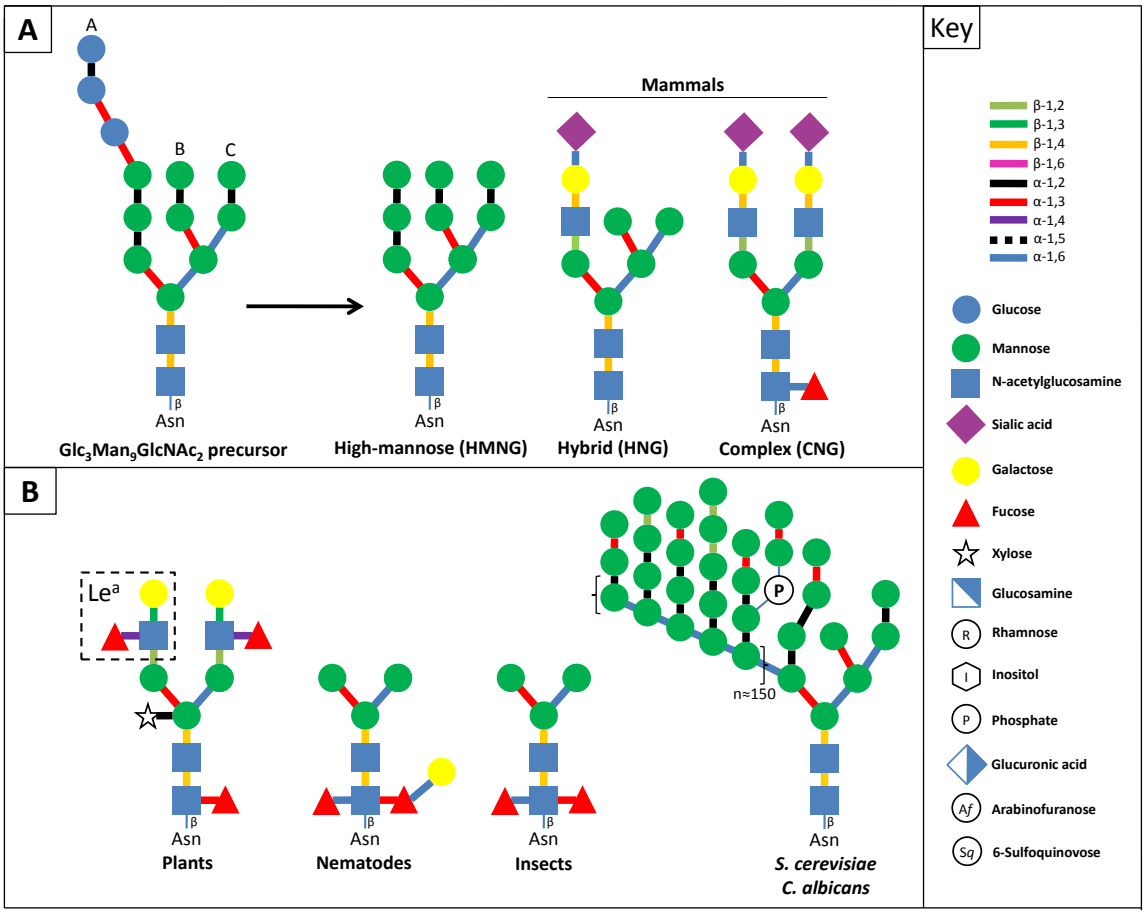


Figure 16: Eukaryotic N-Glycans structures.

A. Structure of the ubiquitous $Glc_3Man_9GlcNAc_2$ N-glycan precursor, and the 3 classes of mature N-glycans. B. Typical structures of mature N-glycans found in various eukaryotic organisms. The Lewis antigen (Le^a), Fuc-α-1,4(Gal-β-1,3)-GlcNAc trisaccharide often found on plant or food allergens is shown in the insert. Note: the very high mannose type N-glycan present in yeasts is similar between *S. cerevisiae* and *C. albicans*, with minor variations. Indeed, the *S. cerevisiae* one lack the β-1,2 linked mannosyl residues and contains a long α-1,6 branch which is about 150 residues long, compared to the 50-150 residue long one found in *C. albicans*.

C-mannosylation is a much rarer event. It has been observed in mammals, mostly humans, but also in other animals (Furmanek and Hofsteenge 2000; Munte et al. 2008; Buettner et al. 2013), but not in plant, yeast, fungi or prokaryotes. It consists in the formation of a α- C-C bond between a mannosyl moiety and C_2 atom of the indolyl moiety of tryptophan, on the first tryptophan of the conserved motif W-x-x-W (Löffler et al. 1996). Literature on that topic is very scarce, but this modification seems to be common for proteins involved in immunity, such as complement proteins (Hofsteenge et al. 1999), or interleukin-12 (Doucey et al. 1999). Its role is still unclear, but examples indicate its involvement in secretion (Goto et al. 2014) and activity tuning, since the secreted Cys subdomains of Muc5A/C and Muc5B lung mucin protein have been found to be retained in the ER if unmannosylated (Perez-Vilar et al. 2004). C-mannosylation has also been observed

in viruses. The Ebola viral protein sGP, which is the first reported example of viral C-mannosylation, seems not to be affected by this post-traductional modification, as no significant change in expression, fold or activity has been observed when unmannosylated (Falzarano et al. 2007). However, further investigations are necessary to elucidate the specific roles of this kind of glycosylation, its biosynthesis regulations, and how this very uncommon glycan linkage is degraded, particularly by bacteria.

3.2.1.2. Plant mannosides

With the exception of O-mannosides, which are absent from ground plants and algae, but largely present in yeast and fungi, and C-mannosylation, which has still not been described for them, plant mannosides share the same kind of mannosides as described above for mammals. However, in addition to some little structural discrepancies of which we are about to discuss, plant do possess specific mannosyl containing molecules, the hemicellulosic β-linked mannans. Both kinds of plant mannosides will be described here.

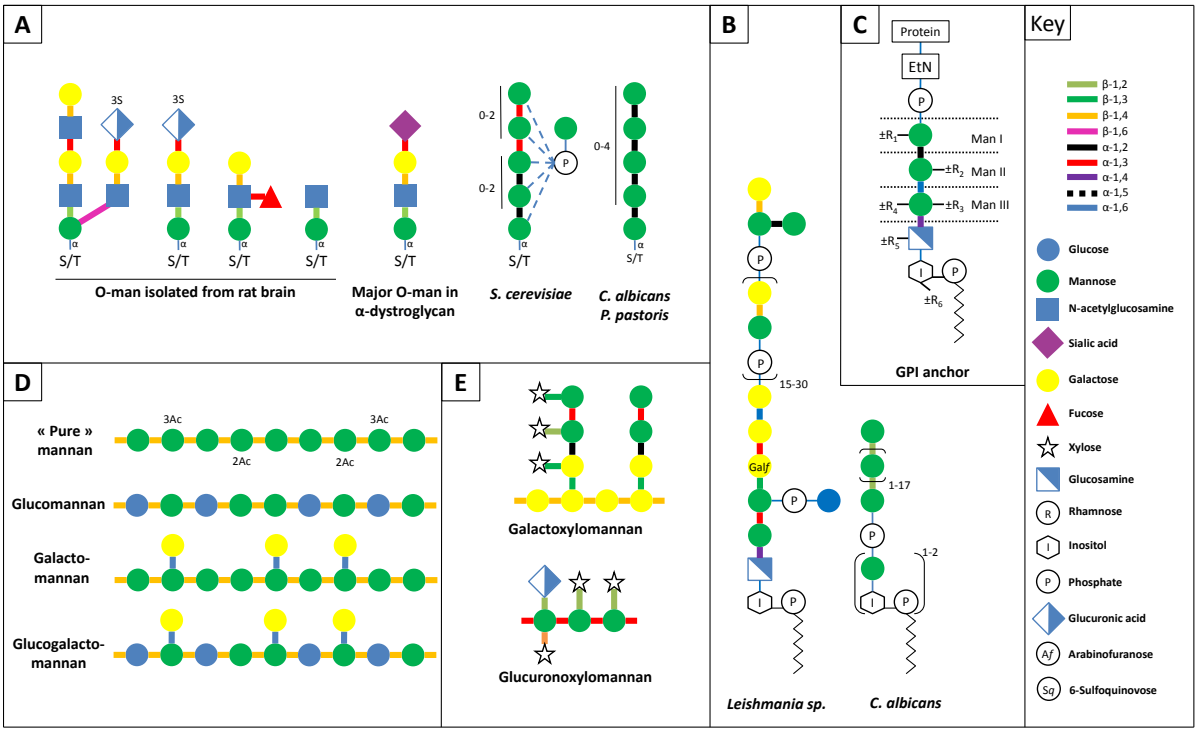


Figure 17: Other structures of eukaryotic mannosides.

A. Structure of eukaryotic O-mannans. B. Typical structures of lipophosphoglycans. C. Highly conserved structure of the GlycosylPhosphatidylInositol (GPI) anchor found in all eukaryotes. EtN: Ethanolamine. D. Structures of plant mannans, with examples of each subclass. E. Structures of fungal capsular polysaccharides.

The first steps in plant N-glycan synthesis are identical to what is observed for mammals, and also relies on the formation of the $Glc_3Man_9GlcNac_2$ precursor in the ER (Pattison and Amtmann 2009; Gomord et al. 2010; Song et al. 2011). But, in addition to the ubiquitous N-x-S/T sequon used for protein attachment, an unusual N-x-C sequon has also been described (Matsui et al. 2011). The maturation process also occurs in the Golgi's

apparatus, but plant N-glycans displays some specific structural features. For instance, β-1,2-xylosyl residues, linked to the β-1,4-mannosyl residue of the core pentasaccharide, and α-1,3-fucosyl ones linked to the reducing end N-acetylglucosamine are typically found. The Lewis a epitope [Galβ1-3-(Fucα1-4)GlcNAc] is also found at the extremities of plant complex type N-glycans branches (fig. 16B). This epitope, found on what is called 'secreted type N-glycan', has been found in many foodstuffs (Wilson et al. 2001) and pollen allergens (Maeda et al. 2005). In plants, protists, archaea, eubacteria and fungi, β-1,4 linked galactosyl and sialyl residues, which are the signature of mammalian complex N-glycans, are absent (Zeleny et al. 2006). However, the unicellular green algae *Chlamydomonas reinhardtii* CC-125 produces mammalian-like N-glycans, containing both the β-1,4-Gal and the sialylated complex glycan structure together with a plant-like core α-1,3 fucosylation (Mamedov and Yusibov 2011). The so-called 'truncated' type N-glycan Man_3-Xyl-$GlcNAc_2$-Fuc (also called paucimannose or vacuolar) is the most commonly found type of N-glycans in vacuolar storage proteins in seeds (Kimura and Matsuo 2000), but contrary to mammalian N-glycans, plants often carry HMNG, and in some cases, exclusively. For instance, algea such as *Codium fragile*, *Chondrus ocellatus*, *Sargassum piluliferum* or *Zostera marina* contains exclusively HMNG, being devoided of complex ones (Yoshiie et al. 2012). Contrary to mammals, for which CNG defects are almost associated with diseases (termed congenital disorders of glycosylation), it seems to be less the case for plants, for which N-glycan processing in the Golgi may be dispensable, even if it associates with diseases under certain stress conditions (Strasser 2014). Free forms of N-glycans have also been detected in plant, during all stages of their development at micromolar concentrations (Maeda and Kimura 2014). They originate mostly from misfolded proteins processing, but they have also been found ubiquitously in all plant tissues associated with auxin-like function, participating for instance in elongation of stems and maturation of fruits, which suggest a specific role for these molecules (Meli et al. 2010).

Plants do produce GPI-anchored proteins (Borner et al. 2003). They are involved in many biological functions, such as cell surface synthesis and remodelling (Liu et al. 2013) or pollen tube-female gametophyte interaction (Capron et al. 2008). However, compared to their mammalian or yeast counterparts, plant GPI-anchored proteins were only characterized recently (Borner et al. 2003). They are structurally very similar to the ones borne by animals or yeasts (Schultz et al. 1998). However, the presence of a galactosyl residue linked to the β-1,4 mannosyl of the core pentasaccharide seems to be plant specific (Ellis et al. 2010).

In plant cell wall, cellulose microfibrils are associated to a dense network of hemicelluloses, pectins, structural glycoproteins and lignin. A large variability on the different proportions of these polymers is observed upon the species, tissue and developmental stage. Hemicellulose constituents are heteroxylans, mixed-linkages glucans,

and heteromannans, all containing β-linked backbones. Hemicelluloses roughly form one third of the total mass of the plant cell wall (Pauly et al. 2013). Contrary to cellulose, which is a non-ramified linear polymer reaching a degree of polymerisation (DP) of 7,000-15,000, hemicelluloses are shorter polysaccharides of DP 100-3000 that can be extensively ramified, which explain why galactan does not fall into that group, as it is more a pectin side chain rather than a main chain. Here, we will focus on the heteromannan sub-group, which refers to a group of hemicelluloses containing mannosyls residues that can be associated to glucosyl and/or galactosyl moities. Mannans are the most abundant hemicellulosic component of softwoods, with a widespread distribution in plant tissues. They are also found in some algae (Domozych et al. 2012). Mannans can be divided into four groups. Linear mannan (also called 'pure mannan'), is made of a main chain composed of more than 90% of β-1,4 linked mannopyranosyl residues. In glucomannan, the main chain also contains β-1,4 linked glucosyl units in various amounts. Both linear mannan and glucomannan can bear additional α-1,6 linked galactosyl branches, to form galactomannan and galactoglucomannan, respectively (Scheller and Ulvskov 2010) (fig. 17D). Species such as *Orchidaceae* also harbour 2- and 3-*O*-acetylated forms of mannans (acemannan). Glucomannan is the major component of softwoods, with a DP of 200 and a ratio Man:Glc of 3:1. It is less present in hardwoods, constituting 3-5% of cell wall, also being shorter (DP70) and with a Man:Glc ratio of 1,5:1 to 2:1 (Hongshu et al. 2002). Galactomannans however, are mainly found in endospermic tissue of seeds of *Leguminosae*, underlining that they are not only present as structuring elements in the plant cell wall, but are also present as energy storage compounds, called Cell Wall Storage Polysaccharides (CWSP) (Buckeridge 2010). The abundance of galactomannans in *Leguminosae*, which constitute a significant part of human diet, makes them dietary fibres as well as other plant cell wall constituents. In plants, galactomannans were also described as signalling molecules for growth and development (Liepman et al. 2007). Moreover, the rheological properties of galactomannan are different to those of the other mannans. Indeed, the presence of galactosyl residues allows for a higher hydrophilic behaviour, making galactomannan strongly involved in water retention in order to avoid seed drying (Ferreira et al. 2009).

3.2.1.3. Specificities of yeast and fungal mannosides

Yeasts and fungi produce essentially the same kind of mannosides as the other eukaryotes, with the exception of hemicelluloses which are specific to plants and C-mannosides which have been only identified in multicellular animal organisms.

However, it is worth noting that the vast majority of yeast and fungal glycoconjugates contains substantially much more amounts of mannosyl units. Indeed, yeasts and fungi are devoided of complex N-glycans. The N-glycan precursor synthesis in the ER is extremely conserved in all eukaryotes, but trimming steps in the Golgi's apparatus do vary, producing

in yeast and fungi huge structures containing so much mannosyl units that it is termed mannan (also sometimes termed mannoproteins) (Munro 2001; Hall and Gow 2013) (fig. 16B). Structurally, this N-glycan is close to the High mannose type found in other eukaryotes. The main difference resides in the presence of a long Man*p*-α-1,6 linked side chain on the central α-1,3 bisecting branch of the core pentasaccharide, which serves as a scaffold for Man*p*-α-1-2 ramifications. These ramifications are then capped by additional α-1,3-linked mannosyl residues. In addition, branches linked through a phosphoester linkage can be released by acidic treatment, and form the acid-labile part of the glycan (which explains why these proteins are sometimes called phosphopeptidomannan).

A specific feature of fungal N-glycans compared to yeasts, is the presence of β-1,2-linked mannosyl residues held on the α-1,2-linked ramifications (Shibata et al. 2003). More, an additional N-acetylglucosaminyl residue α-1,4 bound to the β-1,4 linked mannosyl of the core pentasaccharide has been reported in the fruiting body of the basidiomycete *Coprinopsis cinerea* (Buser et al. 2010). This observation indicates that the substitution on this mannosyl is reign dependant, plant adding β-xylose, animals adding β-GlcNAc, and fungi adding α-GlcNAc. Moreover, the N-glycans of the opportunistic pathogen *Aspergillus fumigates* harbours Gal*f* residues. This galactose containing glycan (termed galactomannan) forms up to 14% of the extracellular matrix (Schmalhorst et al. 2008), and has been implicated in *A. fumigatus* virulence, just as for the Leishmanian lipophosphoglycan that will be detailed in the next pages (Loussert et al. 2010). Fucosylated oligomannose N-glycans have been recently reported in several fungal species belonging to the phylum *Basidiomycota*, indicating that fucosylation seems to be a common feature of fungal N-glycans (Grass et al. 2011). The long α-1,6 side chain of fungal N-glycans seems to be shorter than the yeast one or even absent, since the one of the pathogenic yeast *C. albicans* is on average 41 mannosyl residues long, compared to the 100-150 residue long one found in *S. cerevisiae* (Masuoka 2004). Finally, it seems that fungal N-glycans are long, and closer to yeast ones when under budding, unicellular form, while vegetative mycelium forms mainly HMNG, closer to the one bared by the other eukaryotes (Buser et al. 2010).

O-mannosylation is largely present in yeast and fungi, and both reigns share common structural features that differentiates them from the other eukaryotes. Indeed, the O-mannosides of yeast and fungi are less complex than their animal counterparts (Hall and Gow 2013). They are almost exclusively composed of mannosyl residues, forming a chain of α-1,2 and α-1,3 linked, 1-5 residue long. However, other sugars such as Gal*p*, Gal*f* and Glc*p* can be present in the main chain or in ramifications, especially in filamentous fungi (Goto 2007). In the yeast *S. cerevisiae*, a phosphomannosyl residue can be added on any of the mannosyl residues forming the main chain (fig. 17A).

Synthesis of GPI-anchored proteins is essential for *S. cerevisiae* survival and growth (Fujita and Jigami 2008). The GPI moiety contain either a diacylglycerol with very long saturated fatty acid at the sn-2 position (Fankhauser et al. 1993), or a ceramide, more precisely a phytosphingosine with a $C_{26:0}$ fatty acid, or a phytosphingosine containing a monohydroxylated $C_{26:0}$ (Conzelmann et al. 1992). Lipid remodeling steps are intimately related to lipid raft association in biological membranes, thus directly participating in activity tuning (Wang et al. 2013). Regarding the glycan part of the GPI anchor, in *S. cerevisiae*, it mostly contains mannosyl residues α-1,2 or α-1,3 linked to the ManI, but the presence of phosphoethanolaminyl residues on the other branching points is variable, and appears to be removed on mature proteins (Fujita and Jigami 2008) (Fig. 17C). In fungi, *C. albicans* is hypothesized to possess twice the number of GPI-anchored proteins of *S. cerevisiae* (Richard and Plaine 2007). GPI-anchors are essentially identical to that of the other eukaryotes, and their biosynthesis follows the same pathway. However, compared to other eukaryotes, in yeasts and fungi, some mature GPI-anchored proteins can undergo an additional maturation step. The GPI-anchor may be cleaved off the protein, between the glucosamine and the first mannosyl moiety, in order to direct it to the cell wall, and covalently attach it to the β-1,6 glucan (Van Der Vaart et al. 1996). A specific signal in the propeptide of the nascent protein seems to be required to retain the GPI-anchored protein to the plasma membrane, while those which do not possess it are directed to the cell wall (Frieman and Cormack 2003). This signal is composed of two basic amino acids in the four amino acids upstream of the ω-site of propeptide cleavage in yeast, while in fungi this signal is only composed of a single basic amino acid at the ω-1 or ω-2 sites (Ouyang et al. 2013).

Fungi are also composed of a cell wall containing diverse polysaccharides. Among them, phospholipomannan (PLM) is a mannose containing glycoconjugate, first described for *C. albicans* (Trinel et al. 2002) (Fig. 17B). Its glycan part is formed by a linear chain of 3 to 20 β-1,2 linked mannosyl residues, depending on the serotype. The lipid part is formed by a phytoceramide associating a C_{18}/C_{20} phytosphingosine and C_{26}, or mainly C_{24} hydroxy fatty acids, similarly to GPI anchors. The linker region between the lipid and the glycan part is unusual, composed of a –man*p*-α-1-P-6-O-man*p*-α-1,2-inositol-1-P-lipid (M(IP)) motif, or a repeated version, containing two inositol phosphate residues $M(IP)_2$ (Trinel et al. 2005).

Some pathogenous fungi also contain a capsule, composed of a variety of polysaccharides. It is mainly formed by association of glucuronoxylomannan (GXM) and galactoxylomannan (GalXM) (Zaragoza et al. 2009) (fig. 17E). In the pathogenic fungi *Cryptococcus neoformans*, GXM forms a long polysaccharide of 1.7×10^6 Da that comprises more than 90% of the capsule's polysaccharide mass, and mediates multiple deleterious effects on host immune function (Zaragoza et al. 2009). GXM is composed of a α-1,3-mannan chain, which is branched by additional β-1,2 glucuronic acid residues each 3

mannosyl units. Then, depending on the serotype, additional β-1,2 or β-1,4 xylosyl residues can be added, and mannosyl residues may be 6-O-acetylated (Cherniak and Sundstrom 1994). On the other hand, GalXM is much shorter (1.0×10^5 Da), and constitutes 5-8% of the capsular mass. It is formed by a α-1,6-galactan backbone on which four potential short oligosaccharidic branches can be added. These branches all consist in an α-1,3-D-Man*p*-α-1,4-D-Man*p*-α-1,4-D-Gal*p* trisaccharide, which holds variable amounts of β-1,2 or β-1,3 xylosyl residues.

3.2.1.4. Protozoan mannosides

In some protozoans, and particularly in the *Leishmania* parasite species, a particular class of mannosides is found in the glycocalyx. This molecule shares structural similarities both with the GPI-anchor and the bacterial lipopolysaccharide (LPS), we will describe later. This structure is known as lipophosphoglycan (LPG) (Beverley and Turco 1998). Both LPG and LPS are intimately involved in host-pathogen interactions, as will be detailed later. LPS looks like a GPI anchor since it is composed of a lipid tail allowing anchorage in the membrane bilayer, a phosphoinositol linker and a glycan moiety. However, contrarily to GPI anchors, the glycan part is much larger, and does not hold a protein (fig. 17B). LPG molecules can be divided in four parts: i) a lipid tail (monoalkyl-lysophosphatidylinositol with saturated C_{22} to C_{24} alkyl groups in *Leishmania* species), ii) a core heptasaccharide, containing galactosyl, glucosyl-α-1-phosphate, mannosyl and glucosaminyl moieties, iii) a central multiple repeated β-D-gal*p*-1,4-D-man*p*-phosphate disaccharide (15-30 repeats), and iiii) a Gal-β-1,4(man-α-1,2)-man-α-1-phosphate cap. This molecule plays a key role in parasite invasion and survival, mostly through the notable presence of a Gal*f* unit in the heptasaccharide core (Oppenheimer et al. 2011). The Gal*p*-β-1,4-Man*p*-α-1-phosphate disaccharide is also largely participating in antigenicity, and can be directly linked to GPI anchored proteins involved in *Trypanosomatidae* mediated diseases (Descoteaux and Turco 1999).

3.2.2. Prokaryotic mannosides

For a long time, bacteria & archaea were considered to be devoided of post-translational modifications other than phosphorylation. Finally it is not the case, as N- and O-glycans have been found in many bacteria and archaea (Lommel and Strahl 2009; Calo et al. 2010; Nothaft and Szymanski 2010; Larkin and Imperiali 2011; Nothaft and Szymanski 2013). However, it is obvious that N- and O-glycans of prokaryotic origin are necessarily different of those of eukaryotes, since they lack both the ER and the Golgi apparatus, place of their assembly and maturation processes in eukaryotes.

Many bacterial N-glycans have been reported, but most of their structural patterns remain incompletely elucidated, and to date, no bacterial N-glycan structure containing

mannose has been reported. On the contrary, genome analysis predicted N-glycosylation to be a common post-translational modification in archaea (Kaminski et al. 2013). Archaeal N-glycosylated proteins share several common features with bacteria and eukaryotes. Similarly to bacteria, archaea produce a wide diversity of N-glycan structures, for which an increasing number is being reported (Kärcher et al. 1993; Voisin et al. 2005). There is however no apparent structural uniformity between the different N-glycan generated, which is a specific feature that seems to be associated with the different habitats of these organisms (Calo et al. 2010). Contrary to bacteria, among the panel of reported archaeal structures, a great number contain mannosyl residues. For instance, the S-layer glycoprotein from *Methanothermus fervidus* has been reported to hold an hexasaccharide β-linked to Asn (Kärcher et al. 1993). The reported hexasaccharide is α-D-3-O-MeMan*p*-1,6-α-D-3-O-MeMan*p*-(1,2-α-D-Man*p*)$_3$-1,4-D-GalNAc. These structures, also often contains *N*-acetylated carbohydrates such as GlcNAc, GalNAc, or ManNAc, and sulfated sugars, uronic acids, furanose forms of galactose and a unique 2-acetamido-2,4-dideoxy-5-O-Methyl-hexosulo-1,5-pyranose, the first reported example of aldulose in a N-glycan structure (Ng et al. 2011). Nevertheless, some structurally characterized archaeal N-glycans are structurally close to the ones bared by eukaryotes. Indeed, the S-layer glycoprotein from *Sulfolobus acidocaldarius* has been reported to contain N-glycans linked to Asn residues through a chitobiose moiety, the largest one being 6-sulfoquinovose-Glc$_1$Man$_2$GlcNAc$_2$ (Peyfoon et al. 2010) (fig. 18A). Similarly, a *Thermoplasma acidophilum* plasma membrane glycoprotein has been found to contain a glycan highly mannosylated with α-1,2, α-1,3 and α-1,6 linkages, linked to the Asn residues of proteins through a mannochitobiose trisaccharide (Yang and Haug 1979) (fig.18A). These glycans play a key structural role, since the S-layer glycoprotein participates to cell shaping (Eichler and Adams 2005). Disruption of genes involved in N-glycan build-up and processing or chemical interference with these processes gave insight in the functional roles of these glycans. It appeared that bacitracin treatment (inhibitor of C_{55}-isoprenyl pyrophosphate dephosphorylation) impaired flagella expression in *Methanococcus deltae* (Kalmokoff et al. 1992). But inconsistency on the vital role on N-glycosylation appeared, since some archaeal species could survive without any active N-glycosylation pathway (Chaban et al. 2006), while others could not (Meyer and Albers 2013). However, it clearly appeared that the presence of such glycan coating is strongly associated with the physical challenge of survival in harsh environments. For instance, depending on the salt concentration, variability in N-glycan pattern occur in the halophilic archaeon *Haloferax volcanii* (Guan et al. 2012), while the extreme holophile *Halobacterium halobium* produce sulfated or uronic acids containing N-glycans, in order to adapt itself to the presence of high salt concentrations (Mengele and Sumper 1992). Finally, one can note that in archaea, N-glycans are attached onto proteins at the same N-x-S/T (x≠P) sequon as for eukaryotes, even if an additional N-x-N/L/V sequon has also been reported in *Halobacterium halobium* (Zeitler et al. 1998).

O-glycosylation also occurs both in bacteria and archaea. For the latter however, O-glycan literature is extremely scarce, and only reports examples for *Halobacterium salinarum* and *Haloferax volcanii*, where a Glc-α-1,2-Gal disaccharide is found attached to Ser and Thr (Mescher and Strominger 1976; Sumper et al. 1990). In bacteria on the contrary, O-glycosylation has been reported multiple times, and an increasing amount of structural details are available. Most of the known O-glycan containing mannosyl residues were identified from actinomycetes glycoproteins, with a particular focus onto the human pathogen *Mycobacterium tuberculosis*. Indeed, many of them are immunologically active molecules that make an important contribution to virulence and to the host-bacteria interaction, as will be discussed later. Protein O-mannosylation in actinomycetes resembles O-mannosylation in yeasts. For instance, the *Mycobacterium tuberculosis* Apa protein has been found to be O-mannosylated on multiple Ser/Thr in Pro rich C- and N-terminal domains with 1 to 3 α-1,2 linked mannosyl residues (Dobos et al. 1996), even if glycans containing up to 10 mannosyl residues and a α-1,3 linkage have been identified in the MPB83 protein from *Mycobacterium bovis* (Michell et al. 2003). Several other bacteria were found to produce O-glycans, including *Corynebacterium glutamicum* (Hartmann et al. 2004), *Streptomyces coelicor* (Wehmeier et al. 2009), *Streptococcus gordonii* (Takamatsu et al. 2004), and even a distant Gram negative bacterium such as *Bacteroides fragilis* (Fletcher et al. 2009) (fig. 18B).

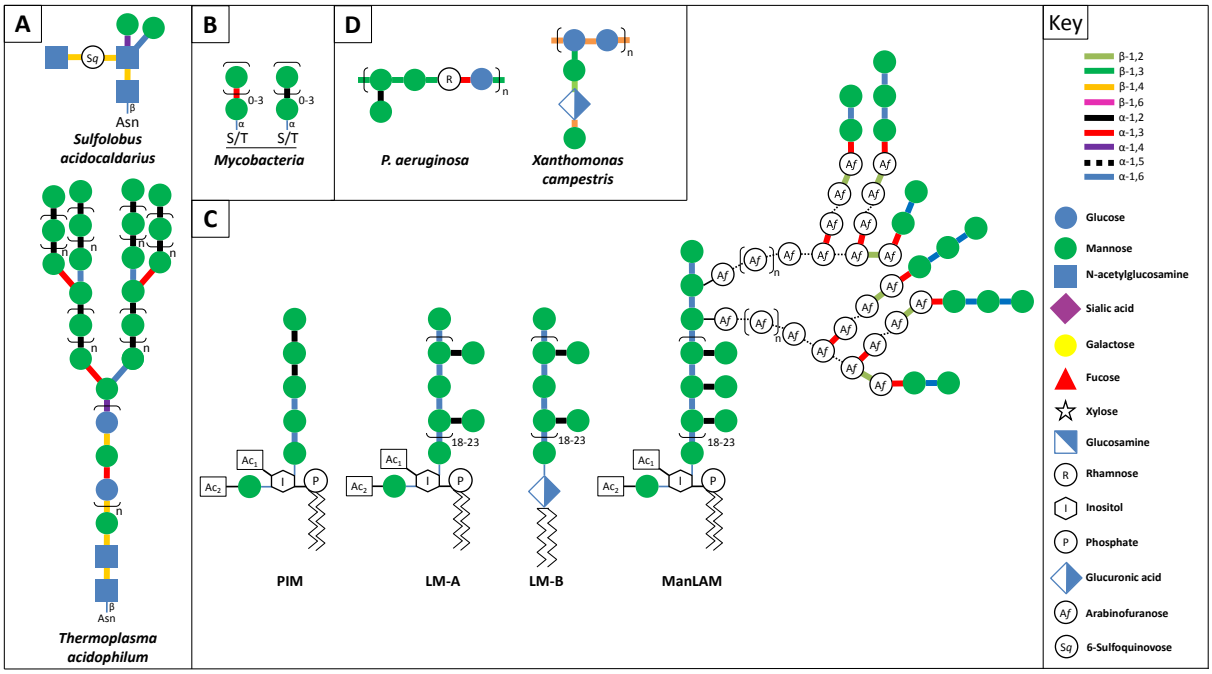


Figure 18: Structure of mannosides from prokaryotes.
A. Structure of archaeal N-glycans. B. Bacterial O-glycans. C. Lipoglycans found in Actinomycetes. D. Bacterial capsular polysaccharides.

GPI-anchored proteins have been identified in archaeal species but to date not in bacteria (Kobayashi et al. 1997; Eisenhaber et al. 2001). Little is known about their structure, function and biosynthesis processes. However, it seems that it is closely related to the

eukaryotic GPI anchor both in regards of its structure and function, since many of the archaea where GPI anchored proteins have been identified are closely related to eukaryotes. However, it has been shown that *Methanosarcina barkeri* produces glucosaminyl archaetidylinositol, a closely related to glycosyl phosphatidylinositol molecule found in eukaryote GPI-anchor, suggesting that as for N-glycans, some little structural discrepancies are present between eukaryotes and prokaryotes (Nishihara et al. 1992).

Mannoglycolipids are present in many prokaryotes, but the most striking example certainly resides in the pathogen *Mycobacterium Tuberculosis*. The mycobacterial cell envelope is a unique structure in that it forms more than 60% of the cell dry weight. One of the features of this envelope is its exceptionally high content of mannosylated molecules. Among them, phosphatidyl-myo-inositol mannosides (PIMs), lipomannan (LM) and lipoarabinomannan (LAM) are of particular interest because of their strong involvement in pathogencity (Kaur et al. 2009) (fig. 18C). They are structurally related to the LPS found in certain Leishmanian species, as introduced above. A deeply detailed view of the structures and biosynthesis of these components can be found in the review paper of Kaur D et al (Kaur et al. 2009). These molecules share a common phosphatidyl-inositol lipid anchor with some variations regarding the number, location and nature of the fatty acids (Gilleron et al. 2008). PIMs contain a mannosylated extention of 1 to 6 mannosyl residues, named PIM_{1-6}. The inositol molecule can hold α-man*p* on the 2-position, while position 6 holds a chain of 5 α-1,6 and α-1,2 mannosyl residues. Two additional acyl chains can be added onto position 3 of the inositol and the C_6-OH of the mannosyl linked to position 2, to yield Ac_1PIM_n and Ac_2PIM_n, respectively (fig. 18C). The LM molecule is slightly different of PIMs in that the mannosyl chain linked on the 6 position of the inositol is much longer (20-25 residues) and only contains α-1,6 linkages (Kaur et al. 2007). Moreover, this chain can hold α-1,2 ramifications, but the branching positions remain unelucidated. LAM contains in addition a arabinan motif linked on some non-terminal mannosyl residues in the mannan core (Shi et al. 2006). The arabinan polymer contains around 60-70 arabinofuranosyl residues depending on the species, consisting in a main chain of α-1,5 linkages with further ramifications of α-1,3 linkages. This structure can be terminally α-1,5 mannosylated, or hold α-1,2 Man_2 or Man_3 capping the Ara*f* chains to form the ManLAM molecule found in *M. tuberculosis*. These glycolipids are conserved among Corynebacteria, but another LM type has been identified both in *M. tuberculosis* and *C. glutamicum* (Lea-Smith et al. 2008). It is closely related to the LM described above (renamed thereafter LM-A) which is termed LM-B. The glycan part is similar to LM-A, but it is anchored through a glucuronic acid diacylglycerol. The relevance of this new mannolipid and its importance for this pathogen remains however to be assessed.

Contrarily to its cell wall, the mycobacterial capsule contains 95-99% proteins and glycans (Ortalo-Magné et al. 1995). Among the latter, three types are encountered: α-1,4-

glucan, arabinomannan, and mannan. α-1,4 glucan composes 80% of the total carbohydrate content and its molecular mass reaches 100.000 Da. Arabinomanan is structurally similar to the lipid anchored LAM, while mannan is made of α-1,6 man*p* residues (Ortalo-Magné et al. 1995). This mannan chain holds some branches consisting of α-1,2 linked mannosyl residues, making it structurally closely related to mannan chain of lipid anchored arabinomannan.

Many prokaryotes secrete extracellular components, such as proteins, signaling molecules, or polysaccharides. A good example of the latter is found in the biofilm secreted by certain prokaryotes. For instance, the opportunistic pathogen *Pseudomonas aeruginosa* produces *t*he Psl polysaccharide, which consists in a repeated pentasaccharide containing D-mannosyl, D-glucosyl and L-rhamnosyl residues (Byrd et al. 2009), with a molecular mass of 0.5-2 × 10^6 Da (fig. 18D). Another example is the biofilm formed by the plant pathogen *Xanthomonas campestris*. This bacterium produces a polymer called xanthan, which participates in plant invasion and virulence of the bacterium. It is however much more known for its applications in food industry as a thickener or viscosifier. Xanthan is formed by a main chain of cellulose, β-1,3 branched every 2 glucosyl units by the trisaccharide β-D-man*p*-1,4-D-GlcA-1,2-D-man*p* (fig. 18D). The two mannosyl residues are derivatized by additional pyruvic and acetyl groups (Crossman and Dow 2004).

3.3. Eukaryotic Mannosides Recognition by bacteria

As we saw above, almost any living organism synthesizes mannosylated glycoconjugates and/or polysaccharides, which are exposed to the outside environment. Any time a cell meets another cell, it is necessarily facing the glycan coat surrounding it (Boraston et al. 2003). In the present paper, we will only consider the interactions occurring between bacteria and the mannosides they may encounter in their environment. Bacterial-mannoside interactions may either induce a 'positive' reaction, such as in symbiosis or mutualism which generally does not imply glycan degradation, or a negative one, such as in pathogenism, parasitism or even commensalism, where bacteria feed on the glycans harboured by other linving cells, triggering defense mechanisms.

Depending on the ecosystem where they live, bacteria are exposed to different mannoside structures. In all cases, carbohydrate recognition involves specific proteins which can be directly coupled to degradation mechanisms, or induce a cascade of signal transduction processes. In carbohydrate harvesting processes, mannoside recognition can be directly associated to degradation, i.e when the receptor is physically connected to the breakdowning activity, like in the case of carbohydrate binding modules (CBM), which enhance the enzymatic efficiency by guiding the catalytic domain towards its substrate (Zhang et al. 2014). CBMs that are specific for β-mannosides are classified in 8 CAZY families (CBM16, 23, 27, 35, 59 and CBM26 harbored by β-mannanases of the GH5 and GH26

families, CBM13 found in plant lectins that binds mannose, CBM29 found in a non-catalytic component from *Piromyces equi*, and the CBM_NC appended to GH26 mannanase from *Cellvibrio japonicus* Ueda107).

In bacteria, mannoside recognition and degradation also involve membrane anchored receptors, which activate a signal cascade ultimately initiating or enhancing the production of the appropriate glycan catabolic enzymes. These receptors are multi-component systems, encoded together with mannoside degrading enzymes by polysaccharide utilization loci (PUL) (Terrapon and Henrissat 2014b). Mannoside specific PULs have been recently characterized, mainly in *Bacteroidetes* species but not only (Martens et al. 2008a; Sonnenburg et al. 2010b; Senoura et al. 2011; Martens et al. 2011; Kawahara et al. 2012b; McNulty et al. 2013; Cuskin et al. 2015b). Like all PULs, they code for polysaccharide utilization systems which resemble the starch utilization system (Sus) found in *Bacteroides thetaiotaomicron* VPI-5482 (Reeves et al. 1997; Shipman et al. 2000b; Cho et al. 2001). Sus-like systems are multiple cell-envelope associated protein complexes used for sensing, binding, and subsequent depolymerization of complex carbohydrates. Starch recognition is mainly performed by the membrane associated SusC and SusD proteins, providing 60% of the total starch binding affinity, the rest being provided by additional recognition mediated by SusE and SusF (Shipman et al. 2000b). The crystal structure of the complex SusE-SusF and starch was recently released, highlighting the fact that they are formed by tandem repeats of starch binding modules, providing a strong affinity toward starch, and allows for accommodation of the 3D starch structure (Cameron et al. 2012b). Such a mechanism, but targeting yeast mannosides by *Bacteroides tethaiotamicron*, has been recently deepely characterized. In that case, two proteins (a SusD like protein and a surface glycan binding protein specific for mannose) are involved in mannoside recognition and sequestration, (Cuskin et al. 2015b). Also, the 3D structure of the binding element of a probable β-mannan degradation pathway of the thermophilic anaerobic bacterium *Caldanaerobius polysaccharolyticus* ATCC BAA-17 has been described (Chekan et al. 2014). In that case, the mannobiose and mannotriose recognition involve a solute-binding component of an ABC transporter.

As introduced, in all cases, activation of these mannoside sequestration and degradation pathways would be under the control of regulatory proteins, whether SusR-like or hybrid two component systems (HTCS). They are activated when oligosaccharide resuting from the primary degradation of the polymer by constitutively expressed mannanases are sensed. Such a system allows for the bacterium to rapidly metabolize the generated oligosaccharides.

For the mannoside utilization systems we just described, recognition of plant and yeast mannans by bacteria has a clear goal, their assimilation. However, interactions

between bacteria and eukaryotic mannosides are also implicated in cell adhesion, in most of cases prior to invasion. For that purpose, many bacteria harbor carbohydrate binding proteins belonging to the class of lectins called *Fimbriae* (or pili). These protein structures are not classified in CBM modules of the CAZy database, because they are not contiguous amino acid sequence within a carbohydrate-active enzyme. Fimbriae are formed by polymerization of pilin proteins. They are widespread in Gram negative bacteria and are responsible for adhesion to host cells, mostly through binding of glycoproteins (Lebeer et al. 2010). Different classes of fimbriae have been reported, but the archetypal one is type 1 fimbriae. Its structure forms an extracellular appendage whose top protein, FimH, has been demonstrated to strongly bind mannosyl residues (Bouckaert et al. 2005; Wellens et al. 2008; Korea et al. 2011). Type 1 fimbriae have been implicated in the specific adhesion of various enterobacteria to human epithelial cells (Grzymajło et al. 2013). They were also found associated with membranous cells, a class of cells found in the Peyer's patches in the gut epithelium, which are dedicated to antigen-transportation through the specific involvement of glycoprotein 2, a highly N-glycosylated protein (Ohno and Hase 2010). In addition, type 1 fimbriae is critical for the proper attachement of bacterial cells to mannose containing exacellular polysaccharides found in biofilms (Rodrigues and Elimelech 2009).

3.4.Mannosides Degradation by micro-organisms

Degradation of mannosides is widespread in the microbial world. Mannoside hydrolysis involves various enzymes, of which mannanases and mannosidases are the main effectors. Mannosidases are enzymes involved in the degradation of terminal mannosyl residues, while mannanases are endo-acting enzymes, degrading structures of high polymerization degrees. β–mannanases and β-mannosidases, involved in plant β-mannan catabolism or in degradation of particular motives of yeast or mammal N-glycans, are found in many glycoside hydrolase families of the CAZy classification, namely the families GH5, GH26, GH1, and GH2. α-mannosidases and α-mannanases, acting specifically on mammal and yeast N-glycans, are classified in the GH families GH76, GH31, GH38, GH47, GH63, GH92, GH99 and GH125.

In addition, recently, new mannoside degradation pathways have been reported, involving β-mannoside phosphorylases classified in the GH130 family. Created in 2011, it is the sole enzyme family implicated in mannoside breakdown by phosphorolysis. Contrary to what happens for glycoside hydrolases, for glycoside phosphorylases the cleavage of the glycan interosidic bonds occurs with concomitant phosphorylation of the glycosyl residue released from the glycan non reducing end. The number of sequences clustered in the GH130 family has been rapidly expanding these last years, containing today no less than 653 members, separated in three subfamilies (Ladevèze et al. 2013). Like all mannoside degrading GHs, GPs of the GH130 family act synergistically with GHs of various specificities.

Indeed, as described in the first section of this paper, mannosides are rarely composed exclusively of mannosyl residues. On the contrary, they are most of the time associated to other glycosyl residues, linked with various linkages. Therefore, in order to complete mannoside degradation, bacteria produce a highly diverse panel of glycan breakdowning activities, of which the specificities are directly related to the mannoside structures they have to face in their environment.

3.4.1. Soil and spring bacteria

Plant and algal β-mannans are one of the preferred substrates of bacteria living in soil and water (Moreira and Filho 2008). β-mannan degradation is generaly carried out by a cocktail of β-mannanases and β-mannosidases (Stoll et al. 1999), of which some have been crystallized (Le Nours et al. 2005). Additional GH1, GH3, GH5, GH9, GH30, or GH116 β-glucosidases, GH4, GH27, GH36, GH57, GH97 or GH110 α-galactosidases (Luonteri et al. 1998) are often associated to ensure complete assimilation of gluco- and galactomannans (Duffaud et al. 1997). These enzymes are often permanently expressed at basal level, and associated with CBMs in order to sequester polymeric substrates that cannot be internalized by the cells because of their high polymerization degree (Zhang et al. 2014).

Recently, two PULs involved in β-mannan degradation by *Caldanaerobius polysaccharolyticus* ATCC BAA-17, a thermophilic bacterium isolated from hot spring sediment, have been discovered by transcriptomics (Chekan et al. 2014). Both PULs function in tandem in order to complete mannan metabolization. The organization of the proposed model is typical of mannan assimilation pathway and regulation. Each of the two PULs contains a GH5 β-mannanase, Man5A and Man5B. First, Man5A, a membrane anchored protein expressed at basal level, produces large manno-oligosaccharides, which are sensed by transcriptional regulators encoded by both mannan PULs (Cann et al. 1999). These also encode for permeases, which are probably involved in the import of these manno-oligosaccharides. The Man5B protein, the other β-mannanase, lacks a signal peptide, which also suggests that the subsequent steps of mannan metabolization occur intracellularly. Interestingly, each PUL contains a putative β-manno-oligosaccharide phosphorylases belonging to the recently created GH130 family.

Another mannoside degrading pathway by soil bacteria, also involving GH130 enzymes, was recently discovered in *Thermoanaerobacter* sp.X-514, a soil bacteria specific for β-1,2 linked manno-oligosaccharides (Chiku et al. 2014). The genomic environment of the two GH130 encoding genes lead Chiku *et al.* to suggest that the two β-1,2-oligomannan phosphorylases are part of a salvage pathway for GDP-mannose biosynthesis, as they are surrounded by a GH5 β-glycoside hydrolase and a GT4 mannosyltransferase (fig. 19 C2). Indeed, in the proposed model, an ABC transporter found in the same gene cluster would

import β-1,2 manno-oligosaccharides gathered from phosphopeptidomannan found in the yeast *C. albicans*, or from intracellular β-1-2 manno-oligosaccharides of *Leishmania mexicana* (Ralton et al. 2003). The specificity of these GH130 β-1,2 mannobiose and β-1,2 manno-oligosaccharides phosphorylases, named Teth514_1789 and Teth514_1788 respectively, would allow for production of α-D-Mannose-1-phosphate from the imported β-1,2 manno-oligosaccharides, thus feeding the GT4 GDP-mannosyltransferase. This study constitutes the sole reported example of mannoside degradation mechanism coupled to a biosynthetic pathway.

The GH130 mannosyl-glucose phosphorylase *Rm*MGP, isolated from the alkaline hot spring *Rodothermus marinus* ATCC 43812 bacterium (Jaito et al. 2014), degrades the β-1,4 bond of the β-D-man*p*-1,4-D-glc (Man-Glc) motif of plant glucomannans. Based on the the analysis of the *Rm*MGP containing PUL, *Rm*MGP would be involved in β-mannan degradation, similarly to what proposed for their *Bacteroides fragilis* and *Rumminococcus albus* (human gut and ruminal bacteria) horthologs, we are about to discuss. The genomic environment is conserved between these bacteria, and contains a cellobiose 2-epimerase (CE) and two GH26 mannanases. In this model (fig. 19 A1), β-1,4-mannobiose units are generated from mannans by GH26 mannanases, which are imported through a conserved transporter and subsequently converted into Man-Glc by the CE. The Man-Glc disaccharide is degraded by *Rm*MGP into Man1P and glucose, which feeds the central metabolism. The generated Man1P molecules are then converted into mannose-6-phosphate by phosphomannomutase, then into fructose-6-phosphate by phosphomannose isomerase.

3.4.2. Plant associated bacteria

In addition to the hemicellulosic β-mannans, plant associated bacteria also have a privilegiate acces to plant N-glycans. Synergic processes are also required in order to efficiently degrade these complex structures. Recently, the full characterization of the bacterial plant pathogen *Xanthomonas campestris pv. Campestris* ATCC 33913 N-glycan degradation PUL-like operon has been published (Dupoiron et al. 2015), demonstrating *in vitro* the sequencial involvement of glycoside hydrolases in N-glycan degradation.

This PUL-like structure is composed of eight genes organized as an operon, encoding for the eight NixE to NixL proteins involved in the synergetic degradation of the plant type, α-1,3 fucosylated, β-1,2-xylosylated $Man_3XylGlNAc_2Fuc$ N-glycan. In this study, the eight targeted genes were cloned and expressed, and individually characterized, in order to elucidate their precise role in the deconstruction of this complex glycan. These analyses conducted to the assignement of α-fucosidase, β-*N*-acetylglucosaminidase, β-mannosidase, β-xylosidase, α-mannosidase and β-galactosidase activities to each of the Nix proteins. All Nix enzymes contain a signal peptide, but characterization of type-II secretion system

mutants revealed that NixG is intracellular, while NixK and NixI are the sole excreted glycoside hydrolases of this system, the other ones being periplasmic. Basing on this very complete analysis, the authors proposed a functional model for plant N-glycan degradation (fig. 19 A2). In this model, the extracellular GH92 NixK removes the α-1,3 mannosyl residue, followed by removal of the β-1,2 xylosyl residue by the GH3 NixI. Then, contrary to what happens for human pathogens, a hypothetical asparaginase, Aspg, would release the glycan from the protein. TonB-dependent transporters would import the $Man_2GlcNAc_2Fuc$ into the perisplam for further deconstruction. The GH125 NixJ would remove the α-1,6 mannosyl residue, followed by action of the GH2 β-1,4 mannanase. The resulting monomers and chitobiose molecules would be translocated to the cytosol through specific transporters, where NixG would hydrolyse the chitobiose molecule. Ultimately, the released GlcNAc molecules would enhance the expression of the Nix operon.

3.4.3. Mammal gut bacteria

Glycan degradation by gut bacteria has been extensively studied. In particular, the human gut microbiota has been at the centre of attention these last years, as it is directly involved in human health. *Bacteroidetes* gut bacteria are prominent glycan degraders in this ecosystem, due to the extreme diversity of CAZymes they produce, allowing them for feeding on various diet and host polysaccharides (Martens et al. 2009a).

As the other bacteria cited above, mammal gut bacteria have a direct acces to plant mannosides, which are part of 'dietary fibers'. In the human gut, dietary mannans are mainly found in the cell wall of grains and nuts. Several examples of plant β-mannan degradation pathways have been reported in the literature. For instance, *Bacteroides fragilis* NCTC 9343, a prominant human gut bacterium, possesses a PUL dedicated to mannan assimilation. Constituting an operon, the *BF0771-BF0774* genes encode for the putative GH26 mannanase ManA, the GH130 mannosylglucose phosphorylase *Bf*MGP (Senoura et al. 2011), a putative sugar/cation symporter, and a cellobiose 2-epimerase (Ojima et al. 2011). As for *Rm*MGP, *Bf*MGP phosphorolyzes β-D-Man*p*-1,4-D-Glc into Man1P and glucose in presence of P_i. This study, which leads to the creation of the GH130 family in the CAZy database, unravelled a unique mannan assimilation pathway. The authors proposed a model similar to the one decribed for *Rm*MGP, in which the GH26 mannanase produces mannobiose units from mannan degradation, which are translocated by the BF0073 symporter to undergo subsequent epimerization into Man-Glc, followed by its phosphorolysis into Man1P and glucose. Then, the end products reach the central metabolim either directly (for glucose), or after being converted into Mannose-6-Phosphate and Fructose-6-Phosphate by phosphomanno mutase and phosphomanno isomerase (fig. 19 A1).

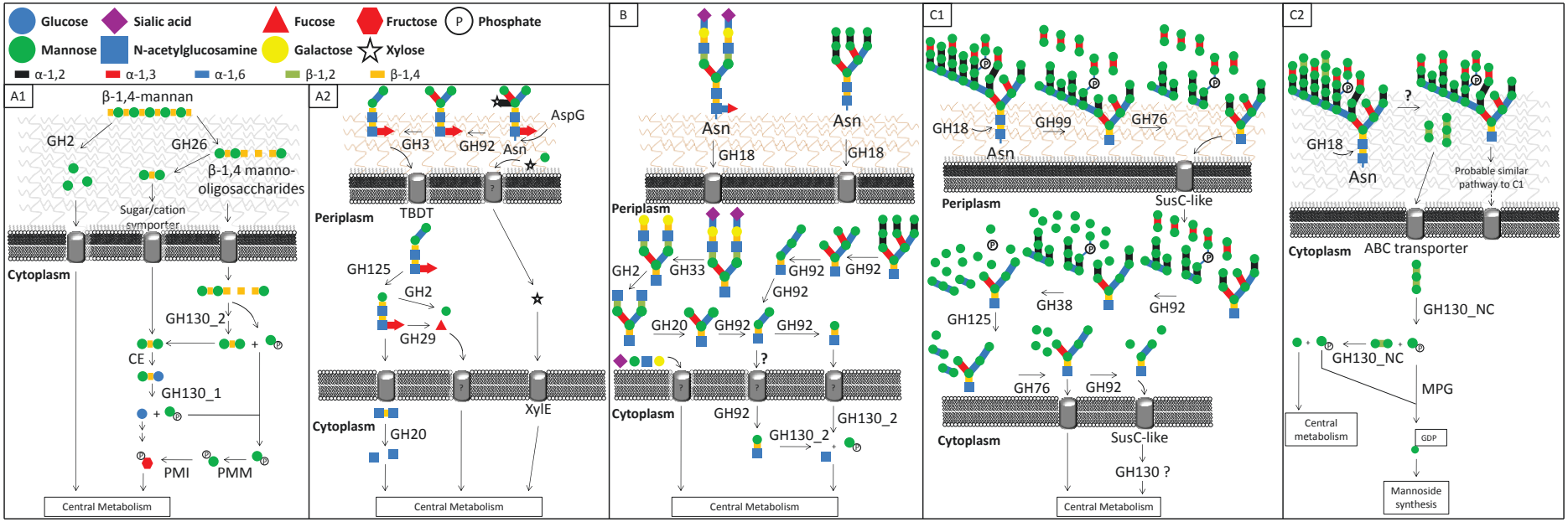


Figure 19: Pathways for mannoside degradation.

A1. Plant β-1,4-Mannan degradation by a Gram positive bacterium (example : *Rumminococcus albus* 7 (Kawahara et al. 2012)), A2. Plant N-glycan degradation by a Gram negative bacterium (example: *Xanthomonas campestris pv. campestris* (Dupoiron et al. 2015)). B. Human High mannose and complex N-glycans degradation by Gram negative bacteria (example *Bacteroides fragilis NCTC 9343* (Nihira et al. 2013, Ladevèze et al. 2013)).

C1. *S. cerevisisae* N-glycan degradation by a Gram negative bacterium (example: *Bacteroides thetaiotaomicron* VPI-5482 (Cuskin et al. 2015)), C2. *C. albicans* β-1,2-Mannoside containing N-glycan degradation coupled with GDP-Man synthesis, for a Gram positive bacterium (example : *Thermoanaerobacter sp.X514* (Chiku et al. 2014)).

CE: Cellobiose 2-Epimerase; PMM: PhosphoMannose Mutase; PMI: PhosphoMannose Isomerase; FruP: Fructose-6-phosphate; MPG: MannosylPhosphate Glucuronyltransferase, TBDT: TonB Dependant Transporter.

Such a pathway has then be identified in other mammal gut bacteria, like *Ruminococcus albus* 7, a ruminal anaerobic bacterium which efficiently degrades plant β-mannan using two synergistic GH130 mannoside phosphorylases (Kawahara et al. 2012b). These two enzymes, *Ra*MP1 and *Ra*MP2, act onto β-D-manp-1,4-D-glc and β-1,4-manno-oligosaccharides of DP<4 respectively. *Ra*MP1, the *Bf*MGP hortholog, is thought to participate to the same mannan degradation pathway together with a GH26 β-mannanase and a cellobiose 2-epimerase (fig. 19 A1). *Ra*MP2, on the contrary, appeared to be different. In addition to be able to process β-1,4 linked manno-oligosaccharides until DP4, it showed a much more wide tolerance to acceptor sugars in reverse-phosphorolysis reactions. Although it is able to phosphorolyse Man-Glc, its natural substrate was demonstrated to be the imported β-1,4 linked manno-oligosaccharides generated by the GH26 mannanase. These manno-oligosaccharides would therefore be processed by *Ra*MP2 to yield α-D-mannose-1-phosphate and β-1,4 linked manno-oligosaccharides of reduced chain length, ultimately β-1,4-mannobiose, converted into Man-Glc and then processed by *Ra*MP1.

Relying on their host to ensure their survival, gut bacteria developed other strategies to survive in case of host's starvation. When diet components are lacking, gut bacteria such as *Bacteroides thetaiotaomicron* VPI-5482 have been demonstrated to mine on host glycans, especially those found on the heavily glycosylated mucins proteins continuously secreted by the host's epithelium (Martens et al. 2008a).

In addition to plant hemicellulosic mannans, many gut bacteria are able to feed on host N-glycans. Recently several studies targeting N-glycan metabolisation by human gut bacteria were published, allowing to revisitate this process, which might be correlated to inflammatory bowel diseases. Here again, these studies revealed that N-glycan metabolisation both involves glycoside hydrolases and phosphorylases (Renzi et al. 2011; Nihira et al. 2013b; Ladevèze et al. 2013).

In particular, several Bacteroidetes bacteria harbor GH130 enzymes involved in N-glycan metabolization. The human gut symbiont *B. thetaiotaomicron* VPI-5482 possesses a mannoside phosphorylase specific for the disaccharide β-D-man*p*-1,4-D-Glc*p*NAc, a N-glycan core motif. The analysis of the sourrounding genes present in the *Bt*1033 containing PUL revealed the presence of four GH92, GH20, GH33 and GH18 glycoside hydrolases assigned to α-mannosidase, β-*N*-acetylhexosaminidase, exo-α-sialidase, and endo-β-*N*-acetylglucosaminidase activities, which would act synergistically to completely breakdown host N-glycans. In addition, this gene cluster encodes a major facilitator superfamily transporter, a protein of unknown function, an outer membrane lipoprotein and a TonB-dependent oligosaccharide transporter. Based on these elements, Nihira *et al*., elaborated a model of N-glycan metabolization (fig. 19B). The presence of a putative sialidase, together

with endo-β-*N*-acetylglucosaminidase activities, orientated the authors to a specific degradation of complex type host N-glycans. The proposed model is therefore in complete accordance with the characteristics of Sus-like systems encountered in Bacteroidetes species (Terrapon and Henrissat 2014b). First, the N-glycan would be bound by the SusC/D homologs and released from the protein by action of GH18 endo-β-*N*-acetylglucosaminidase. Contrary to plant N-glycan degradation, the deglycosylation of the host protein woud keep one *N*-acetylglucosamine residues bound onto the protein. Sequestration of the N-glycan into the periplasm would allow for subsequent deconstruction by the action of whether characterized GH92 α-mannosidase or the hypothesized α-sialidase, β-galactosidase and β-*N*-acetylhexosaminidase (Zhu et al. 2010). The resulting β-D-Man*p*-1,4-GlcNAc disaccharide would finally be internalized by a predicted transporter before being processed by the GH130 (fig. 19B).

Analysis of GH130 containing PUL-like structures at the scale of the human gut microbiome allowed for the redefinition of the GH130 family in three sequence clusters. The GH130_1 subfamily, which contains *Ra*MP1, *Rm*MGP, and *Bf*MGP is specifically encountered in PULs containing GH5 mannanases and GH26 mannosidases, and was therefore associated with plant β-mannan degradation. On the contrary, the GH130_2 subfamily, which contains *Bt*1033 and Uhgb_MP, another mannoside phosphorylase belonging to an unknown human gut bacterium related to *Bacteroidetes*, are contained in PULs together with GH families harbouring the required activities necessary for degradation of mature and immature complex and high mannose type N-glycans, namely GH92 α-mannosidases, GH18 *N*-acetyglucosaminidase or GH97 α-glucosidases. Specifically, Uhgb_MP, identified by high throughput functional screening of the human gut metagenome (Tasse et al. 2010b), presented high affinity for β-D-man*p*-1,4-D-Glc*p*NAc and β-D-man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc, the core trisaccharide of all N-glycans (Ladevèze et al. 2013) (fig. 19B). Intriguingly, this enzyme showed wide promiscuous reactions, being efficient in the degradation and synthesis by reverse-phosphorolysis of various β-1,4 manno-oligosaccharides, and is the sole known mannoside phosphoryase active on β-mannan. All the other GH130 sequences, which do not show sufficient homology to be grouped under a unique subfamily, are grouped in a non-classified group of sequences (GH130_NC). These genes are contained in PULs associating mixed functional synergetic activities, indicating that other mannoside degradation pathways exist (including the β-1,2 mannoside degradation pathway of *Thermoanaerobacter* sp. X-514, which falls into this group).

N-glycan foraging also occurs in mammal oral cavity. Indeed, the pathogen *Capnocytophaga canimorsus* 5 possesses a large transmembranary multi-protein complex involved in deglycosylation of complex N-glycans held on human IgG (Renzi et al. 2011). This organism contains a PUL encoding for 5 proteins, GpdC to GpdG, sharing homology with Sus

binding proteins. All are membrane anchored proteins associated in a stable complex. This complex has been demonstrated to deglycosylate human IgG *in vitro*, the catalytic protein being GpdG, a *N*-acetylglucosaminidase. One of the multiproteic constituent, GpdC, is a porin-like protein, probably involved in import of the released N-glycan to the periplasmic space. In addition, SiaC, a periplasm orientated membrane anchored sialidase would remove the capping sialic acids of the complex and hybrid N-glycans. Still unidentified periplasmic glycoside hydrolases would likely hydrolyze the internalized N-glycan.

Finally, the human gut bacterium *B. Thetaiotaomicron* VPI-5482 was also found to feed on the yeast mannan, a component of the human diet (Cuskin et al. 2015b), thanks to highly complex enzymatic machinery encoded by three PULs. The biochemical characterization of the 15 proteins encoded by these PULs allowed elaborating the precise mechanism by which this bacterium is so efficient to metabolise yeast mannans in the human gut. These PULs, orchestrated around SusC/D homologs to bind and sequester mannosides, mainly code for GH92, GH76, GH99 or GH125 α-mannosidases. The degradation model is similar to the one of *Bt*1033 for complex N-glycan foraging (fig. 19 C1). However, in that case, many extracellular α-manosidases are involved in the limited, but sufficient, degradation of α-1,2 side chains, thus suppressing the sterical hindrance that would restrain GH18 endo-β-*N*-acetylglucosaminidase from releasing the glycans from the protein, and the subsequent translocation to the periplasm. Additional periplasmic α-1,2, α-1,3 and α-1,6 mannosidases have been specifically demonstrated to act in a sequential manner, in order to progressively release mannosyl units that are imported to the cytosol. Here again, these PULs also encode for still uncharacterized GH130 enzymes, which thus may be involved in yeast mannan degradation.

3.5. Conclusion

The widespread distribution of mannosides in nature is at the scale of their importance in biological processes. Being present in all reigns, these glycans are central elements of life, playing a key role as structural molecules or energy sources and in the cell-cell interplay.

To face their structural diversity, bacteria developed many complex catabolic pathways, involving various carbohydrate active enzymes to efficiently catalyze glycan assimilation and ensure metabolic regulation. These strategies, especially those designed by *Bacteroidetes* species, rely on dedicated PULs for complex glycan assimilation. These particular genomic loci, which are often exchanged between bacteria by horizontal gene transfers (Lozupone et al. 2008; Tasse et al. 2010b), are at the basis of their successful distribution in ecosystems. If the role of glycoside hydrolases in mannoside catabolism is known for many years, an increasing number of studies continuously reveal the involvement

of mannoside phosphorylases in plant, mammal, and probably also yeast mannoside breakdown. Many anaerobic bacteria rely on their specific use in order to optimize energy consumption during glycan metabolisation. Using phosphorolysis rather than hydrolysis may also serve other purposes. Indeed, *Bacteroides fragilis* and other *Bacteroidetes* species are lacking the phosphotransferase sugar import system, (Brigham and Malamy 2005), an active transport relying on the immediate phosphorylation of imported simple sugars which generates a continuous flow of metabolites towards the cell, preventing carbohydrate leakage. Therefore, phosphorylation appears as serving a dual interest, saving energy and maximizing carbohydrate speed entry. Moreover, sugar-1-phosphate molecules are pivotal metabolites linking catabolic processes to anabolism, as most of them are substrates of nucleotidyltransferases that yield nucleotide-activated sugars that can readily be used by bacteria to produce a broad array of glycans.

However, the role of glycoside phosphorylases in glycan foraging is probably underestimated, due to the difficulty to differenciate them from real glycoside hydrolases and Leloir glycosyltransferases by using only sequence-based functional genomics and metagenomics.

More generally, bacterial metabolization of mannosides is susciting an increasing interest, and continuously reveals novel pathways of mannose foraging. Particular efforts should be dedicated to the study of human glycan catabolism of pathogenic bacteria with the ultimate goal of controlling deleterious bacterial-host interactions.

Bibliography

Aebi M, Bernasconi R, Clerc S, Molinari M (2010) N-glycan structures: recognition and processing in the ER. Trends Biochem Sci 35:74–82. doi: 10.1016/j.tibs.2009.10.001

Agans R, Rigsbee L, Kenche H, Michail S, Khamis HJ, Paliy O (2011) Distal gut microbiota of adolescent children is different from that of adults: Gut microbiota of adolescents differs from that of adults. FEMS Microbiol Ecol 77:404–412. doi: 10.1111/j.1574-6941.2011.01120.x

André I, Potocki-Véronèse G, Barbe S, Moulis C, Remaud-Siméon M (2014) CAZyme discovery and design for sweet dreams. Curr Opin Chem Biol 19:17–24. doi: 10.1016/j.cbpa.2013.11.014

Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta 1473:4–8.

Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, Antolín M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A, Melo Minardi R, M'rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P (2011) Enterotypes of the human gut microbiome. Nature 473:174–180. doi: 10.1038/nature09944

Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI (2004) The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci U S A 101:15718–15723. doi: 10.1073/pnas.0407076101

Bansil R, Turner BS (2006) Mucin structure, aggregation, physiological functions and biomedical applications. Curr Opin Colloid Interface Sci 11:164–170. doi: 10.1016/j.cocis.2005.11.001

Barrett E, Ross RP, O'Toole PW, Fitzgerald GF, Stanton C (2012) γ-Aminobutyric acid production by culturable bacteria from the human intestine. J Appl Microbiol 113:411–417. doi: 10.1111/j.1365-2672.2012.05344.x

Bartsch H, Nair J (2006) Chronic inflammation and oxidative stress in the genesis and perpetuation of cancer: role of lipid peroxidation, DNA damage, and repair. Langenbecks Arch Surg Dtsch Ges Für Chir 391:499–510. doi: 10.1007/s00423-006-0073-1

Belkaid Y, Bouladoux N, Hand TW (2013) Effector and memory T cell responses to commensal bacteria. Trends Immunol 34:299–306. doi: 10.1016/j.it.2013.03.003

Béra-Maillet C, Mosoni P, Kwasiborski A, Suau F, Ribot Y, Forano E (2009) Development of a RT-qPCR method for the quantification of Fibrobacter succinogenes S85 glycoside hydrolase transcripts in the rumen content of gnotobiotic and conventional sheep. J Microbiol Methods 77:8–16. doi: 10.1016/j.mimet.2008.11.009

Bettler E, Samain E, Chazalet V, Bosso C, Heyraud A, Joziasse DH, Wakarchuk WW, Imberty A, Geremia AR (1999) The living factory: in vivo production of N-acetyllactosamine containing carbohydrates in E. coli. Glycoconj J 16:205–212.

Beverley SM, Turco SJ (1998) Lipophosphoglycan (LPG) and the identification of virulence genes in the protozoan parasite Leishmania. Trends Microbiol 6:35–40. doi: 10.1016/S0966-842X(97)01180-3

Bjursell MK, Martens EC, Gordon JI (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. doi: 10.1074/jbc.M606509200

Bolam DN, Koropatkin NM (2012) Glycan recognition by the Bacteroidetes Sus-like systems. Curr Opin Struct Biol 22:563–569. doi: 10.1016/j.sbi.2012.06.006

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781. doi: 10.1042/BJ20040892

Boraston AB, Revett TJ, Boraston CM, Nurizzo D, Davies GJ (2003) Structural and thermodynamic dissection of specific mannan recognition by a carbohydrate binding module, TmCBM27. Struct Lond Engl 1993 11:665–675.

Borner GHH, Lilley KS, Stevens TJ, Dupree P (2003) Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. Plant Physiol 132:568–577. doi: 10.1104/pp.103.021170

Bouckaert J, Berglund J, Schembri M, De Genst E, Cools L, Wuhrer M, Hung C-S, Pinkner J, Slättegård R, Zavialov A, Choudhury D, Langermann S, Hultgren SJ, Wyns L, Klemm P, Oscarson S, Knight SD, De Greve H (2005) Receptor binding studies disclose a novel class of high-affinity inhibitors of the Escherichia coli FimH adhesin. Mol Microbiol 55:441–455. doi: 10.1111/j.1365-2958.2004.04415.x

Bouskra D, Brézillon C, Bérard M, Werts C, Varona R, Boneca IG, Eberl G (2008) Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. Nature 456:507–510. doi: 10.1038/nature07450

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254.

Bravo JA, Forsythe P, Chew MV, Escaravage E, Savignac HM, Dinan TG, Bienenstock J, Cryan JF (2011) Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. Proc Natl Acad Sci U S A 108:16050–16055. doi: 10.1073/pnas.1102999108

Brigham CJ, Malamy MH (2005) Characterization of the RokA and HexA broad-substrate-specificity hexokinases from Bacteroides fragilis and their role in hexose and N-acetylglucosamine utilization. J Bacteriol 187:890–901. doi: 10.1128/JB.187.3.890-901.2005

Brugman S, Klatter FA, Visser JTJ, Wildeboer-Veloo ACM, Harmsen HJM, Rozing J, Bos NA (2006) Antibiotic treatment partially protects against type 1 diabetes in the Bio-Breeding diabetes-prone rat. Is the gut flora involved in the development of type 1 diabetes? Diabetologia 49:2105–2108. doi: 10.1007/s00125-006-0334-0

Buckeridge MS (2010) Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. Plant Physiol 154:1017–1023. doi: 10.1104/pp.110.158642

Buettner FFR, Ashikov A, Tiemann B, Lehle L, Bakker H (2013) C. elegans DPY-19 is a C-mannosyltransferase glycosylating thrombospondin repeats. Mol Cell 50:295–302. doi: 10.1016/j.molcel.2013.03.003

Buser R, Lazar Z, Kaser S, Kunzler M, Aebi M (2010) Identification, Characterization, and Biosynthesis of a Novel N-Glycan Modification in the Fruiting Body of the Basidiomycete Coprinopsis cinerea. J Biol Chem 285:10715–10723. doi: 10.1074/jbc.M109.076075

Byrd MS, Sadovskaya I, Vinogradov E, Lu H, Sprinkle AB, Richardson SH, Ma L, Ralston B, Parsek MR, Anderson EM, Lam JS, Wozniak DJ (2009) Genetic and biochemical analyses of the Pseudomonas aeruginosa Psl exopolysaccharide reveal overlapping roles for polysaccharide synthesis enzymes in Psl and LPS production. Mol Microbiol 73:622–638. doi: 10.1111/j.1365-2958.2009.06795.x

Calo D, Kaminski L, Eichler J (2010) Protein glycosylation in Archaea: Sweet and extreme. Glycobiology 20:1065–1076. doi: 10.1093/glycob/cwq055

Cameron EA, Kwiatkowski KJ, Lee B-H, Hamaker BR, Koropatkin NM, Martens EC (2014) Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan competition in the gut by separately promoting enhanced sensing and catalysis. mBio 5:e01441–01414. doi: 10.1128/mBio.01441-14

Cameron EA, Maynard MA, Smith CJ, Smith TJ, Koropatkin NM, Martens EC (2012a) Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. J Biol Chem 287:34614–34625. doi: 10.1074/jbc.M112.397380

Cameron EA, Maynard MA, Smith CJ, Smith TJ, Koropatkin NM, Martens EC (2012b) Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. J Biol Chem 287:34614–34625. doi: 10.1074/jbc.M112.397380

Campagnolo M, Campa C, Zorzi RD, Wuerges J, Geremia S (2008) X-ray studies on ternary complexes of maltodextrin phosphorylase. Arch Biochem Biophys 471:11–19. doi: 10.1016/j.abb.2007.11.023

Cann IK, Kocherginskaya S, King MR, White BA, Mackie RI (1999) Molecular cloning, sequencing, and expression of a novel multidomain mannanase gene from Thermoanaerobacterium polysaccharolyticum. J Bacteriol 181:1643–1651.

Cao Y, Rocha ER, Smith CJ (2014) Efficient utilization of complex N-linked glycans is a selective advantage for Bacteroides fragilis in extraintestinal infections. Proc Natl Acad Sci U S A. doi: 10.1073/pnas.1407344111

Capron A, Gourgues M, Neiva LS, Faure J-E, Berger F, Pagnussat G, Krishnan A, Alvarez-Mejia C, Vielle-Calzada J-P, Lee Y-R, Liu B, Sundaresan V (2008) Maternal control of male-gamete delivery in Arabidopsis involves a putative GPI-anchored protein encoded by the LORELEI gene. Plant Cell 20:3038–3049. doi: 10.1105/tpc.108.061713

Cecchini DA, Laville E, Laguerre S, Robe P, Leclerc M, Doré J, Henrissat B, Remaud-Siméon M, Monsan P, Potocki-Véronèse G (2013) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. PloS One 8:e72766. doi: 10.1371/journal.pone.0072766

Chaban B, Voisin S, Kelly J, Logan SM, Jarrell KF (2006) Identification of genes involved in the biosynthesis and attachment of Methanococcus voltae N-linked glycans: insight into N-linked glycosylation pathways in Archaea. Mol Microbiol 61:259–268. doi: 10.1111/j.1365-2958.2006.05226.x

Champion E, Guérin F, Moulis C, Barbe S, Tran TH, Morel S, Descroix K, Monsan P, Mourey L, Mulard LA, Tranier S, Remaud-Siméon M, André I (2012) Applying pairwise combinations of amino acid mutations for sorting out highly efficient glucosylation tools for chemo-enzymatic synthesis of bacterial oligosaccharides. J Am Chem Soc 134:18677–18688. doi: 10.1021/ja306845b

Chang A, Singh S, Phillips GN, Thorson JS (2011) Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. Curr Opin Biotechnol 22:800–808. doi: 10.1016/j.copbio.2011.04.013

Chekan JR, Kwon IH, Agarwal V, Dodd D, Revindran V, Mackie RI, Cann I, Nair SK (2014) Structural and Biochemical Basis for Mannan Utilization by Caldanaerobius polysaccharolyticus Strain ATCC BAA-17. J Biol Chem 289:34965–34977. doi: 10.1074/jbc.M114.579904

Chen C, Van der Borght J, De Vreese R, D'hooghe M, Soetaert W, Desmet T (2014) Engineering the specificity of trehalose phosphorylase as a general strategy for the production of glycosyl phosphates. Chem Commun Camb Engl 50:7834–7836. doi: 10.1039/c4cc02202e

Cherniak R, Sundstrom JB (1994) Polysaccharide antigens of the capsule of Cryptococcus neoformans. Infect Immun 62:1507–1512.

Chesebro B, Trifilo M, Race R, Meade-White K, Teng C, LaCasse R, Raymond L, Favara C, Baron G, Priola S, Caughey B, Masliah E, Oldstone M (2005) Anchorless prion protein results in infectious amyloid disease without clinical scrapie. Science 308:1435–1439. doi: 10.1126/science.1110837

Chiku K, Nihira T, Suzuki E, Nishimoto M, Kitaoka M, Ohtsubo K, Nakai H (2014) Discovery of Two β-1,2-Mannoside Phosphorylases Showing Different Chain-Length Specificities from Thermoanaerobacter sp. X-514. PloS One 9:e114882. doi: 10.1371/journal.pone.0114882

Cho KH, Cho D, Wang GR, Salyers AA (2001) New regulatory gene that contributes to control of Bacteroides thetaiotaomicron starch utilization genes. J Bacteriol 183:7198–7205. doi: 10.1128/JB.183.24.7198-7205.2001

Christensen HR, Frøkiaer H, Pestka JJ (2002) Lactobacilli differentially modulate expression of cytokines and maturation surface markers in murine dendritic cells. J Immunol Baltim Md 1950 168:171–178.

Cobucci-Ponzano B, Perugino G, Strazzulli A, Rossi M, Moracci M (2012) Thermophilic glycosynthases for oligosaccharides synthesis. Methods Enzymol 510:273–300. doi: 10.1016/B978-0-12-415931-0.00015-X

Conzelmann A, Puoti A, Lester RL, Desponds C (1992) Two different types of lipid moieties are present in glycophosphoinositol-anchored membrane proteins of Saccharomyces cerevisiae. EMBO J 11:457–466.

Coutinho P, Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. In: Recent advances in Carbohydrate Bioengineering, Gilbert HJ, Davies G, Henrissat B, Svensson B, editors. Cambridge: The Royal Society of Chemistry, pp p. 3–12

Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. J Mol Biol 328:307–317.

Crossman L, Dow JM (2004) Biofilm formation and dispersal in Xanthomonas campestris. Microbes Infect Inst Pasteur 6:623–629. doi: 10.1016/j.micinf.2004.01.013

Cryan JF, Kaupmann K (2005) Don't worry "B" happy!: a role for GABA(B) receptors in anxiety and depression. Trends Pharmacol Sci 26:36–43. doi: 10.1016/j.tips.2004.11.004

Cui L, Neoh H, Iwamoto A, Hiramatsu K (2012) Coordinated phenotype switching with large-scale chromosome flip-flop inversion observed in bacteria. Proc Natl Acad Sci U S A 109:E1647–1656. doi: 10.1073/pnas.1204307109

Cummings JH, Stephen AM (2007) Carbohydrate terminology and classification. Eur J Clin Nutr 61 Suppl 1:S5–18. doi: 10.1038/sj.ejcn.1602936

Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Pudlo NA, Porter NT, Urs K, Thompson AJ, Cartmell A, Rogowski A, Hamilton BS, Chen R, Tolbert TJ, Piens K, Bracke D, Vervecken W, Hakki Z, Speciale G, Munōz-Munōz JL, Day A, Peña MJ, McLean R, Suits MD, Boraston AB, Atherly T, Ziemer CJ, Williams SJ, Davies GJ, Abbott DW, Martens EC, Gilbert HJ (2015a) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature 517:165–169. doi: 10.1038/nature13995

Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Pudlo NA, Porter NT, Urs K, Thompson AJ, Cartmell A, Rogowski A, Hamilton BS, Chen R, Tolbert TJ, Piens K, Bracke D, Vervecken W, Hakki Z, Speciale G, Munōz-Munōz JL, Day A, Peña MJ, McLean R, Suits MD, Boraston AB, Atherly T, Ziemer CJ, Williams SJ, Davies GJ, Abbott DW, Martens EC, Gilbert HJ (2015b) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature 517:165–169. doi: 10.1038/nature13995

Daudé D, Topham CM, Remaud-Siméon M, André I (2013) Probing impact of active site residue mutations on stability and activity of Neisseria polysaccharea amylosucrase. Protein Sci Publ Protein Soc 22:1754–1765. doi: 10.1002/pro.2375

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ (2014) Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563. doi: 10.1038/nature12820

De Groot PWJ, Ram AF, Klis FM (2005) Features and functions of covalently linked proteins in fungal cell walls. Fungal Genet Biol FG B 42:657–675. doi: 10.1016/j.fgb.2005.04.002

Dejea C, Wick E, Sears CL (2013) Bacterial oncogenesis in the colon. Future Microbiol 8:445–460. doi: 10.2217/fmb.13.17

Descoteaux A, Turco SJ (1999) Glycoconjugates in Leishmania infectivity. Biochim Biophys Acta 1455:341–352.

Desmet T, Soetaert W, Bojarová P, Křen V, Dijkhuizen L, Eastwick-Field V, Schiller A (2012) Enzymatic glycosylation of small molecules: challenging substrates require tailored catalysts. Chem Weinh Bergstr Ger 18:10786–10801. doi: 10.1002/chem.201103069

Dharmani P, Srivastava V, Kissoon-Singh V, Chadee K (2009) Role of intestinal mucins in innate host defense mechanisms against pathogens. J Innate Immun 1:123–135. doi: 10.1159/000163037

Dobos KM, Khoo KH, Swiderek KM, Brennan PJ, Belisle JT (1996) Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of Mycobacterium tuberculosis. J Bacteriol 178:2498–2506.

Domozych DS, Ciancia M, Fangel JU, Mikkelsen MD, Ulvskov P, Willats WGT (2012) The Cell Walls of Green Algae: A Journey through Evolution and Diversity. Front Plant Sci 3:82. doi: 10.3389/fpls.2012.00082

Donovan SM, Wang M, Li M, Friedberg I, Schwartz SL, Chapkin RS (2012) Host-Microbe Interactions in the Neonatal Intestine: Role of Human Milk Oligosaccharides. Adv Nutr Int Rev J 3:450S–455S. doi: 10.3945/an.112.001859

Doucey MA, Hess D, Blommers MJ, Hofsteenge J (1999) Recombinant human interleukin-12 is the second example of a C-mannosylated protein. Glycobiology 9:435–441.

Drissi F, Merhej V, Angelakis E, El Kaoutari A, Carrière F, Henrissat B, Raoult D (2014) Comparative genomics analysis of Lactobacillus species associated with weight gain or weight protection. Nutr Diabetes 4:e109. doi: 10.1038/nutd.2014.6

Duca FA, Sakar Y, Lepage P, Devime F, Langelier B, Doré J, Covasa M (2014) Replication of obesity and associated signaling pathways through transfer of microbiota from obese-prone rats. Diabetes 63:1624–1636. doi: 10.2337/db13-1526

Duffaud GD, McCutchen CM, Leduc P, Parker KN, Kelly RM (1997) Purification and characterization of extremely thermostable beta-mannanase, beta-mannosidase, and alpha-galactosidase from the hyperthermophilic eubacterium Thermotoga neapolitana 5068. Appl Environ Microbiol 63:169–177.

Dukowicz AC, Lacy BE, Levine GM (2007) Small intestinal bacterial overgrowth: a comprehensive review. Gastroenterol Hepatol 3:112–122.

Dupoiron S, Zischek C, Ligat L, Carbonne J, Boulanger A, Dugé de Bernonville T, Lautier M, Rival P, Arlat M, Jamet E, Lauber E, Albenne C (2015) The N-glycan Cluster from Xanthomonas campestris pv. campestris: a Toolbox for Sequential Plant N-glycan Processing. J Biol Chem. doi: 10.1074/jbc.M114.624593

Eichler J, Adams MWW (2005) Posttranslational Protein Modification in Archaea. Microbiol Mol Biol Rev 69:393–425. doi: 10.1128/MMBR.69.3.393-425.2005

Eisenhaber B, Bork P, Eisenhaber F (2001) Post-translational GPI lipid anchor modification of proteins in kingdoms of life: analysis of protein sequence data from complete genomes. Protein Eng 14:17–25.

El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol 11:497–504. doi: 10.1038/nrmicro3050

Ellis M, Egelund J, Schultz CJ, Bacic A (2010) Arabinogalactan-proteins: key regulators at the cell surface? Plant Physiol 153:403–419. doi: 10.1104/pp.110.156000

Falzarano D, Krokhin O, Van Domselaar G, Wolf K, Seebach J, Schnittler H-J, Feldmann H (2007) Ebola sGP--the first viral glycoprotein shown to be C-mannosylated. Virology 368:83–90. doi: 10.1016/j.virol.2007.06.015

Fankhauser C, Homans SW, Thomas-Oates JE, McConville MJ, Desponds C, Conzelmann A, Ferguson MA (1993) Structures of glycosylphosphatidylinositol membrane anchors from Saccharomyces cerevisiae. J Biol Chem 268:26365–26374.

Ferguson MA, Homans SW, Dwek RA, Rademacher TW (1988) Glycosyl-phosphatidylinositol moiety that anchors Trypanosoma brucei variant surface glycoprotein to the membrane. Science 239:753–759.

Ferreira C d. S, Piedade MTF, Tine MAS, Rossatto DR, Parolin P, Buckeridge MS (2009) The role of carbohydrates in seed germination and seedling establishment of Himatanthus sucuuba, an Amazonian tree with populations adapted to flooded and non-flooded conditions. Ann Bot 104:1111–1119. doi: 10.1093/aob/mcp212

Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, Oppermann U (2002) Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. J Biol Chem 277:25677–25684. doi: 10.1074/jbc.M202160200

Fletcher CM, Coyne MJ, Villa OF, Chatzidaki-Livanis M, Comstock LE (2009) A general O-glycosylation system important to the physiology of a major human intestinal symbiont. Cell 137:321–331. doi: 10.1016/j.cell.2009.02.041

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. doi: 10.1038/nrmicro1817

Frieman MB, Cormack BP (2003) The omega-site sequence of glycosylphosphatidylinositol-anchored proteins in Saccharomyces cerevisiae can determine distribution between the membrane and the cell wall. Mol Microbiol 50:883–896.

Fujinaka H, Nakamura J, Murase D, Souno H, Kobayashi H (2007) Use of monosaccharide posphates for improving the intestinal function. Google Patents

Fujita M, Jigami Y (2008) Lipid remodeling of GPI-anchored proteins and its function. Biochim Biophys Acta 1780:410–420. doi: 10.1016/j.bbagen.2007.08.009

Fujita M, Kinoshita T (2010) Structural remodeling of GPI anchors during biosynthesis and after attachment to proteins. FEBS Lett 584:1670–1677. doi: 10.1016/j.febslet.2009.10.079

Fukata M, Abreu MT (2008) Role of Toll-like receptors in gastrointestinal malignancies. Oncogene 27:234–243. doi: 10.1038/sj.onc.1210908

Furmanek A, Hofsteenge J (2000) Protein C-mannosylation: facts and questions. Acta Biochim Pol 47:781–789.

Fushinobu S (2010) Unique sugar metabolic pathways of bifidobacteria. Biosci Biotechnol Biochem 74:2374–2384. doi: 10.1271/bbb.100494

Gerritsen J, Smidt H, Rijkers GT, de Vos WM (2011) Intestinal microbiota in human health and disease: the impact of probiotics. Genes Nutr 6:209–240. doi: 10.1007/s12263-011-0229-7

Ghosh TS, Gupta SS, Nair GB, Mande SS (2013) In silico analysis of antibiotic resistance genes in the gut microflora of individuals from diverse geographies and age-groups. PloS One 8:e83823. doi: 10.1371/journal.pone.0083823

Gilleron M, Jackson M, Nigou J, Puzo G (2008) Structure, activities and biosynthesis of the Phosphatidyl-myo-Inositol-based lipoglycans. In: The Mycobacterial Cell Envelope. Daffe´ and J.-M. Reyrat, pp pp.75–105

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE (2006) Metagenomic analysis of the human distal gut microbiome. Science 312:1355–1359. doi: 10.1126/science.1124234

Goedl C, Griessler R, Schwarz A, Nidetzky B (2006) Structure-function relationships for Schizophyllum commune trehalose phosphorylase and their implications for the catalytic mechanism of family GT-4 glycosyltransferases. Biochem J 397:491–500. doi: 10.1042/BJ20060029

Goedl C, Nidetzky B (2009) Sucrose phosphorylase harbouring a redesigned, glycosyltransferase-like active site exhibits retaining glucosyl transfer in the absence of a covalent intermediate. Chembiochem Eur J Chem Biol 10:2333–2337. doi: 10.1002/cbic.200900429

Goedl C, Sawangwan T, Mueller M, Schwarz A, Nidetzky B (2008) A high-yielding biocatalytic process for the production of 2-O-(alpha-D-glucopyranosyl)-sn-glycerol, a natural osmolyte and useful moisturizing ingredient. Angew Chem Int Ed Engl 47:10086–10089. doi: 10.1002/anie.200803562

Goehler LE, Park SM, Opitz N, Lyte M, Gaykema RPA (2008) Campylobacter jejuni infection increases anxiety-like behavior in the holeboard: possible anatomical substrates for viscerosensory modulation of exploratory behavior. Brain Behav Immun 22:354–366. doi: 10.1016/j.bbi.2007.08.009

Gomord V, Fitchette A-C, Menu-Bouaouiche L, Saint-Jore-Dupas C, Plasson C, Michaud D, Faye L (2010) Plant-specific glycosylation patterns in the context of therapeutic protein production: PMP-specific glycosylation patterns. Plant Biotechnol J 8:564–587. doi: 10.1111/j.1467-7652.2009.00497.x

Goto M (2007) Protein O-glycosylation in fungi: diverse structures and multiple functions. Biosci Biotechnol Biochem 71:1415–1427. doi: 10.1271/bbb.70080

Goto Y, Niwa Y, Suzuki T, Dohmae N, Umezawa K, Simizu S (2014) C-mannosylation of human hyaluronidase 1: Possible roles for secretion and enzymatic activity. Int J Oncol. doi: 10.3892/ijo.2014.2438

Grass J, Pabst M, Kolarich D, Pöltl G, Léonard R, Brecker L, Altmann F (2011) Discovery and structural characterization of fucosylated oligomannosidic N-glycans in mushrooms. J Biol Chem 286:5977–5984. doi: 10.1074/jbc.M110.191304

Grześkowiak Ł, Collado MC, Mangani C, Maleta K, Laitinen K, Ashorn P, Isolauri E, Salminen S (2012) Distinct Gut Microbiota in Southeastern African and Northern European Infants: J Pediatr Gastroenterol Nutr 54:812–816. doi: 10.1097/MPG.0b013e318249039c

Grzymajło K, Ugorski M, Kolenda R, Kędzierska A, Kuźmińska-Bajor M, Wieliczko A (2013) FimH adhesin from host unrestricted Salmonella Enteritidis binds to different glycoprotein ligands expressed by enterocytes from sheep, pig and cattle than FimH adhesins from host restricted Salmonella Abortus-ovis, Salmonella Choleraesuis and Salmonella Dublin. Vet Microbiol 166:550–557. doi: 10.1016/j.vetmic.2013.07.004

Guan Z, Naparstek S, Calo D, Eichler J (2012) Protein glycosylation as an adaptive response in Archaea: growth at different salt concentrations leads to alterations in Haloferax volcanii S-layer glycoprotein N-glycosylation. Environ Microbiol 14:743–753. doi: 10.1111/j.1462-2920.2011.02625.x

Hall RA, Gow NAR (2013) Mannosylation in *Candida albicans* : role in cell wall function and immune recognition: *Candida* wall mannosylation. Mol Microbiol 90:1147–1161. doi: 10.1111/mmi.12426

Hamura K, Saburi W, Abe S, Morimoto N, Taguchi H, Mori H, Matsui H (2012) Enzymatic characteristics of cellobiose phosphorylase from Ruminococcus albus NE1 and kinetic mechanism of unusual substrate inhibition in reverse phosphorolysis. Biosci Biotechnol Biochem 76:812–818. doi: 10.1271/bbb.110954

Hamura K, Saburi W, Matsui H, Mori H (2013) Modulation of acceptor specificity of Ruminococcus albus cellobiose phosphorylase through site-directed mutagenesis. Carbohydr Res 379:21–25. doi: 10.1016/j.carres.2013.06.010

Hartmann M, Barsch A, Niehaus K, Pühler A, Tauch A, Kalinowski J (2004) The glycosylated cell surface protein Rpf2, containing a resuscitation-promoting factor motif, is involved in intercellular communication of Corynebacterium glutamicum. Arch Microbiol 182:299–312. doi: 10.1007/s00203-004-0713-1

Helenius A, Aebi M (2004) ROLES OF N-LINKED GLYCANS IN THE ENDOPLASMIC RETICULUM. Annu Rev Biochem 73:1019–1049. doi: 10.1146/annurev.biochem.73.011303.073752

Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, Davies G (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc Natl Acad Sci U S A 92:7090–7094.

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

Hidaka M, Honda Y, Kitaoka M, Nirasawa S, Hayashi K, Wakagi T, Shoun H, Fushinobu S (2004) Chitobiose phosphorylase from Vibrio proteolyticus, a member of glycosyl transferase family 36, has a clan GH-L-like (alpha/alpha)(6) barrel fold. Struct Lond Engl 1993 12:937–947. doi: 10.1016/j.str.2004.03.027

Hofsteenge J, Blommers M, Hess D, Furmanek A, Miroshnichenko O (1999) The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues. J Biol Chem 274:32786–32794.

Hongshu Z, Jinggan Y, Yan Z (2002) The glucomannan from ramie. Carbohydr Polym 47:83–86. doi: 10.1016/S0144-8617(00)00344-1

Hooper LV, Midtvedt T, Gordon JI (2002) How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu Rev Nutr 22:283–307. doi: 10.1146/annurev.nutr.22.011602.092259

Hooper LV, Stappenbeck TS, Hong CV, Gordon JI (2003) Angiogenins: a new class of microbicidal proteins involved in innate immunity. Nat Immunol 4:269–273. doi: 10.1038/ni888

Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI (2001) Molecular analysis of commensal host-microbial relationships in the intestine. Science 291:881–884. doi: 10.1126/science.291.5505.881

Huddleston JR (2014) Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. Infect Drug Resist 7:167–176. doi: 10.2147/IDR.S48820

Hungate RE (1969) Chapter {IV} A Roll Tube Method for Cultivation of Strict Anaerobes. In: Norris JR, Ribbons DW (eds). Academic Press, pp 117 – 132

Hwang S, Gwon S-Y, Kim MS, Lee S, Rhee K-J (2013) Bacteroides fragilis Toxin Induces IL-8 Secretion in HT29/C1 Cells through Disruption of E-cadherin Junctions. Immune Netw 13:213–217. doi: 10.4110/in.2013.13.5.213

Ikezawa H (2002) Glycosylphosphatidylinositol (GPI)-anchored proteins. Biol Pharm Bull 25:409–417.

Jahn M, Stoll D, Warren RAJ, Szabó L, Singh P, Gilbert HJ, Ducros VMA, Davies GJ, Withers SG (2003) Expansion of the glycosynthase repertoire to produce defined manno-oligosaccharides. Chem Commun Camb Engl 1327–1329.

Jaito N, Saburi W, Odaka R, Kido Y, Hamura K, Nishimoto M, Kitaoka M, Matsui H, Mori H (2014) Characterization of a thermophilic 4-O-β-D-mannosyl-D-glucose phosphorylase from Rhodothermus marinus. Biosci Biotechnol Biochem 78:263–270. doi: 10.1080/09168451.2014.882760

Jayani RS, Saxena S, Gupta R (2005) Microbial pectinolytic enzymes: A review. Process Biochem 40:2931–2944. doi: 10.1016/j.procbio.2005.03.026

Juge N (2012) Microbial adhesins to gastrointestinal mucus. Trends Microbiol 20:30–39. doi: 10.1016/j.tim.2011.10.001

Juste C, Kreil DP, Beauvallet C, Guillot A, Vaca S, Carapito C, Mondot S, Sykacek P, Sokol H, Blon F, Lepercq P, Levenez F, Valot B, Carre W, Loux V, Pons N, David O, Schaeffer B, Lepage P, Martin P, Monnet V, Seksik P, Beaugerie L, Ehrlich SD, Gibrat J-F, Van Dorsselaer A, Dore J (2014) Bacterial protein signals are associated with Crohn's disease. Gut. doi: 10.1136/gutjnl-2012-303786

Kaci G, Lakhdari O, Doré J, Ehrlich SD, Renault P, Blottière HM, Delorme C (2011) Inhibition of the NF-kappaB pathway in human intestinal epithelial cells by commensal Streptococcus salivarius. Appl Environ Microbiol 77:4681–4684. doi: 10.1128/AEM.03021-10

Kalmokoff ML, Koval SF, Jarrell KF (1992) Relatedness of the flagellins from methanogens. Arch Microbiol 157:481–487.

Kaminski L, Lurie-Weinberger MN, Allers T, Gophna U, Eichler J (2013) Phylogenetic- and genome-derived insight into the evolution of N-glycosylation in Archaea. Mol Phylogenet Evol 68:327–339. doi: 10.1016/j.ympev.2013.03.024

Kärcher U, Schröder H, Haslinger E, Allmaier G, Schreiner R, Wieland F, Haselbeck A, König H (1993) Primary structure of the heterosaccharide of the surface glycoprotein of Methanothermus fervidus. J Biol Chem 268:26821–26826.

Kaur D, Guerin ME, Škovierová H, Brennan PJ, Jackson M (2009) Chapter 2 Biogenesis of the Cell Wall and Other Glycoconjugates of Mycobacterium tuberculosis. In: Advances in Applied Microbiology. Elsevier, pp 23–78

Kaur D, McNeil MR, Khoo K-H, Chatterjee D, Crick DC, Jackson M, Brennan PJ (2007) New insights into the biosynthesis of mycobacterial lipomannan arising from deletion of a conserved gene. J Biol Chem 282:27133–27140. doi: 10.1074/jbc.M703389200

Kawahara R, Saburi W, Odaka R, Taguchi H, Ito S, Mori H, Matsui H (2012a) Metabolic Mechanism of Mannan in a Ruminal Bacterium, Ruminococcus albus, Involving Two Mannoside Phosphorylases and Cellobiose 2-Epimerase: DISCOVERY OF A NEW CARBOHYDRATE PHOSPHORYLASE, -1,4-MANNOOLIGOSACCHARIDE PHOSPHORYLASE. J Biol Chem 287:42389–42399. doi: 10.1074/jbc.M112.390336

Kawahara R, Saburi W, Odaka R, Taguchi H, Ito S, Mori H, Matsui H (2012b) Metabolic Mechanism of Mannan in a Ruminal Bacterium, Ruminococcus albus, Involving Two Mannoside Phosphorylases and Cellobiose 2-Epimerase: DISCOVERY OF A NEW CARBOHYDRATE PHOSPHORYLASE, -1,4-MANNOOLIGOSACCHARIDE PHOSPHORYLASE. J Biol Chem 287:42389–42399. doi: 10.1074/jbc.M112.390336

Kelly D, Campbell JI, King TP, Grant G, Jansson EA, Coutts AGP, Pettersson S, Conway S (2004) Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA. Nat Immunol 5:104–112. doi: 10.1038/ni1018

Kim E, Coelho D, Blachier F (2013) Review of the association between meat consumption and risk of colorectal cancer. Nutr Res N Y N 33:983–994. doi: 10.1016/j.nutres.2013.07.018

Kimura Y, Matsuo S (2000) Changes in N-linked oligosaccharides during seed development of Ginkgo biloba. Biosci Biotechnol Biochem 64:562–568.

Kim Y-K, Kitaoka M, Krishnareddy M, Mori Y, Hayashi K (2002) Kinetic studies of a recombinant cellobiose phosphorylase (CBP) of the Clostridium thermocellum YM4 strain expressed in Escherichia coli. J Biochem (Tokyo) 132:197–203.

Kittl R, Withers SG (2010) New approaches to enzymatic glycoside synthesis through directed evolution. Carbohydr Res 345:1272–1279. doi: 10.1016/j.carres.2010.04.002

Kobayashi T, Nishizaki R, Ikezawa H (1997) The presence of GPI-linked protein(s) in an archaeobacterium, Sulfolobus acidocaldarius, closely related to eukaryotes. Biochim Biophys Acta 1334:1–4.

Korea C-G, Ghigo J-M, Beloin C (2011) The sweet connection: Solving the riddle of multiple sugar-binding fimbrial adhesins in Escherichia coli: Multiple E. coli fimbriae form a versatile arsenal of sugar-binding lectins potentially involved in surface-colonisation

and tissue tropism. BioEssays News Rev Mol Cell Dev Biol 33:300–311. doi: 10.1002/bies.201000121

Ladevèze S, Tarquis L, Cecchini DA, Bercovici J, André I, Topham CM, Morel S, Laville E, Monsan P, Lombard V, Henrissat B, Potocki-Véronèse G (2013) Role of glycoside phosphorylases in mannose foraging by human gut bacteria. J Biol Chem 288:32370–32383. doi: 10.1074/jbc.M113.483628

Lagier J-C, Million M, Hugon P, Armougom F, Raoult D (2012) Human Gut Microbiota: Repertoire and Variations. Front Cell Infect Microbiol. doi: 10.3389/fcimb.2012.00136

Laine RA (1994) A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. Glycobiology 4:759–767.

Larkin A, Imperiali B (2011) The expanding horizons of asparagine-linked glycosylation. Biochemistry (Mosc) 50:4411–4426. doi: 10.1021/bi200346n

Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, Klinter S, Pudlo NA, Urs K, Koropatkin NM, Creagh AL, Haynes CA, Kelly AG, Cederholm SN, Davies GJ, Martens EC, Brumer H (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature 506:498–502. doi: 10.1038/nature12907

Lay C, Rigottier-Gois L, Holmstrøm K, Rajilic M, Vaughan EE, de Vos WM, Collins MD, Thiel R, Namsolleck P, Blaut M, Doré J (2005) Colonic microbiota signatures across five northern European countries. Appl Environ Microbiol 71:4153–4155. doi: 10.1128/AEM.71.7.4153-4155.2005

Lea-Smith DJ, Martin KL, Pyke JS, Tull D, McConville MJ, Coppel RL, Crellin PK (2008) Analysis of a New Mannosyltransferase Required for the Synthesis of Phosphatidylinositol Mannosides and Lipoarbinomannan Reveals Two Lipomannan Pools in Corynebacterineae. J Biol Chem 283:6773–6782. doi: 10.1074/jbc.M707139200

Lebeer S, Vanderleyden J, De Keersmaecker SCJ (2010) Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. Nat Rev Microbiol 8:171–184. doi: 10.1038/nrmicro2297

Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jørgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clément K, Doré J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, de Vos WM, Zucker J-D, Raes J, Hansen T, MetaHIT consortium, Bork P, Wang J, Ehrlich SD, Pedersen O (2013) Richness of human gut microbiome correlates with metabolic markers. Nature 500:541–546. doi: 10.1038/nature12506

Le Nours J, Anderson L, Stoll D, Stålbrand H, Lo Leggio L (2005) The structure and characterization of a modular endo-beta-1,4-mannanase from Cellulomonas fimi. Biochemistry (Mosc) 44:12700–12708. doi: 10.1021/bi050779v

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels 6:41. doi: 10.1186/1754-6834-6-41

Liepman AH, Nairn CJ, Willats WGT, Sørensen I, Roberts AW, Keegstra K (2007) Functional genomic analysis supports conservation of function among cellulose synthase-like a gene family members and suggests diverse roles of mannans in plants. Plant Physiol 143:1881–1893. doi: 10.1104/pp.106.093989

Li L-L, McCorkle SR, Monchy S, Taghavi S, van der Lelie D (2009) Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. Biotechnol Biofuels 2:10. doi: 10.1186/1754-6834-2-10

Liu L, Shang-Guan K, Zhang B, Liu X, Yan M, Zhang L, Shi Y, Zhang M, Qian Q, Li J, Zhou Y (2013) Brittle Culm1, a COBRA-like protein, functions in cellulose assembly through binding cellulose microfibrils. PLoS Genet 9:e1003704. doi: 10.1371/journal.pgen.1003704

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 25:402–408. doi: 10.1006/meth.2001.1262

Löffler A, Doucey MA, Jansson AM, Müller DR, de Beer T, Hess D, Meldal M, Richter WJ, Vliegenthart JF, Hofsteenge J (1996) Spectroscopic and protein chemical analyses demonstrate the presence of C-mannosylated tryptophan in intact human RNase 2 and its isoforms. Biochemistry (Mosc) 35:12005–12014. doi: 10.1021/bi9610515

Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J 432:437–444. doi: 10.1042/BJ20101185

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014a) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–495. doi: 10.1093/nar/gkt1178

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014b) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–495. doi: 10.1093/nar/gkt1178

Lommel M, Strahl S (2009) Protein O-mannosylation: Conserved from bacteria to humans. Glycobiology 19:816–828. doi: 10.1093/glycob/cwp066

Loussert C, Schmitt C, Prevost M-C, Balloy V, Fadel E, Philippe B, Kauffmann-Lacroix C, Latgé JP, Beauvais A (2010) In vivo biofilm composition of Aspergillus fumigatus. Cell Microbiol 12:405–410. doi: 10.1111/j.1462-5822.2009.01409.x

Lozupone CA, Hamady M, Cantarel BL, Coutinho PM, Henrissat B, Gordon JI, Knight R (2008) The convergence of carbohydrate active gene repertoires in human gut microbes. Proc Natl Acad Sci U S A 105:15076–15081. doi: 10.1073/pnas.0807339105

Luonteri E, Tenkanen M, Viikari L (1998) Substrate specificities of Penicillium simplicissimum alpha-galactosidases. Enzyme Microb Technol 22:192–198.

Macdonald SS, Blaukopf M, Withers SG (2014) N-Acetyl glucosaminidases from CAZy family GH3 are really glycoside phosphorylases, thereby explaining their use of histidine as an acid/base catalyst in place of glutamic acid. J Biol Chem. doi: 10.1074/jbc.M114.621110

Maeda M, Kamamoto M, Hino K, Yamamoto S, Kimura M, Okano M, Kimura Y (2005) Glycoform analysis of Japanese cedar pollen allergen, Cry j 1. Biosci Biotechnol Biochem 69:1700–1705. doi: 10.1271/bbb.69.1700

Maeda M, Kimura Y (2014) Structural features of free N-glycans occurring in plants and functional features of de-N-glycosylation enzymes, ENGase, and PNGase: the presence of unusual plant complex type N-glycans. Front Plant Sci. doi: 10.3389/fpls.2014.00429

Mahowald MA, Rey FE, Seedorf H, Turnbaugh PJ, Fulton RS, Wollam A, Shah N, Wang C, Magrini V, Wilson RK, Cantarel BL, Coutinho PM, Henrissat B, Crock LW, Russell A, Verberkmoes NC, Hettich RL, Gordon JI (2009) Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc Natl Acad Sci U S A 106:5859–5864. doi: 10.1073/pnas.0901529106

Malbert Y, Pizzut-Serin S, Massou S, Cambon E, Laguerre S, Monsan P, Lefoulon F, Morel S, André I, Remaud-Simeon M (2014) Extending the Structural Diversity of α-Flavonoid Glycosides with Engineered Glucansucrases. ChemCatChem n/a–n/a. doi: 10.1002/cctc.201402144

Mamedov T, Yusibov V (2011) Green algae Chlamydomonas reinhardtii possess endogenous sialylated N-glycans. FEBS Open Bio 1:15–22. doi: 10.1016/j.fob.2011.10.003

Martens EC, Chiang HC, Gordon JI (2008a) Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. Cell Host Microbe 4:447–457. doi: 10.1016/j.chom.2008.09.007

Martens EC, Chiang HC, Gordon JI (2008b) Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. Cell Host Microbe 4:447–457. doi: 10.1016/j.chom.2008.09.007

Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009a) Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. J Biol Chem 284:24673–24677. doi: 10.1074/jbc.R109.022848

Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009b) Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem 284:24673–24677. doi: 10.1074/jbc.R109.022848

Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, Gordon JI (2011) Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. PLoS Biol 9:e1001221. doi: 10.1371/journal.pbio.1001221

Maruyama Y, Nakajima T (2000) The aman6 gene encoding a yeast mannan backbone degrading 1,6-alpha-D-mannanase in Bacillus circulans: cloning, sequence analysis, and expression. Biosci Biotechnol Biochem 64:2018–2020. doi: 10.1271/bbb.64.2018

Masuoka J (2004) Surface glycans of Candida albicans and other pathogenic fungi: physiological roles, clinical uses, and experimental challenges. Clin Microbiol Rev 17:281–310.

Matsui T, Takita E, Sato T, Kinjo S, Aizawa M, Sugiura Y, Hamabata T, Sawada K, Kato K (2011) N-glycosylation at noncanonical Asn-X-Cys sequences in plant cells. Glycobiology 21:994–999. doi: 10.1093/glycob/cwq198

Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell 122:107–118. doi: 10.1016/j.cell.2005.05.007

McNulty NP, Wu M, Erickson AR, Pan C, Erickson BK, Martens EC, Pudlo NA, Muegge BD, Henrissat B, Hettich RL, Gordon JI (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:e1001637. doi: 10.1371/journal.pbio.1001637

McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B, Oozeer R, Cools-Portier S, Gobert G, Chervaux C, Knights D, Lozupone CA, Knight R, Duncan AE, Bain JR, Muehlbauer MJ, Newgard CB, Heath AC, Gordon JI (2011) The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl Med 3:106ra106. doi: 10.1126/scitranslmed.3002701

Meli VS, Ghosh S, Prabha TN, Chakraborty N, Chakraborty S, Datta A (2010) Enhancement of fruit shelf life by suppressing N-glycan processing enzymes. Proc Natl Acad Sci 107:2413–2418. doi: 10.1073/pnas.0909329107

Mengele R, Sumper M (1992) Drastic differences in glycosylation of related S-layer glycoproteins from moderate and extreme halophiles. J Biol Chem 267:8182–8185.

Mescher MF, Strominger JL (1976) Purification and characterization of a prokaryotic glucoprotein from the cell envelope of Halobacterium salinarium. J Biol Chem 251:2005–2014.

Meyer BH, Albers S-V (2013) Hot and sweet: protein glycosylation in Crenarchaeota. Biochem Soc Trans 41:384–392. doi: 10.1042/BST20120296

Michell SL, Whelan AO, Wheeler PR, Panico M, Easton RL, Etienne AT, Haslam SM, Dell A, Morris HR, Reason AJ, Herrmann JL, Young DB, Hewinson RG (2003) The MPB83 antigen from Mycobacterium bovis contains O-linked mannose and (1-->3)-mannobiose moieties. J Biol Chem 278:16423–16432. doi: 10.1074/jbc.M207959200

Mirza O, Skov LK, Sprogøe D, van den Broek LAM, Beldman G, Kastrup JS, Gajhede M (2006) Structural rearrangements of sucrose phosphorylase from Bifidobacterium adolescentis during sucrose conversion. J Biol Chem 281:35576–35584. doi: 10.1074/jbc.M605611200

Monsan P, Remaud-Siméon M, André I (2010) Transglucosidases as efficient tools for oligosaccharide and glucoconjugate synthesis. Curr Opin Microbiol 13:293–300. doi: 10.1016/j.mib.2010.03.002

Moreira LRS, Filho EXF (2008) An overview of mannan structure and mannan-degrading enzyme systems. Appl Microbiol Biotechnol 79:165–178. doi: 10.1007/s00253-008-1423-4

Mukasa R, Umeda M, Endo T, Kobata A, Inoue K (1995) Characterization of glycosylphosphatidylinositol (GPI)-anchored NCAM on mouse skeletal muscle cell line C2C12: the structure of the GPI glycan and release during myogenesis. Arch Biochem Biophys 318:182–190. doi: 10.1006/abbi.1995.1219

Munro S (2001) What can yeast tell us about N-linked glycosylation in the Golgi apparatus? FEBS Lett 498:223–227.

Munte CE, Gäde G, Domogalla B, Kremer W, Kellner R, Kalbitzer HR (2008) C-mannosylation in the hypertrehalosaemic hormone from the stick insect Carausius morosus. FEBS J 275:1163–1173. doi: 10.1111/j.1742-4658.2008.06277.x

Nagae M, Yamaguchi Y (2012) Function and 3D Structure of the N-Glycans on Glycoproteins. Int J Mol Sci 13:8398–8429. doi: 10.3390/ijms13078398

Nagy E, Maier T, Urban E, Terhes G, Kostrzewa M, ESCMID Study Group on Antimicrobial Resistance in Anaerobic Bacteria (2009) Species identification of clinical isolates of Bacteroides by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis 15:796–802. doi: 10.1111/j.1469-0691.2009.02788.x

Nakae S, Ito S, Higa M, Senoura T, Wasaki J, Hijikata A, Shionyu M, Ito S, Shirai T (2013) Structure of Novel Enzyme in Mannan Biodegradation Process 4-O-β-d-Mannosyl-d-Glucose Phosphorylase MGP. J Mol Biol 425:4468–4478. doi: 10.1016/j.jmb.2013.08.002

Nakai H, Hachem MA, Petersen BO, Westphal Y, Mannerstedt K, Baumann MJ, Dilokpimol A, Schols HA, Duus JØ, Svensson B (2010) Efficient chemoenzymatic oligosaccharide synthesis by reverse phosphorolysis using cellobiose phosphorylase and cellodextrin phosphorylase from Clostridium thermocellum. Biochimie 92:1818–1826. doi: 10.1016/j.biochi.2010.07.013

Nakai H, Kitaoka M, Svensson B, Ohtsubo K (2013) Recent development of phosphorylases possessing large potential for oligosaccharide synthesis. Curr Opin Chem Biol 17:301–309. doi: 10.1016/j.cbpa.2013.01.006

Nakajima M, Toyoizumi H, Abe K, Nakai H, Taguchi H, Kitaoka M (2014) 1,2-β-Oligoglucan phosphorylase from Listeria innocua. PloS One 9:e92353. doi: 10.1371/journal.pone.0092353

Naser SA, Sagramsingh SR, Naser AS, Thanigachalam S (2014) Mycobacterium avium subspecies paratuberculosis causes Crohn's disease in some inflammatory bowel disease patients. World J Gastroenterol WJG 20:7403–7415. doi: 10.3748/wjg.v20.i23.7403

Ng SYM, Wu J, Nair DB, Logan SM, Robotham A, Tessier L, Kelly JF, Uchida K, Aizawa S-I, Jarrell KF (2011) Genetic and mass spectrometry analyses of the unusual type IV-like pili of the archaeon Methanococcus maripaludis. J Bacteriol 193:804–814. doi: 10.1128/JB.00822-10

Nidetzky B, Eis C, Albert M (2000) Role of non-covalent enzyme-substrate interactions in the reaction catalysed by cellobiose phosphorylase from Cellulomonas uda. Biochem J 351 Pt 3:649–659.

Nidetzky B, Griessler R, Schwarz A, Splechtna B (2004) Cellobiose phosphorylase from Cellulomonas uda: gene cloning and expression in Escherichia coli, and application of the recombinant enzyme in a "glycosynthase-type" reaction. J Mol Catal B Enzym 29:241–248. doi: 10.1016/j.molcatb.2003.11.014

Nihira T, Saito Y, Kitaoka M, Nishimoto M, Otsubo K, Nakai H (2012) Characterization of a laminaribiose phosphorylase from Acholeplasma laidlawii PG-8A and production of 1,3-β-d-glucosyl disaccharides. Carbohydr Res 361:49–54. doi: 10.1016/j.carres.2012.08.006

Nihira T, Saito Y, Nishimoto M, Kitaoka M, Igarashi K, Ohtsubo K, Nakai H (2013a) Discovery of cellobionic acid phosphorylase in cellulolytic bacteria and fungi. FEBS Lett 587:3556–3561. doi: 10.1016/j.febslet.2013.09.014

Nihira T, Saito Y, Ohtsubo K, Nakai H, Kitaoka M (2014) 2-O-α-D-glucosylglycerol phosphorylase from Bacillus selenitireducens MLS10 possessing hydrolytic activity on β-D-glucose 1-phosphate. PloS One 9:e86548. doi: 10.1371/journal.pone.0086548

Nihira T, Suzuki E, Kitaoka M, Nishimoto M, Ohtsubo K, Nakai H (2013b) Discovery of β-1,4-D-mannosyl-N-acetyl-D-glucosamine phosphorylase involved in the metabolism of N-glycans. J Biol Chem 288:27366–27374. doi: 10.1074/jbc.M113.469080

Nikulina M, Habich C, Flohé SB, Scott FW, Kolb H (2004) Wheat gluten causes dendritic cell maturation and chemokine secretion. J Immunol Baltim Md 1950 173:1925–1933.

Nishihara M, Utagawa M, Akutsu H, Koga Y (1992) Archaea contain a novel diether phosphoglycolipid with a polar head group identical to the conserved core of eucaryal glycosyl phosphatidylinositol. J Biol Chem 267:12432–12435.

Nishimoto M, Kitaoka M (2007) Practical preparation of lacto-N-biose I, a candidate for the bifidus factor in human milk. Biosci Biotechnol Biochem 71:2101–2104.

Nishimoto M, Kitaoka M (2009) One-pot enzymatic production of beta-D-galactopyranosyl-(1-->3)-2-acetamido-2-deoxy-D-galactose (galacto-N-biose) from sucrose and 2-acetamido-2-deoxy-D-galactose (N-acetylgalactosamine). Carbohydr Res 344:2573–2576. doi: 10.1016/j.carres.2009.09.031

Nothaft H, Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol 8:765–778. doi: 10.1038/nrmicro2383

Nothaft H, Szymanski CM (2013) Bacterial Protein N-Glycosylation: New Perspectives and Applications. J Biol Chem 288:6912–6920. doi: 10.1074/jbc.R112.417857

Ogawa H, Fukushima K, Naito H, Funayama Y, Unno M, Takahashi K, Kitayama T, Matsuno S, Ohtani H, Takasawa S, Okamoto H, Sasaki I (2003) Increased expression of HIP/PAP and regenerating gene III in human inflammatory bowel disease and a murine bacterial reconstitution model. Inflamm Bowel Dis 9:162–170.

Ohdan K, Fujii K, Yanase M, Takaha T, Kuriki T (2007) Phosphorylase coupling as a tool to convert cellobiose into amylose. J Biotechnol 127:496–502. doi: 10.1016/j.jbiotec.2006.07.023

Ohno H, Hase K (2010) Glycoprotein 2 (GP2): grabbing the FimH $^{+}$ bacteria into M cells for mucosal immunity. Gut Microbes 1:407–410. doi: 10.4161/gmic.1.6.14078

Ojima T, Saburi W, Sato H, Yamamoto T, Mori H, Matsui H (2011) Biochemical characterization of a thermophilic cellobiose 2-epimerase from a thermohalophilic bacterium, Rhodothermus marinus JCM9785. Biosci Biotechnol Biochem 75:2162–2168. doi: 10.1271/bbb.110456

O'Keefe SJD (2008) Nutrition and colonic health: the critical role of the microbiota. Curr Opin Gastroenterol 24:51–58. doi: 10.1097/MOG.0b013e3282f323f3

O'Neill EC, Field RA (2014) Enzymatic synthesis using glycoside phosphorylases. Carbohydr Res. doi: 10.1016/j.carres.2014.06.010

O'Neill EC, Field RA (2015) Enzymatic synthesis using glycoside phosphorylases. Carbohydr Res 403:23–37. doi: 10.1016/j.carres.2014.06.010

Oppenheimer M, Valenciano AL, Sobrado P (2011) Biosynthesis of galactofuranose in kinetoplastids: novel therapeutic targets for treating leishmaniasis and chagas' disease. Enzyme Res 2011:415976. doi: 10.4061/2011/415976

Ortalo-Magné A, Dupont MA, Lemassu A, Andersen AB, Gounon P, Daffé M (1995) Molecular composition of the outermost capsular material of the tubercle bacillus. Microbiol Read Engl 141 (Pt 7):1609–1620.

Ouwerkerk JP, de Vos WM, Belzer C (2013a) Glycobiome: bacteria and mucus at the epithelial interface. Best Pract Res Clin Gastroenterol 27:25–38. doi: 10.1016/j.bpg.2013.03.001

Ouwerkerk JP, de Vos WM, Belzer C (2013b) Glycobiome: Bacteria and mucus at the epithelial interface. Best Pract Res Clin Gastroenterol 27:25–38. doi: 10.1016/j.bpg.2013.03.001

Ouyang H, Chen X, Lü Y, Wilson IBH, Tang G, Wang A, Jin C (2013) One single basic amino acid at the ω-1 or ω-2 site is a signal that retains glycosylphosphatidylinositol-anchored protein in the plasma membrane of Aspergillus fumigatus. Eukaryot Cell 12:889–899. doi: 10.1128/EC.00351-12

Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC (2010) CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology 20:1574–1584. doi: 10.1093/glycob/cwq106

Pattison RJ, Amtmann A (2009) N-glycan production in the endoplasmic reticulum of plants. Trends Plant Sci 14:92–99. doi: 10.1016/j.tplants.2008.11.008

Paulick MG, Bertozzi CR (2008) The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteins †. Biochemistry (Mosc) 47:6991–7000. doi: 10.1021/bi8006324

Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G (2013) Hemicellulose biosynthesis. Planta 238:627–642. doi: 10.1007/s00425-013-1921-1

Peng W, Pranskevich J, Nycholat C, Gilbert M, Wakarchuk W, Paulson JC, Razi N (2012) Helicobacter pylori β1,3-N-acetylglucosaminyltransferase for versatile synthesis of type 1 and type 2 poly-LacNAcs on N-linked, O-linked and I-antigen glycans. Glycobiology 22:1453–1464. doi: 10.1093/glycob/cws101

Perez-Vilar J, Randell SH, Boucher RC (2004) C-Mannosylation of MUC5AC and MUC5B Cys subdomains. Glycobiology 14:325–337. doi: 10.1093/glycob/cwh041

Peyfoon E, Meyer B, Hitchen PG, Panico M, Morris HR, Haslam SM, Albers S-V, Dell A (2010) The S-Layer Glycoprotein of the Crenarchaeote *Sulfolobus acidocaldarius* Is Glycosylated at Multiple Sites with Chitobiose-Linked *N* -Glycans. Archaea 2010:1–10. doi: 10.1155/2010/754101

Praissman JL, Wells L (2014) Mammalian O-Mannosylation Pathway: Glycan Structures, Enzymes, and Protein Substrates. Biochemistry (Mosc) 53:3066–3078. doi: 10.1021/bi500153y

Prakash S, Tomaro-Duchesneau C, Saha S, Cantor A (2011) The Gut Microbiota and Human Health with an Emphasis on the Use of Microencapsulated Bacterial Cells. J Biomed Biotechnol 2011:1–12. doi: 10.1155/2011/981214

Principato M, Qian B-F (2014) Staphylococcal enterotoxins in the etiopathogenesis of mucosal autoimmunity within the gastrointestinal tract. Toxins 6:1471–1489. doi: 10.3390/toxins6051471

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J-M, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Antolin M, Artiguenave F, Blottiere H, Borruel N, Bruls T, Casellas F, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Forte M, Friss C, van de Guchte M, Guedon E, Haimet F, Jamet A, Juste C, Kaci G, Kleerebezem M, Knol J, Kristensen M, Layec S, Le Roux K, Leclerc M, Maguin E, Melo Minardi R, Oozeer R, Rescigno M, Sanchez N, Tims S, Torrejon T, Varela E, de Vos W, Winogradsky Y, Zoetendal E, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–65. doi: 10.1038/nature08821

Quigley EMM (2010) Prebiotics and probiotics; modifying and mining the microbiota. Pharmacol Res Off J Ital Pharmacol Soc 61:213–218. doi: 10.1016/j.phrs.2010.01.004

Ralton JE, Naderer T, Piraino HL, Bashtannyk TA, Callaghan JM, McConville MJ (2003) Evidence that intracellular beta1-2 mannan is a virulence factor in Leishmania parasites. J Biol Chem 278:40757–40763. doi: 10.1074/jbc.M307660200

Rasmussen MA, Hespell RB, White BA, Bothast RJ (1988) Inhibitory Effects of Methylcellulose on Cellulose Degradation by Ruminococcus flavefaciens. Appl Environ Microbiol 54:890–897.

Reeves AR, Wang GR, Salyers AA (1997) Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. J Bacteriol 179:643–649.

Renzi F, Manfredi P, Mally M, Moes S, Jenö P, Cornelis GR (2011) The N-glycan Glycoprotein Deglycosylation Complex (Gpd) from Capnocytophaga canimorsus Deglycosylates Human IgG. PLoS Pathog 7:e1002118. doi: 10.1371/journal.ppat.1002118

Resta SC (2009) Effects of probiotics and commensals on intestinal epithelial physiology: implications for nutrient handling. J Physiol 587:4169–4174. doi: 10.1113/jphysiol.2009.176370

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466:334–338. doi: 10.1038/nature09199

Richard ML, Plaine A (2007) Comprehensive analysis of glycosylphosphatidylinositol-anchored proteins in Candida albicans. Eukaryot Cell 6:119–133. doi: 10.1128/EC.00297-06

Rodrigues DF, Elimelech M (2009) Role of type 1 fimbriae and mannose in the development of *Escherichia coli* K12 biofilm: from initial cell adhesion to biofilm formation. Biofouling 25:401–411. doi: 10.1080/08927010902833443

Ronchera-Oms CL, Jiménez NV, Peidro J (1995) Stability of parenteral nutrition admixtures containing organic phosphates. Clin Nutr Edinb Scotl 14:373–380.

Rossi M, Amaretti A, Raimondi S (2011) Folate production by probiotic bacteria. Nutrients 3:118–134. doi: 10.3390/nu3010118

Salgado H, Moreno-Hagelsieb G, Smith TF, Collado-Vides J (2000) Operons in Escherichia coli: genomic analyses and predictions. Proc Natl Acad Sci U S A 97:6652–6657. doi: 10.1073/pnas.110147297

Sartor RB (2008) Microbial influences in inflammatory bowel diseases. Gastroenterology 134:577–594. doi: 10.1053/j.gastro.2007.11.059

Sartor RB (2006) Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. Nat Clin Pract Gastroenterol Hepatol 3:390–407. doi: 10.1038/ncpgasthep0528

Scaldaferri F, Gerardi V, Lopetuso LR, Del Zompo F, Mangiola F, Boškoski I, Bruno G, Petito V, Laterza L, Cammarota G, Gaetani E, Sgambato A, Gasbarrini A (2013) Gut microbial flora, prebiotics, and probiotics in IBD: their current usage and utility. BioMed Res Int 2013:435268. doi: 10.1155/2013/435268

Scanlan PD, Shanahan F, Clune Y, Collins JK, O'Sullivan GC, O'Riordan M, Holmes E, Wang Y, Marchesi JR (2008) Culture-independent analysis of the gut microbiota in colorectal

cancer and polyposis. Environ Microbiol 10:789–798. doi: 10.1111/j.1462-2920.2007.01503.x

Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol 61:263–289. doi: 10.1146/annurev-arplant-042809-112315

Schirbel A, Kessler S, Rieder F, West G, Rebert N, Asosingh K, McDonald C, Fiocchi C (2013) Pro-Angiogenic Activity of TLRs and NLRs: A Novel Link Between Gut Microbiota and Intestinal Angiogenesis. Gastroenterology 144:613–623.e9. doi: 10.1053/j.gastro.2012.11.005

Schmalhorst PS, Krappmann S, Vervecken W, Rohde M, Muller M, Braus GH, Contreras R, Braun A, Bakker H, Routier FH (2008) Contribution of Galactofuranose to the Virulence of the Opportunistic Pathogen Aspergillus fumigatus. Eukaryot Cell 7:1268–1277. doi: 10.1128/EC.00109-08

Schmeing TM, Huang KS, Kitchen DE, Strobel SA, Steitz TA (2005) Structural Insights into the Roles of Water and the 2' Hydroxyl of the P Site tRNA in the Peptidyl Transferase Reaction. Mol Cell 20:437–448. doi: 10.1016/j.molcel.2005.09.006

Schultz C, Gilson P, Oxley D, Youl J, Bacic A (1998) GPI-anchors on arabinogalactan-proteins: implications for signalling in plants. Trends Plant Sci 3:426–431. doi: 10.1016/S1360-1385(98)01328-4

Schwiertz A, Taras D, Schäfer K, Beijer S, Bos NA, Donus C, Hardt PD (2010) Microbiota and SCFA in lean and overweight healthy subjects. Obes Silver Spring Md 18:190–195. doi: 10.1038/oby.2009.167

Scott KP, Gratz SW, Sheridan PO, Flint HJ, Duncan SH (2013) The influence of diet on the gut microbiota. Pharmacol Res 69:52–60. doi: 10.1016/j.phrs.2012.10.020

Sekirov I, Russell SL, Antunes LCM, Finlay BB (2010) Gut Microbiota in Health and Disease. Physiol Rev 90:859–904. doi: 10.1152/physrev.00045.2009

Senoura T, Ito S, Taguchi H, Higa M, Hamada S, Matsui H, Ozawa T, Jin S, Watanabe J, Wasaki J, Ito S (2011) New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase. Biochem Biophys Res Commun 408:701–706. doi: 10.1016/j.bbrc.2011.04.095

Sharma V, Ichikawa M, Freeze HH (2014) Mannose metabolism: More than meets the eye. Biochem Biophys Res Commun 453:220–228. doi: 10.1016/j.bbrc.2014.06.021

Shibata N, Kobayashi H, Okawa Y, Suzuki S (2003) Existence of novel beta-1,2 linkage-containing side chain in the mannan of Candida lusitaniae, antigenically related to Candida albicans serotype A. Eur J Biochem FEBS 270:2565–2575.

Shi L, Berg S, Lee A, Spencer JS, Zhang J, Vissa V, McNeil MR, Khoo K-H, Chatterjee D (2006) The carboxy terminus of EmbC from Mycobacterium smegmatis mediates chain

length extension of the arabinan in lipoarabinomannan. J Biol Chem 281:19512–19526. doi: 10.1074/jbc.M513846200

Shipman JA, Berleman JE, Salyers AA (2000a) Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. J Bacteriol 182:5365–5372.

Shipman JA, Berleman JE, Salyers AA (2000b) Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. J Bacteriol 182:5365–5372.

Skropeta D (2009) The effect of individual N-glycans on enzyme activity. Bioorg Med Chem 17:2645–2653. doi: 10.1016/j.bmc.2009.02.037

Sommer MOA, Church GM, Dantas G (2010) The human microbiome harbors a diverse reservoir of antibiotic resistance genes. Virulence 1:299–303. doi: 10.4161/viru.1.4.12010

Song W, Henquet MGL, Mentink RA, van Dijk AJ, Cordewener JHG, Bosch D, America AHP, van der Krol AR (2011) N-glycoproteomics in plants: Perspectives and challenges. J Proteomics 74:1463–1474. doi: 10.1016/j.jprot.2011.05.007

Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL (2010a) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell 141:1241–1252. doi: 10.1016/j.cell.2010.05.005

Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL (2010b) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell 141:1241–1252. doi: 10.1016/j.cell.2010.05.005

Sonnenburg JL, Xu J, Leip DD, Chen C-H, Westover BP, Weatherford J, Buhler JD, Gordon JI (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science 307:1955–1959. doi: 10.1126/science.1109051

Sorini C, Falcone M (2013) Shaping the (auto)immune response in the gut: the role of intestinal immune regulation in the prevention of type 1 diabetes. Am J Clin Exp Immunol 2:156–171.

Speciale G, Thompson AJ, Davies GJ, Williams SJ (2014) Dissecting conformational contributions to glycosidase catalysis and inhibition. Curr Opin Struct Biol 28C:1–13. doi: 10.1016/j.sbi.2014.06.003

Stappenbeck TS, Hooper LV, Gordon JI (2002) Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. Proc Natl Acad Sci U S A 99:15451–15455. doi: 10.1073/pnas.202604299

Stilling RM, Dinan TG, Cryan JF (2014) Microbial genes, brain & behaviour - epigenetic regulation of the gut-brain axis. Genes Brain Behav 13:69–86. doi: 10.1111/gbb.12109

Stoll D, Stålbrand H, Warren RA (1999) Mannan-degrading enzymes from Cellulomonas fimi. Appl Environ Microbiol 65:2598–2605.

Strasser R (2014) Biological significance of complex N-glycans in plants and their impact on plant physiology. Front Plant Sci. doi: 10.3389/fpls.2014.00363

Strober W (2011) Adherent-invasive E. coli in Crohn disease: bacterial “agent provocateur.” J Clin Invest 121:841–844. doi: 10.1172/JCI46333

Strugala V, Dettmar PW, Pearson JP (2008) Thickness and continuity of the adherent colonic mucus barrier in active and quiescent ulcerative colitis and Crohn’s disease. Int J Clin Pract 62:762–769. doi: 10.1111/j.1742-1241.2007.01665.x

Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234. doi: 10.1016/j.pep.2005.01.016

Sudo N, Chida Y, Aiba Y, Sonoda J, Oyama N, Yu X-N, Kubo C, Koga Y (2004) Postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice. J Physiol 558:263–275. doi: 10.1113/jphysiol.2004.063388

Sumper M, Berg E, Mengele R, Strobel I (1990) Primary structure and glycosylation of the S-layer protein of Haloferax volcanii. J Bacteriol 172:7111–7118.

Suzuki M, Kaneda K, Nakai Y, Kitaoka M, Taniguchi H (2009) Synthesis of cellobiose from starch by the successive actions of two phosphorylases. New Biotechnol 26:137–142. doi: 10.1016/j.nbt.2009.07.004

Takamatsu D, Bensing BA, Sullam PM (2004) Genes in the accessory sec locus of Streptococcus gordonii have three functionally distinct effects on the expression of the platelet-binding protein GspB. Mol Microbiol 52:189–203. doi: 10.1111/j.1365-2958.2004.03978.x

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 (2010a) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res 20:1605–1612. doi: 10.1101/gr.108332.110

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 (2010b) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res 20:1605–1612. doi: 10.1101/gr.108332.110

Termén S, Tollin M, Rodriguez E, Sveinsdóttir SH, Jóhannesson B, Cederlund A, Sjövall J, Agerberth B, Gudmundsson GH (2008) PU.1 and bacterial metabolites regulate the

human gene CAMP encoding antimicrobial peptide LL-37 in colon epithelial cells. Mol Immunol 45:3947–3955. doi: 10.1016/j.molimm.2008.06.020

Terrapon N, Henrissat B (2014a) How do gut microbes break down dietary fiber? Trends Biochem Sci 39:156–158. doi: 10.1016/j.tibs.2014.02.005

Terrapon N, Henrissat B (2014b) How do gut microbes break down dietary fiber? Trends Biochem Sci 39:156–158. doi: 10.1016/j.tibs.2014.02.005

Thompson AJ, Williams RJ, Hakki Z, Alonzi DS, Wennekes T, Gloster TM, Songsrirote K, Thomas-Oates JE, Wrodnigg TM, Spreitz J, Stütz AE, Butters TD, Williams SJ, Davies GJ (2012) Structural and mechanistic insight into N-glycan processing by endo-α-mannosidase. Proc Natl Acad Sci U S A 109:781–786. doi: 10.1073/pnas.1111482109

Trinel P-A, Delplace F, Maes E, Zanetta J-P, Mille C, Coddeville B, Jouault T, Strecker G, Poulain D (2005) Candida albicans serotype B strains synthesize a serotype-specific phospholipomannan overexpressing a beta-1,2-linked mannotriose. Mol Microbiol 58:984–998. doi: 10.1111/j.1365-2958.2005.04890.x

Trinel P-A, Maes E, Zanetta J-P, Delplace F, Coddeville B, Jouault T, Strecker G, Poulain D (2002) Candida albicans phospholipomannan, a new member of the fungal mannose inositol phosphoceramide family. J Biol Chem 277:37260–37271. doi: 10.1074/jbc.M202295200

Turnbaugh PJ, Gordon JI (2009) The core gut microbiome, energy balance and obesity. J Physiol 587:4153–4158. doi: 10.1113/jphysiol.2009.174136

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI (2009) A core gut microbiome in obese and lean twins. Nature 457:480–484. doi: 10.1038/nature07540

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444:1027–1031. doi: 10.1038/nature05414

Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørlie M, Eijsink VGH (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. Science 330:219–222. doi: 10.1126/science.1192231

Van Der Vaart JM, te Biesebeke R, Chapman JW, Klis FM, Verrips CT (1996) The beta-1, 6-glucan containing side-chain of cell wall proteins of Saccharomyces cerevisiae is bound to the glycan core of the GPI moiety. FEMS Microbiol Lett 145:401–407.

Veneault-Fourrey C, Commun C, Kohler A, Morin E, Balestrini R, Plett J, Danchin E, Coutinho P, Wiebenga A, de Vries RP, Henrissat B, Martin F (2014) Genomic and transcriptomic analysis of Laccaria bicolor CAZome reveals insights into polysaccharides remodelling

during symbiosis establishment. Fungal Genet Biol 72:168–181. doi: 10.1016/j.fgb.2014.08.007

Voisin S, Houliston RS, Kelly J, Brisson J-R, Watson D, Bardy SL, Jarrell KF, Logan SM (2005) Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of Methanococcus voltae. J Biol Chem 280:16586–16593. doi: 10.1074/jbc.M500329200

Walker AW, Sanderson JD, Churcher C, Parkes GC, Hudspith BN, Rayment N, Brostoff J, Parkhill J, Dougan G, Petrovska L (2011) High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. BMC Microbiol 11:7. doi: 10.1186/1471-2180-11-7

Wang Y, Murakami Y, Yasui T, Wakana S, Kikutani H, Kinoshita T, Maeda Y (2013) Significance of Glycosylphosphatidylinositol-anchored Protein Enrichment in Lipid Rafts for the Control of Autoimmunity. J Biol Chem 288:25490–25499. doi: 10.1074/jbc.M113.492611

Wehmeier S, Varghese AS, Gurcha SS, Tissot B, Panico M, Hitchen P, Morris HR, Besra GS, Dell A, Smith MCM (2009) Glycosylation of the phosphate binding protein, PstS, in Streptomyces coelicolor by a pathway that resembles protein O-mannosylation in eukaryotes. Mol Microbiol 71:421–433. doi: 10.1111/j.1365-2958.2008.06536.x

Wellens A, Garofalo C, Nguyen H, Van Gerven N, Slättegård R, Hernalsteens J-P, Wyns L, Oscarson S, De Greve H, Hultgren S, Bouckaert J (2008) Intervening with Urinary Tract Infections Using Anti-Adhesives Based on the Crystal Structure of the FimH–Oligomannose-3 Complex. PLoS ONE 3:e2040. doi: 10.1371/journal.pone.0002040

Westover BP, Buhler JD, Sonnenburg JL, Gordon JI (2005) Operon prediction without a training set. Bioinforma Oxf Engl 21:880–888. doi: 10.1093/bioinformatics/bti123

Wexler HM (2007) Bacteroides: the Good, the Bad, and the Nitty-Gritty. Clin Microbiol Rev 20:593–621. doi: 10.1128/CMR.00008-07

Wilson IB, Zeleny R, Kolarich D, Staudacher E, Stroop CJ, Kamerling JP, Altmann F (2001) Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core alpha1,3-linked fucose and xylose substitutions. Glycobiology 11:261–274.

Wink DA, Hines HB, Cheng RYS, Switzer CH, Flores-Santana W, Vitek MP, Ridnour LA, Colton CA (2011) Nitric oxide and redox mechanisms in the immune response. J Leukoc Biol 89:873–891. doi: 10.1189/jlb.1010550

World Health Organization (2003) Joint WHO/FAO Expert consutation. “Diet, Nutrition and the prevention of Chronic Diseases.”

Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI (2003) A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science 299:2074–2076. doi: 10.1126/science.1080029

Xu J, Chiang HC, Bjursell MK, Gordon JI (2004) Message from a human gut symbiont: sensitivity is a prerequisite for sharing. Trends Microbiol 12:21–28.

Xu J, Gordon JI (2003) Honor thy symbionts. Proc Natl Acad Sci U S A 100:10452–10459. doi: 10.1073/pnas.1734063100

Yamamoto Y, Kawashima D, Hashizume A, Hisamatsu M, Isono N (2013) Purification and characterization of 1,3-β-D-glucan phosphorylase from Ochromonas danica. Biosci Biotechnol Biochem 77:1949–1954. doi: 10.1271/bbb.130411

Yang LL, Haug A (1979) Purification and partial characterization of a procaryotic glycoprotein from the plasma membrane of Thermoplasma acidophilum. Biochim Biophys Acta 556:265–277.

Yang X, Xie L, Li Y, Wei C (2009) More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. PloS One 4:e6074. doi: 10.1371/journal.pone.0006074

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–W451. doi: 10.1093/nar/gks479

Yoshiie T, Maeda M, Kimura M, Hama Y, Uchida M, Kimura Y (2012) Structural Features of N-Glycans of Seaweed Glycoproteins: Predominant Occurrence of High-Mannose Type N-Glycans in Marine Plants. Biosci Biotechnol Biochem 76:1996–1998. doi: 10.1271/bbb.120463

Zaragoza O, Rodrigues ML, De Jesus M, Frases S, Dadachova E, Casadevall A (2009) The capsule of the fungal pathogen Cryptococcus neoformans. Adv Appl Microbiol 68:133–216. doi: 10.1016/S0065-2164(09)01204-0

Zeitler R, Hochmuth E, Deutzmann R, Sumper M (1998) Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent N-glycosylation of the cell surface glycoprotein from Halobacterium halobium. Glycobiology 8:1157–1164.

Zeleny R, Kolarich D, Strasser R, Altmann F (2006) Sialic acid concentrations in plants are in the range of inadvertent contamination. Planta 224:222–227. doi: 10.1007/s00425-005-0206-8

Zhang X, Rogowski A, Zhao L, Hahn MG, Avci U, Knox JP, Gilbert HJ (2014) Understanding How the Complex Molecular Architecture of Mannan-degrading Hydrolases Contributes to Plant Cell Wall Degradation. J Biol Chem 289:2002–2012. doi: 10.1074/jbc.M113.527770

Zhang Y-HP, Lynd LR (2005) Cellulose utilization by Clostridium thermocellum: bioenergetics and hydrolysis product assimilation. Proc Natl Acad Sci U S A 102:7321–7325. doi: 10.1073/pnas.0408734102

Zhu Y, Suits MDL, Thompson AJ, Chavan S, Dinev Z, Dumon C, Smith N, Moremen KW, Xiang Y, Siriwardena A, Williams SJ, Gilbert HJ, Davies GJ (2010) Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont. Nat Chem Biol 6:125–132. doi: 10.1038/nchembio.278

Zocco MA, Ainora ME, Gasbarrini G, Gasbarrini A (2007) Bacteroides thetaiotaomicron in the gut: molecular aspects of their interaction. Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver 39:707–712. doi: 10.1016/j.dld.2007.04.003

Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Booijink CCGM, Troost FJ, Bork P, Wels M, de Vos WM, Kleerebezem M (2012) The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J 6:1415–1426. doi: 10.1038/ismej.2011.212

Thesis goals

As introduced in the previous chapters, CAZymes play a key role in the metabolization of the numerous plant, host and microbial glycans by human gut bacteria. In particular, mannoside recognition and degradation are central elements of interactions between gut microbes, their host, and dietary components. However, despite the recent fascinating studies on mannoside metabolization by gut bacteria, in particular those belonging to the *Bacteroides* genus, much remains to be done in order to identify and characterize some of the main actors of these catabolic pathways. In particular, the role of glycoside phosphorylases in glycoside metabolization by anaerobic bacteria is probably underestimated, because of the lack of studies targeting these original enzymes, which are both able to breakdown and to create glycosidic linkages.

In this context, the main goal of my thesis was to identify the function of enzymes classified in the glycoside hydrolase GH130 family, that are produced by highly prevalent cultivated and uncultivated *Bacteroides* species. When this project started, in October 2011, the picture of the GH130 family was the following: only one GH130 member was functionally characterized (*Bf*MGP, a monospecific enzyme acting on β-D-Mannopyranosyl-1,4-D-glucose, a disaccharide released during the last steps of glucomannan breakdown), and four GH130 3D structures were available, with no associated function.

At that stage, a lot of questions arose, to which we were committed to answer during this thesis:

- Are GH130 enzymes dedicated to β-D-Man*p*-1,4-D-Glc*p* phosphorolysis, or do they process various other substrates?
- In particular, are they all specific for mannosides?
- What is the catalytic amino acid of these enzymes, and which mechanism do they use?
- What are the structural determinants of their substrate specificity?
- What are their physiological roles in the human gut ecosystem?
- Would it be possible to control their activity if it is deleterious for human health?
- Does the GH130 family also contain members acting as hydrolases, as it is the case for several GPs containing CAZy families?
- If so, which are the structural differences harbored by GH130 hydrolases and phosphorylases?
- Are the activities displayed by the GH130 members relevant for biotechnological applications?

The work carried out for three years to answer these questions is presented in the four following chapters.

Results

First article

Functional metagenomics is a powerful tool to mine ecosystems by directly focusing on the catalytic activities of interest. It is a valuable strategy to save money, optimize and rationalize research efforts. In our team, the functional metagenomics of the human gut microbiota (Tasse *et al.*, 2010), focusing primarily on dietary fiber deconstruction, allowed one to identify dozens of original enzymes of unknown function, belonging to novel CAZy families. One of them in particular, called latter on Uhgb_MP, caught our attention. First, this enzyme of unknown function appeared to be very prevalent in the human gut. Characterizing it regarding its activity, with the aim of understanding its *in vivo* role, was the first objective of this PhD project. In 2011, Uhgb_MP sequence was classified in the GH130 CAZy family, which contained at the beginning of my PhD, only one functionally characterized member. This enzyme is *Bf*MGP, a mannoside phosphorylase from *B. fragilis* which was in 2011, the unique known phosphorylase producing α-D-mannose-1-phosphate.

In order to understand which role is played by Uhgb_MP and its homologs in carbohydrate foraging, we used primary structure analysis, kinetics, NMR, site-directed mutagenesis, 3D modelling and docking in combination with PUL and GH130 prevalence analyses at the scale of the entire human gut ecosystem. The results are presented in the following paper, published in Journal of Biological Chemistry in 2013. They were partly obtained in collaboration with B. Henrissat and V. Lombard, who hosted me at the AFMB (CNRS Marseilles) during several weeks for creation of the GH130 subfamilies and identification of the GH130 containing PULs.

Moreover, this work highlighted the great substrate promiscuity harbored by Uhgb_MP, which makes it an interesting tool for glycodiversification for industrial purposes. It concerns not only the synthesis of the expensive β-D-mannopyranosyl-1,4-*N*-acetyl-D-glucosamine or β-D-mannopyranosyl-1,4-chitobiose, but also various oligosaccharides containing β-1,4-D-mannopyranosyl units at the non-reductive end, using mannan, a low cost agroresource, as feedstock. These applications have been patented in 2013 (European Patent N° EP13306108.5, international extension N° PCT/EP2014/066565) (Appendix).

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 45, pp. 32370–32383, November 8, 2013

Role of Glycoside Phosphorylases in Mannose Foraging by Human Gut Bacteria*[S]

Received for publication, May 7, 2013, and in revised form, September 11, 2013 Published, JBC Papers in Press, September 16, 2013, DOI 10.1074/jbc.M113.483628

Simon Ladevèze[‡§¶], Laurence Tarquis[‡§¶], Davide A. Cecchini[‡§¶], Juliette Bercovici[‡§¶], Isabelle André[‡§¶], Christopher M. Topham[‡§¶], Sandrine Morel[‡§¶], Elisabeth Laville[‡§¶], Pierre Monsan[‡§¶], Vincent Lombard[‖], Bernard Henrissat[‖], and Gabrielle Potocki-Véronèse[‡§¶1]

From the [‡]Institut National des Sciences Appliquées, Université Paul Sabatier, Institut National Polytechnique, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, Université de Toulouse, 135 Avenue de Rangueil, F-31077 Toulouse, the [§]CNRS, UMR5504, F-31400 Toulouse, the [¶]Institut National de Recherche Agronomique, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, and the [‖]Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, CNRS UMR 7257, 163 Avenue de Luminy, F-13288 Marseille, France

Background: The relations between the gut microbiota, food, and host play a crucial role in human health.
Results: Prevalent bacterial glycoside phosphorylases are able to break down dietary carbohydrates and the *N*-glycans lining the intestinal epithelium.
Conclusion: GH130 enzymes are new targets to study interactions between host and gut microbes.
Significance: Glycoside phosphorylases are key enzymes of host glycan catabolism by gut bacteria.

To metabolize both dietary fiber constituent carbohydrates and host glycans lining the intestinal epithelium, gut bacteria produce a wide range of carbohydrate-active enzymes, of which glycoside hydrolases are the main components. In this study, we describe the ability of phosphorylases to participate in the breakdown of human *N*-glycans, from an analysis of the substrate specificity of UhgbMP, a mannoside phosphorylase of the GH130 protein family discovered by functional metagenomics. UhgbMP is found to phosphorolyze β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc and is also a highly efficient enzyme to catalyze the synthesis of this precious *N*-glycan core oligosaccharide by reverse phosphorolysis. Analysis of sequence conservation within family GH130, mapped on a three-dimensional model of UhgbMP and supported by site-directed mutagenesis results, revealed two GH130 subfamilies and allowed the identification of key residues responsible for catalysis and substrate specificity. The analysis of the genomic context of 65 known GH130 sequences belonging to human gut bacteria indicates that the enzymes of the GH130_1 subfamily would be involved in mannan catabolism, whereas the enzymes belonging to the GH130_2 subfamily would rather work in synergy with glycoside hydrolases of the GH92 and GH18 families in the breakdown of *N*-glycans. The use of GH130 inhibitors as therapeutic agents or functional foods could thus be considered as an innovative strategy to inhibit *N*-glycan degradation, with the ultimate goal of protecting, or restoring, the epithelial barrier.

The human gut microbiota is a dense and complex ecosystem, which plays a crucial role in maintaining human health. The many meta-omic studies carried out in the last few years have shown that certain metabolic diseases, such as obesity, and some inflammatory diseases, such as inflammatory bowel diseases (IBD)[2] (1–4), are associated with structural and functional imbalances of the microbiota (5). The catabolism of complex carbohydrates by gut bacteria plays a key role in the microbial colonization and equilibrium of the digestive tract (6, 7). Indeed, the plant polysaccharides (cellulose, hemicelluloses, pectin, and resistant starch) that compose dietary fiber supply the main source of carbon for gut bacteria, in the form of monosaccharides produced by polysaccharide breakdown, such as glucose, xylose, arabinose, uronic acids, and to a lesser extent, galactose, mannose, and rhamnose (8). In particular, mannose is found in plant cell walls, in mannan and glucomannan, the backbones that consist of a homopolymer of β1,4-linked D-mannopyranosyl residues and a heterogeneous sequence of β1,4-linked D-glucopyranosyl and D-mannopyranosyl units (9). There is, however, a further source of complex carbohydrates in the gut, namely the heavily *O*- and *N*-glycosylated glycoproteins that line the intestinal epithelium to form a protective barrier against pathogens and chemical and mechanical aggression (10). Although host *O*-glycans are rich in *N*-acetylhexosamines (GlcNAc, GalNAc, and Neu5Ac), galactose and fucose, many mature *N*-glycans contain eight α- and β-linked D-mannopyranosyl residues linked to chitobiose (Fig. 1) (11). Human glycan-microbial interactions in the gastrointestinal tract play a crucial role in determining the outcome of relations of both commensals and pathogens with the host. Indeed, alterations in

* This work was supported by the French Ministry of Higher Education and Research and by the French National Institute for Agricultural Research (INRA, CEPIA division and the "Meta-omics of Microbial Ecosystems" research program).
[S] This article contains supplemental Figs. S1–S6 and Tables 1 and 2.
[1] To whom correspondence should be addressed: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, Institut National des Sciences Appliquées, CNRS UMR5504, Institut National de Recherche Agronomique UMR792, 135 Ave. de Rangueil, 31077 Toulouse, France. Tel.: 33-5-61-55-94-87; Fax: 33-5-61-55-94-00; E-mail: veronese@insa-toulouse.fr.

[2] The abbreviations used are: IBD, inflammatory bowel disease; GT, glycosyltransferase; GH, glycoside hydrolase; PL, polysaccharide lyase; GP, glycoside phosphorylase; DP, degree of polymerization; CE, carbohydrate esterase; BfMP, *B. fragilis* NCTC 9343 mannosylglucose phosphorylase; HPAEC-PAD, high performance anion exchange chromatography with pulsed amperometric detection; PDB, Protein Data Bank; *p*NP, *p*-nitrophenyl.

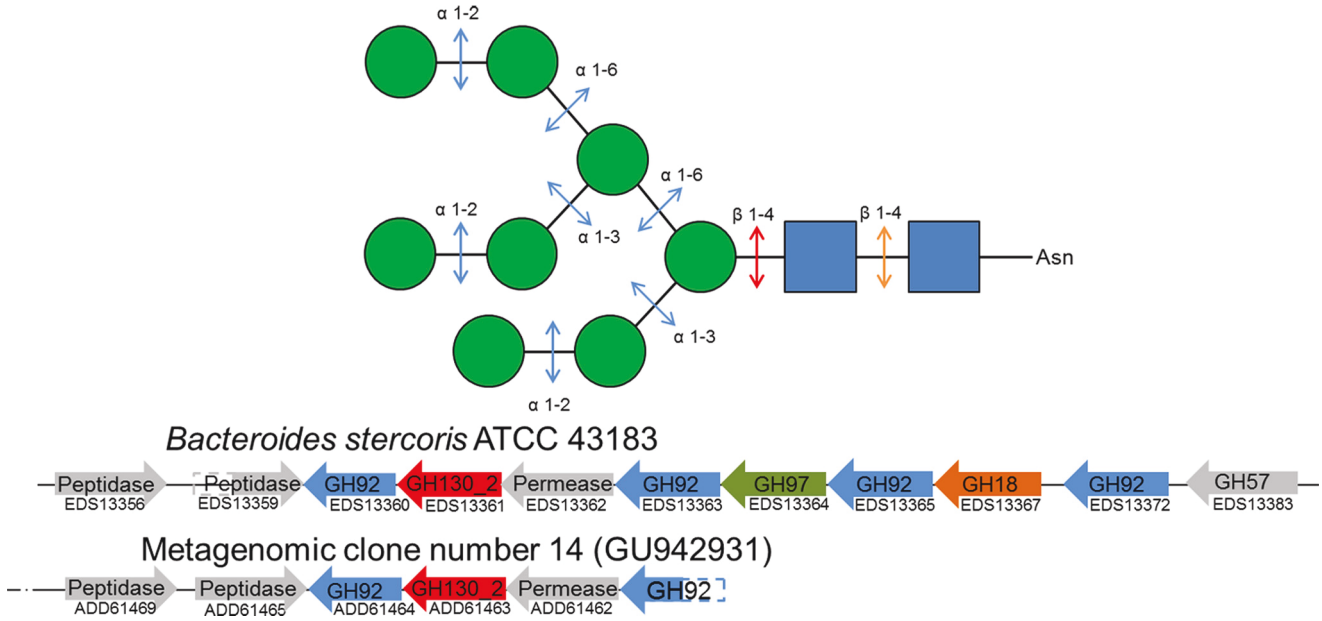


FIGURE 1. **Schematic representation of *N*-glycan processing by glycoside hydrolases and phosphorylases.** *Green spheres,* mannosyl residues; *blue squares, N*-acetyl-D-glucosamine residues. The genomic cluster containing the GH130_2 encoding gene and other ones involved in *N*-glycan processing are shown for *B. stercoris* ATCC 43183 genome, and the UhgbMP encoding metagenomic clone (accession number GU94293). CAZyme encoding genes are colored according to the linkage they break down in the *N*-glycan structure, based on known biochemical data of the families to which they belong.

the structure and/or quantity of host glycans due to biosynthetic defects or microbial degradation alter their barrier function and are thought to be involved in the initiation and the maintenance of mucosal inflammation in IBDs and in the development of intestinal cancer (12).

To breakdown these complex carbohydrates of either plant or human origin, gut bacteria produce a full repertoire of carbohydrate-active enzymes (CAZymes, listed in the CAZy database (13)), of which glycoside hydrolases (GHs) are the main constituents, as revealed by Gill *et al.* (14). In particular, the degradation of dietary mannans requires both endo-β-mannanases and β-mannosidases to release mannose (15). The hydrolysis of *N*-glycans by a broad consortium of endo- and exo-, α-, and β-glycosidases, in particular those produced by the prominent gut bacterium *Bacteroides thetaiotaomicron* (16–20), has also been described. Until August, 2013, only glycoside hydrolases (GHs) have been implicated in *N*-glycan breakdown.

However, other types of CAZymes participate in the breakdown of complex carbohydrates, particularly those of plant origin, by working in synergy with GHs as follows: carbohydrate esterases (CE), polysaccharide lyases (PLs), and glycoside phosphorylases (GPs). The presence of CEs and PLs is readily detectable in gut bacterial genomes, metagenomes, and metatranscriptomes, because they are sufficiently divergent from GHs as to be classified in their own CAZyme families. This is not the case for GPs, which are found both in the glycosyltransferase (GT) and glycoside hydrolase families, depending on the sequence and catalytic mechanism similarities shared with GT and GH archetypes, respectively. The prevalence of GPs and their role in the metabolism of carbohydrates is therefore difficult to evaluate on the basis of sequence data alone, as a given family of GHs (or of GTs) may include both GHs (or GTs) as well as GPs. GPs catalyze the breakdown of a glycosidic linkage from oligosaccharide or polysaccharide substrates with concomitant phosphate glycosylation to yield a glycosyl-phosphate product and a sugar chain of reduced length. These enzymes are also able to perform reverse phosphorolysis (in the so-called "synthetic reaction") to form a glycosidic bond between the glycosyl unit originating from the glycosyl-phosphate, which acts as the sugar donor, and a carbohydrate acceptor (23). Retaining GPs, for which phosphorolysis occurs via overall retention of the substrate anomeric configuration, are found in CAZy families GT4, GT35, and GH13. Inverting GPs are classified in families GH65, GH94, GH112, and GH130. Inverting GH-related GPs and hydrolytic enzymes use a similar single displacement mechanism, differing in the requirement of GPs for a single catalytic residue (the proton donor), and inverting GHs for two catalytic residues. In GP-catalyzed reactions, the reaction begins with the direct nucleophilic attack by phosphate to the glycosidic bond with the aid of the catalytic residue, which donates a proton to the glycosidic oxygen atom and then proceeds through an oxocarbenium cation-like transition state. In GH reaction mechanisms, the nucleophilic attack of the C-1 of the glycoside is performed by a water molecule activated by the catalytic base. The natural structural and functional diversity of GPs thus appears to be highly restricted because of the following: (i) they are found in only 7 of the 226 GH and GT families listed in the CAZy database (March 2013); (ii) approximately only 15 EC entries are currently assigned to GPs (24); (iii) their specificity toward glycosyl phosphates is limited to α- and β-D-glucopyranose 1-phosphate (25–29), which are the most prevalent substrates; α-D-galactopyranose 1-phosphate (30); *N*-acetyl-α-D-glucosamine 1-phosphate (31), and α-D-mannopyranose 1-phosphate (32). In July 2013, α-D-mannopyranose 1-phosphate specificity was described for just one enzyme produced by a human gut bacterium, the *Bacteroides fragilis* NCTC 9343 mannosylglucose phosphorylase (BfMP),

which converts β-D-mannopyranosyl-1,4-D-glucopyranose and phosphate into α-D-mannopyranose 1-phosphate and D-glucose (32). This enzyme has been implicated in the catabolism of linear dietary mannans, assisted by a β-1,4-mannanase and a mannobiose 2-epimerase. During review of this study, Nihira *et al.* (33) reported the discovery of a metabolic pathway for *N*-glycans, which includes a β-D-mannosyl-*N*-acetyl-1,4-D-glucosamine phosphorylase named BT1033, produced by the human gut inhabitant *B. thetaiotamicron* VPI-5482. BfMP and BT1033 are two of the four enzymes to be characterized in the recently created GH130 family, which includes a total of 447 entries, from archaea, bacteria, and eukaryotes. The other characterized GH130 enzymes are RaMP1 and RaMP2 from the ruminal bacterium *Ruminococcus albus* NE1 (34). These enzymes have also been proposed to participate in mannan catabolism in the bovine rumen and are assisted by an endomannanase and an epimerase, via RaMP2- and RaMP1-catalyzed phosphorolysis of β-1,4-manno-oligosaccharides and 4-*O*-β-D-mannopyranosyl-D-glucopyranose, respectively (34). X-ray crystallographic studies show that GH130 enzymes share a 5-fold β-propeller fold. Currently, atomic coordinates data sets for four protein structures are available in the RCSB Protein Data Bank (35) as follows: BACOVA_03624 protein from *Bacteroides ovatus* ATCC 8483 (3QC2); BDI_3141 protein from *Parabacteroides distasonis* ATCC 8503 (3TAW); BT_4094 protein from *Bacteroides thetaiotamicron* VPI-5482 (3R67), and TM1225 protein from *Thermotoga maritima* MSB8 (1VKD). However, no function has yet been attributed to these four proteins, thus limiting the understanding of structure specificity relations for GH130 enzymes and the investigation of their catalytic mechanism.

Recently, the sequence of another GH130 enzyme, which we refer to as UhgbMP (unknown human gut bacterium mannoside phosphorylase, GenBank™ accession number ADD61463.1), was discovered by functional metagenomics of the human gut microbiota (36). The 36.6-kbp metagenomic DNA fragment containing the UhgbMP-encoding gene was taxonomically assigned to an as yet unidentified bacterium belonging to the genus *Bacteroides*. However, UhgbMP itself presents 99% protein sequence identity with the hypothetical protein BACSTE_03540 from *Bacteroides stercoris* ATCC 43183 (accession number EDS13361.1).

Here, we present an integrative approach, based on analyses of UhgbMP substrate specificity and of metagenomic and genomic data at the level of the entire human gut ecosystem, to reveal the role of UhgbMP and of 64 other GH130 enzymes produced by known gut bacteria in the breakdown of host and dietary mannose-containing glycans. In addition, we establish the molecular basis of GH130 enzyme catalysis, supported by the experimental results of rational engineering of UhgbMP and the analysis of its three-dimensional molecular model. Finally, we discuss the potential of this enzyme for the development of therapeutic agents and functional foods to inhibit *N*-glycan degradation, with the ultimate goal of protecting the epithelial barrier in the IBD context.

EXPERIMENTAL PROCEDURES

Recombinant UhgbMP Production and Purification—First, the UhgbMP encoding gene was PCR-amplified from the *Escherichia coli* metagenomic clone (GenBank™ accession number GU942931) using primers forward 5′-AGTATGAGTAGCAAAGTTATTATTCCTTGG-3′ and reverse 5′-TCAGATGATGCTTGTACGTTTGGTAAATTC-3′, by using the Expand Long Template PCR kit (Roche Applied Science).

To allow heterologous *UhgbMP* production in *E. coli* with His_6 tag at the N-terminal extremity, the PCR product was purified and subsequently cloned into the pCR8/GW/TOPO entry vector (Invitrogen), and then into the pDEST17 destination vector (Invitrogen), according to the manufacturer's recommendations. *E. coli* BL21-AI cells (Invitrogen) harboring the UhgbMP-encoding plasmid were cultured at 20 °C for 24 h in ZYM-5052 autoinduction medium (37) supplemented with 100 μg/ml ampicillin, inoculated at $A_{600\,nm}$ 0.1. Cells were harvested and resuspended in 20 mM Tris-HCl, pH 7.0, 300 mM NaCl, and lysed by sonication. Soluble lysate was applied to a TALON resin loaded with cobalt (GE Healthcare) equilibrated in 20 mM Tris-HCl, pH 7.0, 300 mM NaCl. After column washing with 8 volumes of the same buffer supplemented with 10 mM imidazole, the protein was eluted in 20 mM Tris-HCl, pH 7.0, 300 mM NaCl, 150 mM imidazole. Finally, the protein sample was desalted on a PD-10 column (GE Healthcare) and eluted in 20 mM Tris-HCl, pH 7.0, 0.1% Tween 80 (v/v). In these conditions, 84% of UhgbMP remained soluble after 8 days at 4 °C, thus allowing further functional characterization. The purity of the purified wild-type UhgbMP and mutants was evaluated higher than 95% by SDS-PAGE using Any kD™ Mini-PROTEAN® TGX™ Precast Gel (Bio-Rad) (supplemental Fig. 1). After migration, proteins were stained with the PageBlue Protein Staining Solution (Thermo Scientific) according to the manufacturer's recommendations. Protein concentrations were determined by spectrometry using a NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) The NanoDrop® measurement error was 5%. The calculated extinction coefficient of the purified UhgbMP fused to an N-terminal His^6 tag was 76 630 $M^{-1}\cdot cm^{-1}$.

UhgbMP Mutagenesis—Site-directed mutagenesis was performed by using the pDEST17 plasmid harboring the wild-type UhgbMP encoding gene as PCR template and the primers 5′-ATGGCTGTGCCAATACCGTAACCG-3′ and 5′-TACGGTATTGGCACAGCCATAGTAT-3′ for D304N mutant, 5′-GTATGGCTACAACCCACGCGTGTGCT-3′ and 5′-ACGCGTGGGTTGTAGCCATACACCCA-3′ to obtain the D104N mutant, 5′-AACCGTACCAATGCATGGGCGATGT-3′ and 5′-ATGCATTGGTACGGTTCGCGC-3′ for the E273Q mutant, and finally 5′-TGGCGAGGACCCGCGCGT-3′ and 5′-TCCTCGCCATACACCCAGGTACCGA-3′ for the Y103E mutant. The PCR products, amplified with the Phusion® High Fidelity DNA polymerase (New England Biolabs), were purified and digested by DpnI (New England Biolabs) before *E. coli* TOP10 transformation. Protein production and purification were identical for wild-type UgbMP and its variants.

Enzyme Assays—All reactions were carried out at 37 °C (wild-type UhgbMP optimal temperature) in 20 mM Tris-HCl, pH 7.0 (wild-type UhgbMP optimal pH). Syntheses of manno-oligosaccharides from α-D-mannopyranose 1-phosphate were performed with 0.1 mg/ml purified UhgbMP during 24 h at 37 °C in 20 mM Tris-HCl, pH 7.0, with 10 mM of α-D-mannopy-

ranose 1-phosphate (Sigma, reference M1755), and 10 mM of D-glucose, D-mannose, D-galactose, D-fructose, *N*-acetyl-D-glucosamine, β-D-mannopyranosyl-1,4-*N*,*N*′-diacetylchitobiose (Dextra, UK, reference MC0320), L-rhamnose, D-altrose, D-allose, D-fucose, L-fucose, D-mannitol, D-sorbitol, D-lyxose, xylitol, L-xylose, D-xylose, L-arabinose, or D-cellobiose.

Phosphorolysis kinetic parameters were determined with 0.01 mg/ml of purified enzyme by quantifying α-D-mannose 1-phosphate release rate from 0.5 to 10 mM inorganic phosphate (10 mM corresponding to the intracellular concentration of inorganic phosphate that was previously measured in bacteria (38)) and 1–10 mM of β-D-mannopyranosyl-1,4-D-glucose (Carbosynth, UK, reference OM04754), 0.4–4 mM β-1,4-D-mannan (Megazyme, Ireland, reference P-MANCB), 1–20 mM β-D-mannopyranosyl-1,4-D-mannose (Megazyme, Ireland, reference *O*-MBI), 0.1–1 mM *p*NP-β-D-mannopyranose, or 0.05–0.5 mM of β-D-mannopyranosyl-1,4-*N*,*N*′-diacetylchitobiose (Dextra, UK, reference MC0320). Reverse phosphorolysis kinetic parameters were determined with 0.01 mg/ml purified enzyme by quantifying α-D-mannopyranose 1-phosphate consumption rate from 0.5 to 10 mM α-D-mannopyranose 1-phosphate and 5–40 mM of D-mannose, D-glucose, D-galactose, D-fructose, or *N*-acetyl-D-glucosamine, or 0.1–40 mM of *N*,*N*′-diacetylchitobiose (Dextra, UK, reference C8002). The apparent kinetic parameters for phosphorolysis and reverse phosphorolysis at fixed initial concentrations of inorganic phosphate and α-D-mannose 1-phosphate, respectively, were determined by fitting the initial rates of α-D-mannose 1-phosphate release and consumption to the Michaelis-Menten equation. The kinetic parameters for phosphorolysis and synthesis of β-D-mannopyranosyl-1,4-*N*,*N*′-diacetylchitobiose and β-D-mannopyranosyl-1,4-D-mannose were determined by varying carbohydrates, α-D-mannose 1-phosphate or inorganic phosphate concentrations, and by fitting the initial rates of α-D-mannose 1-phosphate release and consumption to the sequential random Bi Bi mechanism equation (39). Nonlinear regression was performed with SigmaPlot enzyme kinetics module, version 1.3 (Systat Software, Inc., San Jose, CA). UhgbMP-specific activity toward β-1,4-D-manno-oligosaccharides, was determined with 0.1 mg/ml purified enzyme by quantifying α-D-mannopyranose 1-phosphate release rate from 10 mM inorganic phosphate and 10 mM β-D-manno-oligosaccharides of polymerization degree 2–6 (Megazyme, Ireland). The percentage of UhgbMP inhibition by exogenous carbohydrates or polyols was measured with 0.1 mg/ml purified enzyme by quantifying α-D-mannopyranose 1-phosphate consumption rate from 10 mM α-D-mannopyranose 1-phosphate as glycosyl donor, with and without 10 mM of L-rhamnose, D-altrose, D-allose, D-fucose, D-mannitol, D-sorbitol, D-lyxose, xylitol, L-xylose, D-xylose, L-arabinose, or D-cellobiose.

The percentage of reverse phosphorolysis activity of the D104N, D304N, and E273Q variants, compared with that of the wild-type enzyme, was determined with 0.1 mg/ml of purified proteins by quantifying α-D-mannopyranose 1-phosphate consumption rate from 10 mM α-D-mannopyranose 1-phosphate as glycosyl donor and 10 mM D-mannose as acceptor. α-D-Mannopyranose 1-phosphate was quantified by using high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Carbohydrates and α-D-mannopyranose 1-phosphate were separated on a 4 × 250 mm Dionex CarboPak PA100 column. A gradient of sodium acetate (from 0 to 150 mM in 15 min) and an isocratic step of 300 mM sodium acetate in 150 mM NaOH was applied at a 1 ml·min^{-1} flow rate. Detection was performed using a Dionex ED40 module with a gold working electrode and a Ag/AgCl pH reference. Finally, the hydrolytic or phosphorolytic behavior of the wild-type UhgbMP and its Y103E variant was assessed by using 1 mM *p*NP-β-D-mannopyranose, in the absence or presence, respectively, of 10 mM inorganic phosphate. The *p*NP release was monitored at $A_{405\,nm}$ on a carry-100 UV-visible spectrophotometer (Agilent Technologies). Between three and five independent experiments were carried out to determine initial activity, kinetic constants, and percentage of inhibition by carbohydrates and polyols of wild-type UhgbMP and its mutants. For all reaction rate measurements, it was checked by HPAEC-PAD that less than 10% of substrate was consumed and that the amount of consumed or released α-D-mannopyranose 1-phosphate increased linearly with time.

NMR Spectroscopy—Freeze-dried reaction media were exchanged twice with 99.9 atom % D_2O and lyophilized. Deuterium oxide was used as the solvent, and sodium 2,2,3,3-tetradeuterio-3-trimethylsilylpropanoate was selected as the internal standard. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker Advance 500-MHz spectrometer using a 5-mm z-gradient TBI probe at 298 K, an acquisition frequency of 500.13 MHz, and a spectral width of 8012.82 Hz. Spectra were acquired and processed using TopSpin 3.0 software. The various signals were assigned by comparison with signals obtained from α-D-mannopyranose 1-phosphate, β-D-mannopyranosyl-1,4-D-mannose (Megazyme, Irleland, reference *O*-MBI), β-D-mannopyranosyl-1,4-D-glucose (Carbosynth, UK, reference OM04754), or β-D-mannopyranosyl-1,4-*N*,*N*′-diacetylchitobiose (Dextra, UK, reference MC0320), used as standards.

Three-dimensional Molecular Modeling—The UhgbMP sequence was submitted to the I-TASSER server for automated protein structure and function prediction (40). The homologous *T. maritima* TM1225 structure (PDB accession code 1VKD) was used to provide spatial restraints. The three-dimensional model of UhgbMP predicted by I-TASSER was then further refined by energy minimization using the CFF91 force field implementation in the DISCOVER module of the InsightII software suite (Accelrys, San Diego). The CFF91 cross-terms, a harmonic bond potential, and a dielectric constant of 1.0 were specified in the energy function. An initial minimization was performed with positional restraints on the protein backbone using a steepest descent algorithm followed by conjugated gradient minimization until the maximum RMS energy gradient was less than 0.5 kcal mol^{-1} Å^{-1}. The system was then fully relaxed without positional restraints. α-D-Mannopyranose 1-phosphate, β-D-mannoheptaose, and β-D-mannopyranosyl-1,4-*N*,*N*′-diacetylchitobiose were manually docked into the active site of UhgbMP. The ligand complexes were then optimized according to the minimization protocol described above with the ligand molecules free to move. Molecular graphics images were produced using PyMOL software (Schrödinger, LLC).

GH130 Multiple Sequence Alignment Analyses—The 369 public sequences of GH130 enzymes listed in the CAZy database in January, 2013, were aligned with MUSCLE version 3.7 (41). A distance matrix was generated from the multiple sequence alignment using the BLOSUM62 amino acid residue substitution matrix. The output result file was subjected to hierarchical clustering using Ward's method (39), and the resulting tree was visualized using DENDROSCOPE 3 (42). Two sequence clusters were clearly apparent, members of which were accordingly assigned to the GH130_1 or GH130_2 subfamilies.

Position-dependent amino acid residue variation in multiple sequence alignment data were analyzed using the Shannon information entropy measure (H_X), calculated using SEQUESTER software. The Shannon entropy (H_X) at residue alignment position (X), corrected for the normalized frequency of residue type occurrence, is computed as shown in Equation 1,

$$H_X = \frac{\sum_{i=1}^{20} p_{(i|X)} \ln p_{(i|X)}}{\sum_{i=1}^{20} p_i \ln p_i} \qquad \text{(Eq. 1)}$$

where $p_{i|X}$ is the conditional probability of residue type i occurrence at alignment position X, and p_i is the normalized probability of residue type i occurrence at any position. To minimize sampling bias, normalized residue type probability values were taken as those documented by Ranganathan and co-workers (43),[3] garnered from sequence data for all natural proteins. Values of H_X lie in the range 0–1; a zero value corresponds to a fully conserved residue position, and a value of unity represents a distribution in which each residue type has an equal chance of occurrence. SEQUESTER appends Shannon entropies at aligned residue positions in a chosen reference protein structure to an atomic coordinate data file for convenient three-dimensional visual display.

GH130 Genomic Context Analysis—The analysis of the habitat of the organisms displaying these 369 GH130 sequences, as referenced in the GOLD database, allowed us to sort out 28 public genomes of human gut bacteria that display GH130 sequences. The genomic context of these 63 sequences, as well as of the UhgbMP and the GenBank™ ADD61810 sequences belonging to the metagenomic sequences GU942931 and GU942945, respectively, was analyzed to identify the CAZy encoding genes that are present in the same multigenic cluster as a GH130_1 or a GH130_2 encoding gene. CAZy encoding genes were searched on the same DNA strand as the GH130 encoding gene, with an increment of 10 kbp maximum upstream or downstream the GH130 sequence. For each glycoside hydrolase (GH), polysaccharide lyase (PL), or CE family, the frequency of co-occurrence in a multigenic cluster with a GH130_1 or a GH130_2 sequence was calculated, pondered by the number of GH130_1 or GH130_2 sequences, and used as edge attributes in a Cytoscape representation. In total, 74 and 62 co-occurrences of GH, PL, or CE sequences with GH130_1 or GH130_2 sequences, respectively, were counted. Prediction of transmembrane topology and signal peptides was performed using PHOBIUS.

[3] O. Rivoire, and R. Ranganathan, unpublished data.

TABLE 1

Apparent kinetic parameters for phosphorolysis and reverse phosphorolysis reactions catalyzed by UhgbMP, calculated from three to five independent experiments

Substrate concentrations were as follows: phosphorolysis, 10 mM inorganic phosphate and 1–20 mM β-D-mannopyranosyl-1,4-D-mannose; 1–10 mM β-D-mannopyranosyl-1,4-D-glucose, 0.4–4 mMβ-1,4-D-mannan, or 0.1–1mM *p*NP-β-D-mannopyranose, and 5 mM inorganic phosphate and 0.05–0.5 mM β-D-mannopyranosyl-1,4-*N,N′*-diacetylchitobiose; reverse phosphorolysis, 10 mM α-D-mannopyranose 1-phosphate and 5–40 mM D-glucose, D-mannose, D-galactose, D-fructose, or *N*-acetyl-D-glucosamine, and 5 mM α-D-mannopyranose 1-phosphate and 0.1–1 mM *N,N′*-diacetylchitobiose.

Phosphorolysis			
Glycoside substrates	$K_{m\ app}$ (mM)	$k_{cat\ app}$ (s^{-1})	$k_{cat\ app}/K_{m\ app}$ ($s^{-1}mM^{-1}$)
β-D-mannopyranosyl-1,4-D-mannose	1.1 ± 0.04	1.1 ± 0.01	1.04
β-D-mannopyranosyl-1,4-D-glucose	47 ± 6	61 ± 6	1.27
β-1,4-D-Mannan	2.2 ± 0.4	3.1 ± 0.3	1.41
β-D-mannopyranosyl-1,4-N,N'-diacetyl chitobiose	0.018 ± 0.003	0.072 ± 0.002	4.0
*p*NP-β-D-mannopyranose	0.4 ± 0.1	0.009 ± 0.0001	0.02
Reverse-phosphorolysis			
Glycoside acceptors	$K_{m\ app}$ (mM)	$k_{cat\ app}$ (s^{-1})	$k_{cat\ app}/K_{m\ app}$ ($s^{-1}mM^{-1}$)
D-glucose	30 ± 6	13 ± 2	0.44
D-mannose	15.3 ± 0.3	11.2 ± 0.1	0.73
D-galactose	16 ± 3	13 ± 1	0.80
D-fructose	2.9 ± 0.9	12.1 ± 0.8	4.24
N-acetyl-D-glucosamine	3.8 ± 0.6	13.5 ± 0.5	3.57
N,N'-diacetyl chitobiose	0.28 ± 0.03	9.3 ± 0.4	33.25

Analysis of GH130 Encoding Gene Prevalence in the Human Gut Microbiome—The 369 GH130 sequences referenced in the CAZy database in January, 2013, were searched by TBLASTN analysis (*E*-value = 0, identity ≥90%) in the catalog of prevalent genes (as defined by Qin *et al.* (3)) encoded on contigs of metagenomic sequences obtained from fecal samples of 162 individuals from the MetaHit cohort and of 139 individuals from the National Institutes of Health HMP cohort in the United States.

RESULTS

Uhgb_MP Substrate and Product Specificity—From the metagenomic sequence contained in the recombinant *E. coli* clone (accession number GU942931), we subcloned the UhgbMP encoding gene to produce a soluble protein tagged with a His_6 tag at the N terminus, with a yield of 32 mg of purified protein per liter of culture. We first characterized the substrate specificity of UhgbMP for carbohydrate phosphorolysis. In contrast to RaMP1 and BfMP, but similar to RaMP2 and to BT1033, UhgbMP exhibits a relaxed specificity toward carbohydrate substrates (Table 1). UhgbMP is able to phosphorolyze β-D-mannopyranosyl-1,4-D-glucopyranose, β-1,4-linked D-manno-oligosaccharides, and mannan (β-D-Man*p*-1,4-(D-Man*p*)$_n$, with n = 1–15), characterized by a notable increase in specific activity with the degree of polymerization (DP) (Fig. 2). UhgbMP is thus to date the only characterized GH130 enzyme that is able to break down mannan, a constituent of hemicellulose in grains and nuts (45). Interestingly, UhgbMP is also able to phosphorolyze β-D-mannopyranosyl-1,4-*N,N′*-diacetylchitobiose (β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc), a signature motif of human *N*-glycans. However, neither cellobiose (β-D-Glc*p*-1,4-D-Glc*p*) nor *N,N′*-diacetylchitobiose (β-D-Glc*p*NAc-1,4-D-Glc*p*NAc) could be phosphorolyzed, indicating that the UhgbMP subsite −1 is highly specific for mannosyl residues.

Finally, no trace of oligosaccharide-phosphate was visible, at any reaction time, on HPAEC-PAD chromatograms of prod-

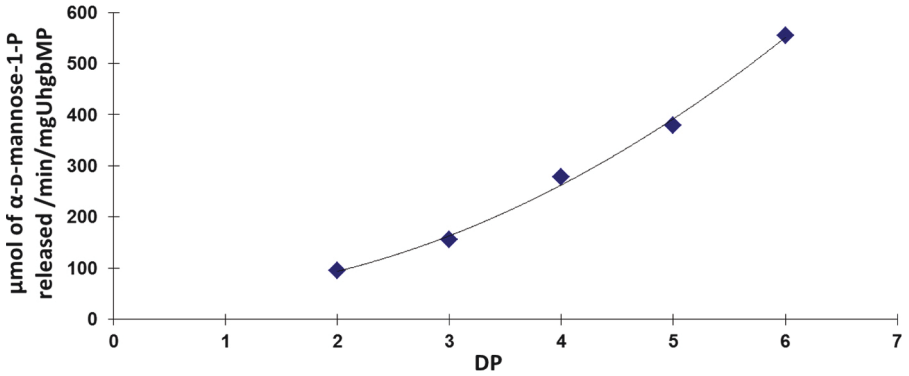


FIGURE 2. **Dependence of UhgbMP-specific activity with the DP of the phosphorolyzed β-1,4-D-manno-oligosaccharides.**

ucts obtained during phosphorolysis reactions, irrespective of the carbohydrate substrate. UhgbMP is thus an exo-acting enzyme, able to breakdown only the first β-mannosidic linkage at the nonreducing end of oligosaccharides.

α-D-Mannopyranose 1-phosphate was therefore tested as a glycosyl donor for reverse phosphorolysis. First, even without any carbohydrate acceptor, UhgbMP produces mannose and, furthermore, manno-oligosaccharides of DP ranging from 1 to 12, indicating that water itself plays the role of first acceptor at the beginning of the reaction. The β-1,4 regio-specific synthesis of manno-oligosaccharides was characterized by ^{1}H and ^{13}C NMR (supplemental Fig. 2). Various carbohydrates or polyols were tested as acceptors (Table 1). D-Glc*p*NAc and β-D-Glc*p*NAc-1,4-D-Glc*p*NAc were the best recognized acceptors, given that the UhgbMP K_m values for these compounds are 4- and 55-fold lower than for D-mannose, respectively. Starting from α-D-mannopyranose 1-phosphate and D-Glc*p*NAc or β-D-Glc*p*NAc-1,4-D-Glc*p*NAc, UhgbMP generates series of manno-oligosaccharides containing D-Glc*p*NAc or β-D-Glc*p*NAc-1,4-D-Glc*p*NAc at their reducing end, with a DP of up to 4. D-Glucose, D-mannose, D-galactose, and D-fructose are also recognized as acceptors.

The kinetic mechanism of UhgbMP was investigated by determining the initial reaction velocities for β-D-Man*p*-1,4-D-Man*p* and β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc phosphorolysis and synthesis at various initial concentrations of substrates. First, double-reciprocal plots of $1/v$ *versus* $1/[\beta\text{-D-Man}p\text{-}1,4\text{-D-Man}p]$ at various concentrations of inorganic phosphate and of $1/v$ *versus* $1/[\text{D-Man}p]$ at various concentrations of α-D-mannopyranose 1-phosphate crossed at a certain point, on the left of the $1/v$ axis (Fig. 3). UhgbMP thus catalyzes the β-D-Man*p*-1,4-D-Man*p* phosphorolysis and synthesis through a sequential Bi Bi mechanism involving the formation of a ternary complex. The equilibrium constant was calculated from the Haldane relationship for a sequential random Bi Bi mechanism (39) (the most probable mechanism fitting with this kinetic dataset) as shown in Equation 2,

$$K = (k_{cat}{}^{p}\, K_{S\ \alpha\text{-D-Man}p\text{-1-P}}\, K_{m\ \text{D-Man}p})/(k_{cat}{}^{s}\, K_{S\ \beta\text{-D-Man}p\text{-1,4-D-Man}p}\, K_{m\ \text{P}_i}) \quad \text{(Eq. 2)}$$

($k_{cat}{}^{s}\, K_{S\ \beta\text{-D-Man}p1,4\text{-D-Man}p}\, K_{m\ \text{Pi}}$), $k_{cat}{}^{p}$ and $k_{cat}{}^{s}$ being k_{cat} for phosphorolytic and synthetic reactions, respectively (Table 2). The value of 2.1 indicates that reaction equilibrium favors phosphorolysis. The kinetic parameters calculated for β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc phosphorolysis and synthesis are clearly different. β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc synthesis is inhibited by β-D-Glc*p*NAc-1,4-D-Glc*p*NAc concentrations higher than 1 mM (Fig. 3). For β-D-Glc*p*NAc-1,4-D-Glc*p*NAc and β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc concentrations lower than 1 and 0.5 mM, respectively, the most probable mechanism is also a mixed-type sequential random Bi Bi mechanism. The *K* value calculated in these conditions is 0.32, a value similar in magnitude to the value 0.658 obtained for D-Man*p*-1,4-β-D-Glc*p* phosphorolysis by RaMP2 (34). Further investigations of UhgbMP mechanism will be necessary to determine the order of substrate binding and product release and to better understand how this enzyme works on its natural substrates, including β-D-Man*p*-1,4-D-Glc*p*NAc.

Finally, with all other compounds tested as acceptors, UhgbMP synthesizes the same $(\text{Man})_n$ series as when α-D-mannopyranose 1-phosphate was the sole substrate. This shows that none of the other tested molecules were good acceptors, because water and additional mannose and manno-oligosaccharide units were used preferentially by the enzyme. However, the presence of cellobiose, D-fucose, L-rhamnose, L- and D-xylose, *N*-acetyl-D-galactosamine (D-Gal*p*NAc), D-altrose, D-allose, xylitol, D-lyxose, or D-mannitol decreased enzyme-specific activity, in some cases quite markedly (Table 3). But concentrations of these compounds did not decrease during reaction, and no significant additional product was produced compared with reaction in the presence of α-D-mannopyranose 1-phosphate as sole substrate. These carbohydrates thus act as UhgbMP inhibitors. Various sugar-phosphates were also tested as glycosyl donors for reverse phosphorolysis as follows: α-D-fructose-1- and -6-phosphate, D-ribose 1-phosphate, α-D-galactosamine 1-phosphate, α-D-glucosamine 1-phosphate, D-mannose 6-phosphate, and α-D-glucopyranosyl 1- and -6-phosphates. None of them was consumed by UhgbMP, and no reaction product appeared on HPAEC-PAD chromatograms after 24 h at 37 °C.

Key Residues Involved in Mannoside Phosphorolysis Mechanism—To investigate the UhgbMP catalytic mechanism, we first built a three-dimensional model of the enzyme using the atomic coordinates of the TM1225 protein from *T. maritima* MSB8 (PDB accession code 1VKD) as a structural template. The enzyme, which adopts a five-bladed β-propeller fold (Fig. 4*A*), was identified from structural genomics initiatives. Of the

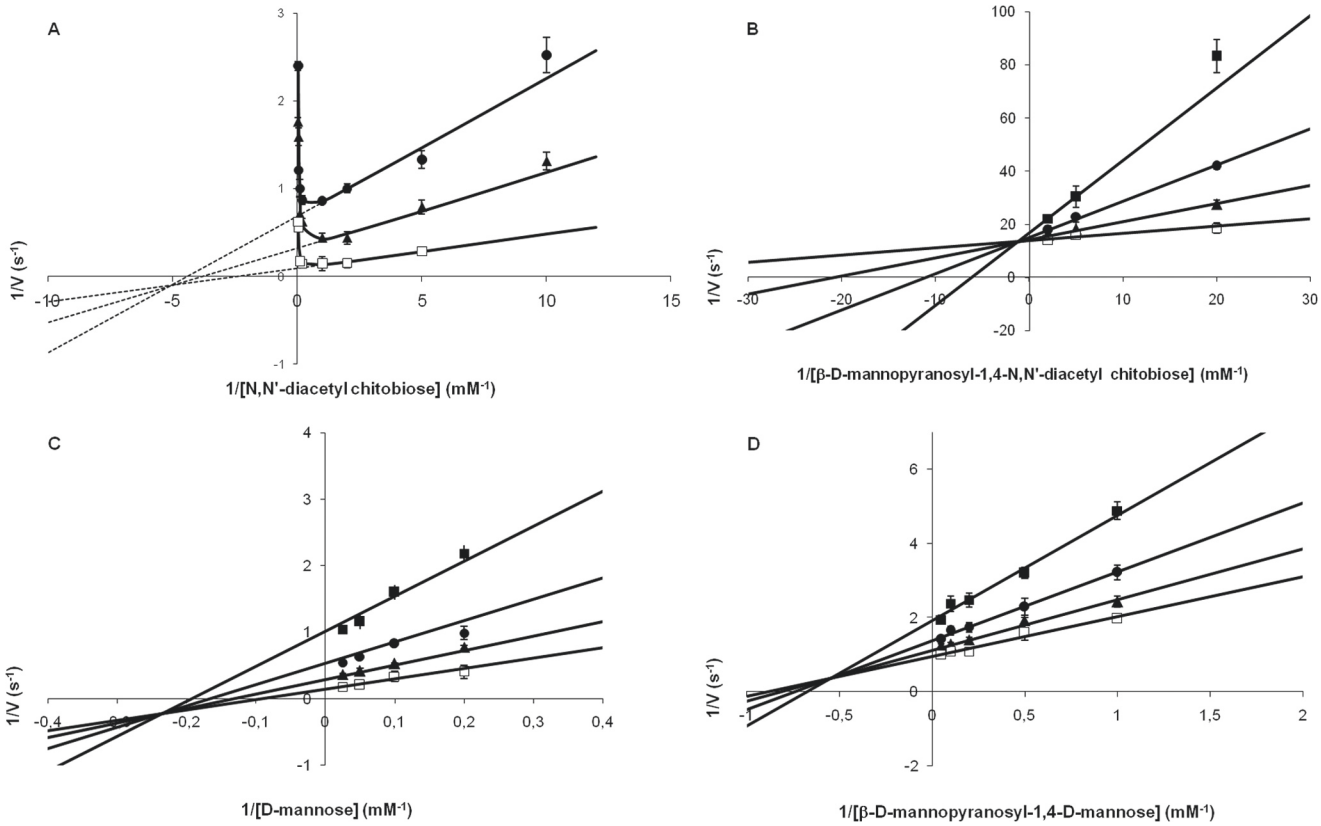


FIGURE 3. **Double-reciprocal plots for β-D-mannopyranosyl-1,4-*N,N′*-diacetylchitobiose synthesis (*A*) and phosphorolysis (*B*) and for β-D-mannopyranosyl-1,4-D-mannose synthesis (*C*) and phosphorolysis (*D*).** The initial velocities of α-D-mannose 1-phosphate consumption and release were measured at various concentrations of substrates. *A* and *C,* concentrations of α-D-mannose 1-phosphate were 0.5 mM (*filled square*), 1 mM (*filled circle*), 2 mM (*filled triangle*), and 5 mM (*open square*). *B* and *D,* concentrations of inorganic phosphate were 0.5 mM (*filled square*), 1 mM (*filled circle*), 2 mM (*filled triangle*) and 5 mM (*open square*). Data are the mean values ± S.D. for three to five independent experiments.

TABLE 2

Kinetic parameters for the phosphorolysis and synthesis of β-D-mannopyranosyl-1,4-D-mannose and β-D-mannopyranosyl-1,4-*N,N′*-diacetylchitobiose

Substrate concentrations were as follows: β-D-mannopyranosyl-1,4-D-mannose phosphorolysis, 1–20 mM β-D-mannopyranosyl-1,4-D-mannose, and 0.5–5 mM inorganic phosphate; β-D-mannopyranosyl-1, 4-D-mannose synthesis, 5–40 mM D-mannose, and 0.5–5 mM α-D-mannopyranose 1-phosphate; β-D-mannopyranosyl-1,4-*N,N′*-diacetylchitobiose phosphorolysis, 0.05–0.5 mM β-D-mannopyranosyl-1,4-*N,N′*-diacetylchitobiose, and 0.5–5 mM inorganic phosphate; β-D-mannopyranosyl-1,4-*N,N′*-diacetylchitobiose synthesis, 0.1–1 mM *N,N′*-diacetylchitobiose, and 1–5 mM α-D-mannopyranose 1-phosphate.

	β-D-Man*p*-1,4-D-Man*p*	
Phosphorolysis	k_{cat}^{P} (s^{-1})	1.17 ± 0.02
	$K_{m\ \beta\text{-D-Man}p\text{-}1,4\text{-D-Man}p}$ (mM)	1.8 ± 0.2
	$K_{m\ Pi}$ (mM)	1.1 ± 0.1
	$K_{S\ \beta\text{-D-Man}p\text{-}1,4\text{-D-Man}p}$ (mM)	1.0 ± 0.1
	$K_{S\ Pi}$ (mM)	0.64 ± 0.04
Synthesis	k_{cat}^{S} (s^{-1})	24 ± 6
	$K_{m\ D\text{-Man}p}$ (mM)	4.2 ± 0.3
	$K_{m\ \alpha\text{-D-Man}p\text{-}1\text{-P}}$ (mM)	1.8 ± 0.5
	$K_{S\ D\text{-Man}p}$ (mM)	27 ± 1
	$K_{S\ \alpha\text{-D-Man}p\text{-}1\text{-P}}$ (mM)	12 ± 4
	β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc	
Phosphorolysis	k_{cat}^{P} (s^{-1})	0.074 ± 0.001
	$K_{m\ \beta\text{-D-Man}p\text{-}1,4\text{-}\beta\text{-D-Glc}p\text{NAc-}1,4\text{-D-Glc}p\text{NAc}}$ (mM)	0.77 ± 0.03
	$K_{m\ Pi}$ (mM)	1 675 ± 180
	$K_{S\ \beta\text{-D-Man}p\text{-}1,4\text{-}\beta\text{-D-Glc}p\text{NAc-}1,4\text{-D-Glc}p\text{NAc}}$ (mM)	6 x 10^{-5} ± 2 x 10^{-6}
	$K_{S\ Pi}$ (mM)	0.13 ± 0.03
Synthesis	k_{cat}^{S} (s^{-1})	18.0 ± 0.3
	$K_{m\ \beta\text{-D-Glc}p\text{NAc-}1,4\text{-D-Glc}p\text{NAc}}$ (mM)	0.56 ± 0.02
	$K_{m\ \alpha\text{-D-Man}p\text{-}1\text{-P}}$ (mM)	42 ± 1
	$K_{S\ \beta\text{-D-Glc}p\text{NAc-}1,4\text{-D-Glc}p\text{NAc}}$ (mM)	0.19 ± 0.01
	$K_{S\ \alpha\text{-D-Man}p\text{-}1\text{-P}}$ (mM)	14.3 ± 0.4

four proteins of known structure that share the same fold, TM1225 aligns with the highest sequence identity to UhgbMP (59%), but it has yet to be functionally characterized.

TABLE 3

Percentage of UhgbMP inhibition by carbohydrates and polyols, quantified by measuring α-D-mannopyranose 1-phosphate consumption rate from 10 mM α-D-mannopyranose 1-phosphate as glycosyl donor, with and without 10 mM of carbohydrates or polyols

Inhibitor	% of inhibition
	0
L-Arabinose	2 ± 0.4
D-Cellobiose	37 ± 1
D-Fucose	12 ± 2
L-Fucose	11 ± 4
L-Rhamnose	31 ± 2
Xylitol	29.4 ± 0.4
D-Lyxose	19 ± 2
L-Xylose	58 ± 1
D-Mannitol	37 ± 4
D-Altrose	84 ± 3
D-Xylose	86.2 ± 0.2
D-Allose	94 ± 3

To identify putative catalytic amino acid residues, we analyzed the multiple alignment of the 369 protein sequences referenced in CAZy family GH130 (January 2013) (supplemental Fig. 3). The most conserved amino acid residue positions were then mapped onto the three-dimensional model of UhgbMP. Of these, Asp-104 and Asp-304 were found to be part of a narrow groove putatively considered to be the active site. Another conserved amino acid residue, Glu-273, present in a short α-helix turn section, was also identified in the groove.

These three amino acid residues were individually mutated to investigate their potential role in the UhgbMP catalytic

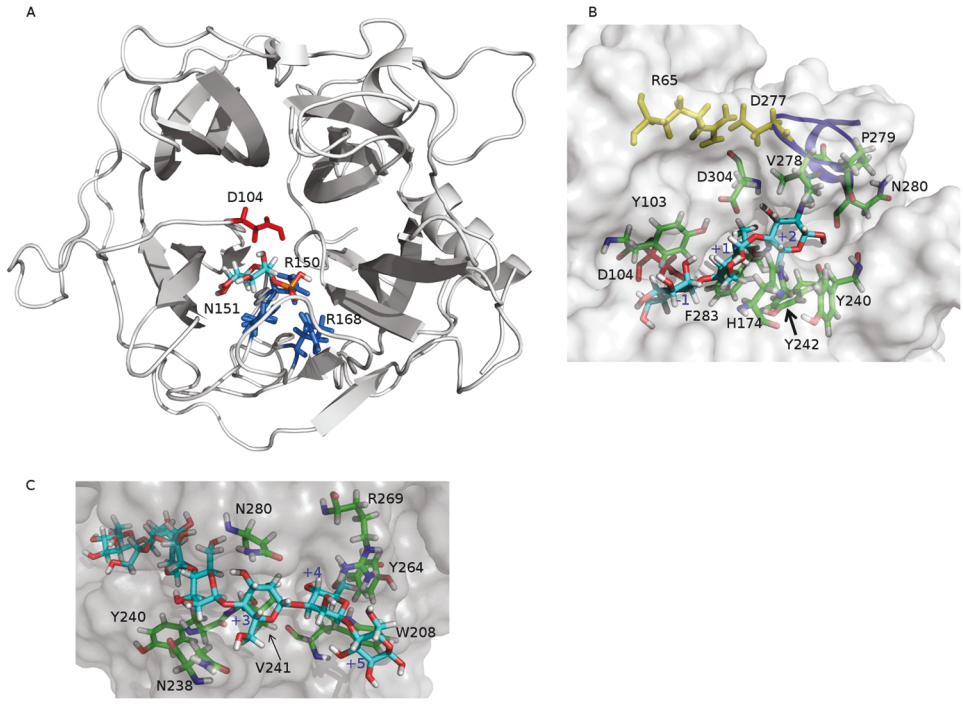


FIGURE 4. *A,* α-D-mannopyranose 1-phosphate docked into the three-dimensional model of UhgbMP. The most probable proton donor, Asp-104, is colored in *red*. Also shown in *blue* are the three phosphate-stabilizing residues, namely Arg-150, Asn-151, and Arg-168. *B,* β-D-mannopyranosyl-1,4-*N*,*N*′-diacetylchitobiose (β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc) docked into the three-dimensional model of UhgbMP. Tyr-103, Asp-304, His-174, Tyr-240, and Phe-283 are residues involved in recognition of the glycosyl residue in the +1 subsite. Asp-304, Tyr-242, Pro-279, Asn-280, and Val-278 are residues involved in binding of the glycosyl residue in the +2 subsite. *Yellow* shows Arg-65 and Asp-277 residues that form a salt bridge locking the Cys-274–Val-281 loop, represented as a *blue line*. *C,* β-1,4-D-mannoheptaose docked into the three-dimensional model of UhgbMP. Tyr-240, Asn-238, and Val-241 are residues involved in the substrate accomodation at the +3 subsite. Asn-280 and Asn-238 are residues involved in oligosaccharide binding at the +4 subsite. Tyr-264, Trp-208, and Arg-269 are residues shaping the +5 subsite.

mechanism. Results showed that mutation of Asp-104 into an asparagine completely abolished UhgbMP activity, although D304N and E273Q mutations retained only 3.9 ± 0.5 and 0.2 ± 0.1% of the native catalytic activity, respectively. These three amino acids were thus considered as candidate catalytic residues. To further elucidate the functional roles of these residues, molecular modeling techniques were used to dock α-D-mannopyranose 1-phosphate into the putative active site groove.

Of the identified docking modes, only one appeared compatible with spatial constraints provided by the UhgbMP reaction mechanism (Fig. 4*A*) as follows: (i) provision for the specific recognition of mannose at the −1 subsite; (ii) substrate interaction with at least one of the (Asp-104, Asp-304, or Glu-273) putative catalytic acidic residue side chains; (iii) presence of a favorable phosphate-binding site allowing reverse phosphorolysis to take place; (iv) presence of binding subsites able to accommodate β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc and β-1,4-D-mannan chains in catalytically productive binding modes with respect to active site residue(s) implicated in the chemical reaction mechanism (Fig. 4, *B* and *C*).

In this binding mode, where the α-D-mannopyranosyl 1-phosphate ring structure is stabilized in the −1 subsite through stacking interactions with Tyr-103, the Asp-104 residue located on the β-face of the catalytic chiral center can act as the unique proton donor during catalysis, whereas Arg-150, Arg-168, and Asn-151 amino acid residues can favorably assist phosphate group positioning consistent with inversion of configuration at C-1 of mannose (Fig. 4*A*). Given the high conservation of these amino acid residues within the GH130 family alignment, we suggest that enzymes contained in this family should share the same single displacement mechanism described for GH-like inverting phosphorylases. In such a mechanism, the aspartic acid corresponding to Asp-104 in UhgbMP would be the sole catalytic residue, having the role of proton donor (Fig. 5). The phosphate group, stabilized through ionic interactions with highly conserved Arg-150, Arg-168, and Asn-151 residues in UhgbMP, would then act as the nucleophile (Fig. 4*A*).

Molecular Basis of Substrate Specificity of GH130 Enzymes—The biochemical data from the kinetic characterization of BfMP, RaMP1, RaMP2, BT1033, and UhgbMP revealed marked specificity differences among the five enzymes. The BfMP and RaMP1 + 1 subsite is highly specific for glucose, whereas RaMP2, BT1033, and UhgbMP display looser specificity both toward carbohydrate substrates for phosphorolysis and acceptors for synthetic reactions. We therefore investigated whether the five characterized proteins could represent different GH130 subfamilies. Indeed, in most of the cases, enzymes classified in a CAZy subfamily share the same substrate and/or product specificity, reflecting a high degree of conservation in their active site (46, 47). A phylogenetic tree was constructed, based on the multiple alignments of the GH130 enzyme sequences. Three clusters of sequences clearly appeared. BfMP and RaMP1 are contained in subfamily GH130_1 (79 sequences, Fig. 6 and supplemental Table 1). UhgbMP, RaMP2, and BT1033 are contained in subfamily GH130_2 (42 sequences), together with

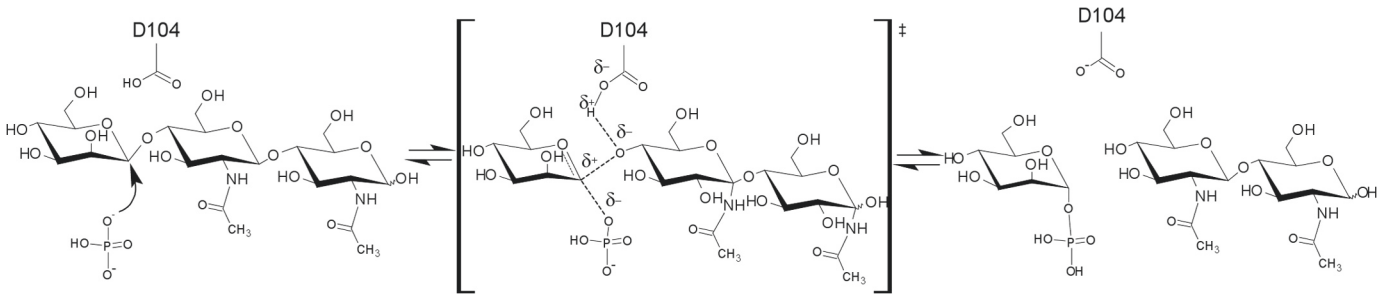


FIGURE 5. **Proposed catalytic mechanism of β-D-mannopyranosyl-1,4-*N*,*N*′-diacetylchitobiose phosphorolysis by UhgbMP, with Asp-104 acting as the proton donor.**

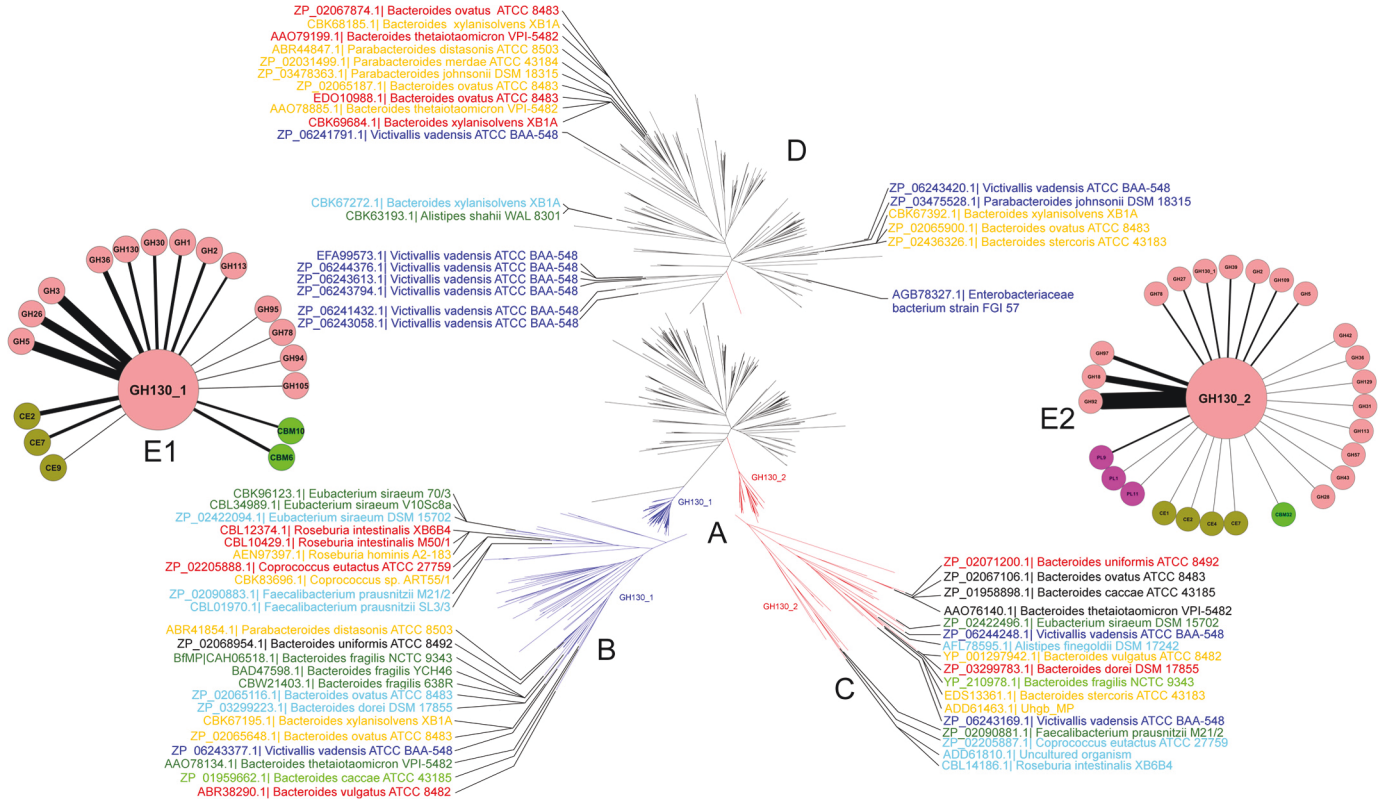


FIGURE 6. *A*, phylogenetic tree of the GH130 family. In this radial phylogram, branches corresponding to subfamilies GH130_1 and GH130_2 are colored in *blue* and *red*, respectively. Branches in *black* correspond to proteins that are not classified into any subfamily (GH130_NC). *B–D*, detailed views of the phylogentic tree of the GH130_1 and GH130_2 subfamilies and of not assigned sequences (GH130_NC). Only sequences corresponding to proteins from human gut bacteria are shown. Label color is relative to their prevalence in the human gut metagenome of 301 different individuals as follows. *Blue*, sequences found in the metagenome of 0–5 individuals; *light blue*, sequences found in the metagenome of 24–40 individuals; *green*, sequences found in the metagenome of 44–59 individuals; *light green*, sequences found in the metagenome of 62–76 individuals; *orange*, sequences found in the metagenome of 83–118 individuals; *red*, sequences found in the metagenome of 121–153 individuals; and *black*, sequences found in the metagenome of 173–197 individuals. *E*, cytoscape representations of CAZy encoding genes surrounding GH130_1 (*E1*) and GH130_2 (*E2*) encoding ones. This representation summarizes the number of co-presence of 65 known GH130 encoding genes and others CAZy ones in operon-like multigenic clusters of human gut bacteria genomes. Each gene encoding for a CAZy family found in a cluster containing a GH130 encoding gene is represented by a *node*. The *width* of the edges connecting the nodes is proportional to the number of times the co-presence of the two CAZymes encoding genes is observed in all the genomes analyzed.

TM1225 (PDB accession code 1VKD). The GH130_NC cluster contains 248 sequences of as yet uncharacterized proteins that are too heterogeneous to permit the creation of univocal subfamilies. This cluster contains the other three proteins of known structure, namely the BDI_3141 protein from *P. distasonis* ATCC 8503, the BT_4094 protein from *B. thetaiotamicron* VPI-5482, and the BACOVA_03624 protein from *B. ovatus* ATCC 8483. However, a key residue position occupied by Tyr-103 in UhgbMP allows the discrimination of the GH130_NC sequences from those of the two other subfamilies. The tyrosine residue at position 103 in UhgbMP, which lies close to the phosphate group binding site in the three-dimensional model of the docked complex of the enzyme with α-D-mannopyranose 1-phosphate, is strictly conserved in subfamilies GH130_1 and _2 but is replaced by a glutamic acid in 234 of the 248 sequences of the GH130_NC group (supplemental Fig. 3). Moreover, residues Arg-150, Arg-168, and Asn-151 that could favorably assist phosphate group positioning in UhgbMP active site are not conserved in the GH_130 NC cluster, although they are perfectly conserved in the GH130_1 and GH130_2 sequences (supplemental Fig. 3). We therefore suspect that the majority of the enzymes classified as GH130_NC

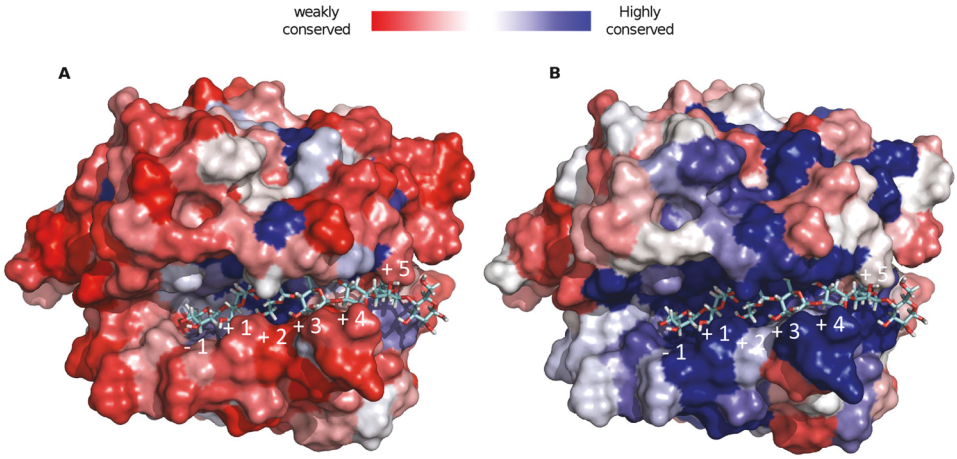


FIGURE 7. **Structural conservation of GH130 (*A*) and GH130_2 (*B*) enzymes, mapped onto the UhgbMP three-dimensional model, in which β-1,4-D-mannoheptaose has been docked.**

are hydrolases and not phosphorylases. The Glu residue corresponding to Tyr-103 in UhgbMP is expected to act as the second catalytic residue, taking on the role of base. To validate this hypothesis, we attempted to transform UhgbMP, which possesses low intrinsic hydrolase activity when assayed in the absence of inorganic phosphate, into a hydrolase, through the replacement of Tyr-Y103 by a glutamic acid. In the presence of 10 mM P_i, the *p*NP-β-D-mannopyranoside breakdown activity of the wild-type enzyme was $6.3 \times 10^{-3} \pm 3 \times 10^{-4}$ μmol·min^{-1}·mg^{-1}, a value similar to that obtained for mutant Y103E ($5.7 \times 10^{-3} \pm 1 \times 10^{-4}$ μmol·min^{-1}·mg^{-1}). In contrast, without any phosphate, the hydrolytic activity of the wild-type was $1.4 \times 10^{-3} \pm 1 \times 10^{-4}$ μmol·min^{-1}·mg^{-1}, although it was $3.8 \times 10^{-3} \pm 3 \times 10^{-4}$ μmol·min^{-1}·mg^{-1} for Y103E. Nevertheless, the Y103E mutant *p*NP release curve *versus* time reached a plateau after only 10 min reaction at 37 °C with and without phosphate, although less than 5% of substrate was consumed (supplemental Fig. 4). This phenomenon was also observed for the wild-type enzyme without phosphate. This indicates that inorganic phosphate is probably involved in maintaining the UhgbMP active site conformation and that the Y103E mutation may alter this conformation. Because the product yields are so weak without phosphate, it remains difficult to conclude that the Y103E mutant is really a hydrolase. Functional and structural investigations will thus be needed to confirm that the GH130_NC cluster contains mannoside hydrolases.

Our attempts to check correct folding of the Y103E variant, as well as that of the wild-type enzyme and of variants D104N, D304N, and E273Q by circular dichroïsm using a J-815 UV Spectrum spectropolarimeter (Jasco) failed, because of the high absorbance of Tween 80 at wavelengths between 200 and 290 nm (48).

To study the active site conservation in family GH130, we projected the results of the multiple alignments of the GH130 protein sequences onto the UhgbMP three-dimensional model, in which β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc or β-1,4-linked D-mannoheptaose (Fig. 7) was docked to enable the mapping of putative binding carbohydrate subsites. Clearly, residues lining the UhgbMP catalytic furrow are highly conserved only among the GH130_2 sequences. The +1 putative subsite appears delimited by Tyr-103, Asp-304, His-174, Tyr-240, and Phe-283 residues that are specifically conserved in the GH130_2 family (Fig. 4*B* and supplemental Fig. 5). In the presence of β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc, Tyr-103 and Asp-304 residues can form H-bond interactions with the *N*-acetyl-D-glucosamine positioned in the +1 subsite. In the presence of β-1,4-linked D-mannoheptaose, the Tyr-103 can interact with the mannosyl residue located in the +1 subsite, whereas Asp-304 establishes a hydrogen bond with the mannosyl moiety at +2 subsite. Asp-304 thus probably plays a key role in substrate binding at +1 and +2 subsites, explaining why its mutation dramatically alters UhgbMP activity.

The Tyr-242 and Phe-283 residues provide additional staking platforms at +1 subsite to stabilize the bound glycosyl unit. At the putative +2 subsite, the *N*-acetyl-D-glucosamine residue at the reducing end of the β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc is found stabilized through hydrogen bonding interactions with Tyr-242, Pro-279, Asn-280, and Asp-304, while establishing van der Waals interactions with Val-278. Interestingly, in our three-dimensional model of the active site, the *N*-acetyl groups of the D-Glc*p*NAc moieties docked at +1 and +2 subsites are found to fit nicely into a pocket formed by the loop Pro-271–Pro-284, which is locked in a "closed conformation" via a salt bridge ionic interaction between Asp-277 and Arg-65 residues (Fig. 4*B* and supplemental Fig. 5). The conformation of this loop may be modified by the E273Q mutation, thus altering oligosaccharide accommodation in the active site. Sequence analysis indicates that such a loop is very well conserved within the subfamily GH130_2 but less within GH130_1 (supplemental Fig. 3). This could thus suggest an involvement of this loop in the specificity determination of GH130_2 enzymes toward manno-oligosaccharides longer than DP2.

The +3 to +6 putative subsites have been mapped using the docked mannoheptosaccharide (Figs. 4*C* and 7 and supplemental Fig. 5). Surprisingly, very few aromatic residues able to provide stacking interactions with the manno-oligosaccharide chain have been found. However, a dense network of likely hydrogen bonding interactions and van der Waals contacts could explain the increase of UhgbMP catalytic efficiency with the polymerization degree of manno-oligosaccharides. In the +3 putative subsite, the mannosyl residue can be stabilized by

interactions with Tyr-240, Asn-238, and Val-241 residues, although at the putative +4 subsite, the sugar moiety is rather engaged in interactions with Asn-280 and Asn-238 residues. Residues Tyr-264, Trp-208, and Arg-269 are seen to establish interactions with the mannosyl moiety located at subsite +5. The length of the funnel binding site is found to accommodate up to +5 subsites. Outside this funnel, the +6 subsite has been putatively defined as sitting on the aromatic Trp-208 residue, which is almost perfectly conserved among the GH130_2 subfamily and not in GH130_1.

Finally, two important differences between GH130_1 and GH130_2 subfamilies, identified by sequence alignment analysis, could explain why GH130_1 enzymes seem to exhibit a very narrow specificity toward β-D-Man*p*-1,4-D-Glc*p*, as observed for RaMP1 and BfMP, although GH130_2 enzymes would be able to act on longer manno-oligosaccharides. First, the UhgbMP loop Gly-121–Gly-125, which defines the extremity of the −1 subsite, is very well conserved within the GH130_2 subfamily, whereas an insertion of 12 residues is observed in all GH130_1 sequences (supplemental Fig. 3). This longer loop could thus prevent the accommodation of long oligosaccharides in the negative subsite of GH130_1 enzymes. Second, we observed that the His-174 residue, which is not conserved in the GH130_1 subfamily and is contained in the UhgbMP Pro-169–Asp-179 loop, interacts with the sugar moiety bound at the +1 subsite. Interestingly, a motif containing five glycine residues was rather found within the GH130_1 subfamily in place of His-174 observed in GH130_2 enzymes that could play a major role in substrate accommodation within the active site (supplemental Fig. 3).

Genomic Context of GH130 Encoding Genes, Focus on Human Gut Bacteria—Differences in substrate specificity of enzymes belonging to the GH130_1 and GH130_2 subfamilies may illustrate different roles in the microbial ecosystem in which they are produced. 23 of the 79 GH130_1, 17 of the 42 GH130_2, and 25 of the 248 GH130_NC enzymes belong to human gut bacteria. To assess how globally prevalent these GH130 encoding genes are in gut microbiomes, we compared the 369 GH130 sequences with the human fecal metagenome sequences currently available, sampled from 162 individuals of the MetaHit cohort (3, 49) and 139 individuals of the NIH Human Microbiome Project (50). No less than 15 GH130_1, 10 GH130_2, and 14 GH130_NC sequences were detected in the fecal metagenome of at least 50 of the 301 individuals (Fig. 6). With the exception of four (one each from pig and chicken gut bacteria and two from cow rumen bacteria), these sequences all belong to human gut bacteria. The UhgbMP sequence was detected in the metagenomes of 93 of the 301 considered individuals. This indicates that it is not a rare gene and that it probably plays a critical role in mannose foraging in the gut. The highest occurrence values (sequences found in 173 to 197 individuals) were found for the GH130_1 sequence of *Bacteroides uniformis* ATCC 8492 and the GH130_2 sequences of *B. ovatus* ATCC 8483, *Bacteroides caccae* ATCC 43185, and *B. thetaiotamicron* VPI-5482 (BT1033 sequence) (GenBank™ accession numbers ZP_02068954.1, ZP_02067106.1, ZP_01958898.1, and AAO76140.1, respectively), which can be considered as common genes according to the Qin *et al.* definition (3), as they were found in more than 50% of the individuals. Interestingly, the mean number of GH130 BLAST hits per Mbp of sequence, obtained against the metagenomes of the 27 IBD patients (suffering either with Crohn disease or ulcerative colitis) of the MetaHit cohort, was 12, 28, and 58% higher than that obtained from the data sampled from the 135 other individuals (healthy or obese individuals), for GH130_1, GH130_2, and GH130_NC respectively (supplemental Table 2). This indicates a higher prevalence of GH130 encoding genes, and particularly of GH130_2 and GH130_NC, in the gut microbiome of IBD patients. This prevalence difference is thin, but it could be significant, considering that the gut microbiome of IBD patients harbors, on average, 25% fewer genes than that of healthy individuals (3).

To explore this question in more depth, we extracted from the GOLD database the 28 public genomes of human gut bacteria presenting sequences contained in the GH130_1, GH130_2, and GH130_NC subfamilies. We analyzed the genomic context of these 65 GH130 encoding genes, with a particular focus on the CAZy-encoding genes surrounding them. For each of these genomes, we highlighted operon-like multigenic clusters encoding CAZymes that act synergistically to break down complex carbohydrate structures, such as those contained in hemicellulose and human *N*-glycans. Interestingly, the 23 GH130_1 sequences are mainly surrounded by sequences of GH3, GH5, and GH26 (Fig. 6). These families contain enzymes that target hemicellulose, in particular β-mannanases and β-mannosidases. In contrast, the 17 GH130_2 sequences are mainly flanked by GH18 and GH92 and, to a lesser extent, GH97 sequences (Fig. 6 and supplemental Fig. 6). GH18 and GH92 contain the endo-β-*N*-acetylglucosaminidases and α1,2-, α1,3-, or α1,6-mannosidases that are implicated in the hydrolysis of human *N*-glycans (16, 17, 19, 20), although the only three characterized GH97 enzymes present α-glucosidase (EC 3.2.1.20) or α-galactosidase (EC 3.2.1.22) activity. Finally, only two multigenic clusters were found to encode both a GH130_2 enzyme and a GH2 enzyme, the GH2 family containing β-mannosidases capable of hydrolyzing the disaccharide β-D-Man*p*-1,4-D-GlcNAc (17).

The multigenic cluster containing UhgbMP encoding gene and that from *B. stercoris* ATCC 43183 containing its homolog BACSTE_03540 (accession number EDS13361.1) are clear examples of this gene organization with several GH92, GH18, and GH97 sequences surrounding a GH130_2 encoding gene (Fig. 1 and supplemental Fig. 6). In these cases, the results of *in silico* detection of signal peptide or transmembrane topology indicate that UhgbMP and BACSTE_03540 are probably intracellular, although the glycoside-hydrolases encoded by the same multigenic clusters would rather be secreted.

DISCUSSION

The biochemical characterization of UhgbMP, a representative of the GH130 enzyme family, revealed its flexibility toward carbohydrate substrates. UhgbMP is capable of catalyzing the phosphorolysis of β-D-mannopyranosyl-1,4-D-glucopyranose, β-1,4-D-manno-oligosaccharides of DP >5 and mannan, as well as the *N*-glycan core oligosaccharide β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc. Based on a sequence analysis of the GH130 family and on a structural model of UhgbMP, supported by the experimental findings presented here and else-

where by Senoura *et al.* (32), Nihira *et al.* (33), and Kawahara *et al.* (34), we propose the creation of two GH130 subfamilies, of which the members probably share the same single displacement mechanism involving a single catalytic acidic residue (corresponding to UhgbMP D104) acting as the proton donor. Subfamily GH130_1 gathers together enzymes (including BfMP and RaMP1) exhibiting narrow specificity toward β-D-Man*p*-1,4-D-Glc. Conversely, enzymes of the GH130_2 subfamily show high promiscuity toward their substrates and products. Their active sites would be sufficiently extensible to accommodate complex carbohydrate structures such as β-D-mannan and β-D-manno-oligosaccharides, with or without β-D-Glc*p*NAc-1,4-D-GlcNAc, D-Glc*p*NAc, or D-glucose at their reducing end. The resolution of crystallographic structures of representatives of each of GH130 clusters in complex with substrates and products will, however, be necessary to confirm the role of the key residues identified here and to deepen our understanding of the reaction mechanisms operating in these enzymes.

Among the 369 enzymes archived in the GH130 family (January, 2013), 65 belong to human gut bacteria. The strong prevalence of the corresponding genes, in particular that coding for UhgbMP, in the human gut metagenome suggests that these enzymes have a major role for foraging mannose in this ecosystem. Here, we have demonstrated that UhgbMP, and probably also the 16 other enzymes in the proposed GH130_2 subfamily produced by gut bacteria (supplemental Table 1), including BT1033, are able to participate in breaking down the host mannosylated glycoproteins lining the intestinal epithelium.

In vivo, incorporation of GH reaction products into metabolic pathways is energy-consuming, although direct metabolism of the glycoside phosphates synthesized by GPs does not require ATP. Phosphorolysis reactions mediated by GPs would thus be more advantageous than hydrolysis under certain physiological conditions, for example during intensive use of carbohydrate resources or in anoxic environments such as the gastrointestinal tract, where ATP cannot be efficiently produced by the respiratory chain-linked phosphorylation process (51).

In common with other GPs, especially those previously identified in microorganisms having a facultatively anaerobic lifestyle like *Bifidobacterium* sp., *Lactobacillus* sp., or *Clostridium* sp (50), enzymes in the proposed GH130_1 and GH130_2 families can therefore be considered as catabolic enzymes, working in tandem with catabolic GHs in dietary and host mannoside breakdown.

Based on the analysis of the genomic context of GH130_2 encoding genes in gut bacteria, and accordingly with Nihira *et al.* (33), we propose that mannoside phosphorolysis catalyzed by GH130_2 enzymes acts in concert with GH activities of enzymes in the GH18 and GH92 families, to completely break down *N*-glycans, as proposed in Fig. 1. The role of the GH97s whose genes belong to the same multigenic systems as those coding for the proposed GH130_2 enzymes (for example, in *B. stercoris* ATCC 43183) is less clear. Family GH97 contains only three characterized enzymes, of which two are α-glucosidases with α-1,4-link specificities (52, 53). However, the functional diversity of GH97 enzymes has not yet been thoroughly explored, and if these enzymes were previously thought to contribute to dietary carbohydrates in the human gut, they may well also be active in the breakdown of *N*-glycans, for example by liberating, like the GH99 enzymes (18), the α-1,3-linked glucosyl residue at the nonreductive end of the immature *N*-glycans. In the particular case of the unknown gut bacterium producing UhgbMP and of *B. stercoris* ATCC 43183 producing its homolog BACSTE_03540, and based on the results of *in silico* detection of signal peptides, we assume that the UhgbMP and BACSTE_03540 physiological role would be the intracellular phosphorolysis of short oligosaccharides that can be internalized in the cell (di- or trisaccharides (54)), like the β-D-Man*p*-1,4-D-Glc*p*NAc disaccharide or the β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc trisaccharide resulting from extracellular hydrolysis of *N*-glycans by glycoside hydrolases belonging to the GH18, GH92, and maybe also GH97 families. This deglycosylation arsenal thus allows one to deprotect the glycoproteins of the intestinal epithelium. Their protein part could also thus be degraded by the endopeptidases, similarly to those putatively encoded by the genes ADD61465.1 and ADD61469.1, which belong to the same metagenomic DNA fragment than the UhgbMP encoding gene, so as to perforate the intestinal epithelium (Fig. 1). However, further studies will be needed to confirm the physiological role of these enzymes, like metabolomic and transcriptomic analyses of gut bacteria producing GH130_2 enzymes (like *B. stercoris* ATCC 43183) in the presence of *N*-glycans as carbon source.

The results presented here show that *in vitro*, UhgbMP is also able to effectively degrade manno-oligosaccharides and mannan itself. This is also probably the case for other gut bacterial enzymes of the proposed GH130_2 subfamily, which show a similar active site topology, allowing the accommodation of long oligosaccharides. The intestinal bacteria that produce GH130_2 enzymes would thus be able to use dietary hemicelluloses and their hydrolysis products as carbon sources if needed, which would give them a competitive advantage with respect to the other bacteria, to colonize and maintain themselves in the intestinal tract. The use of mannose-rich carbohydrates (plant or yeast mannans, linear β-1,4-linked manno-oligosaccharides, or β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc) as functional foods could thus allow the metabolism of the *N*-glycan-degrading bacteria to be bypassed and, eventually, may reduce the degradation of the epithelial barrier for therapeutic applications, especially for patients suffering from IBD. Indeed, in their gut metagenome, the known GH130_2 and GH130_NC encoding genes seem highly prevalent, even if enzyme assays on fecal samples, as well as metatranscriptomic and metaproteomic studies, should be performed as evidence that GH130 could be biomarkers of IBDs. However, this approach presents a risk of overfeeding this specific type of bacteria and increasing their prevalence in the gut, and thus the potential for *N*-glycan breakdown, in the case of reduced intake of exogenous mannose-rich carbohydrates. The utilization of GH130 enzyme inhibitors, like D-altrose, D-xylose, and D-allose for UhgbMP or suicide inhibitors that could be specifically designed thanks to the available or future structural data, would thus be the preferred strategy in therapeutic contexts.

Finally, we have shown here that the relaxed specificity displayed by UhgbMP toward acceptor substrates can be used for the stereo- and region-selective synthesis of original glyco-con-

jugates and oligosaccharides. In particular, this enzyme is able to synthesize β-1,4-linked D-manno-oligosaccharides, which have been reported to present prebiotic properties (55), even if their use as functional food may be risky, from our point of view. UhgbMP is also highly effective for production of *N*-glycan core oligosaccharides, such as β-D-Man*p*-1,4-D-GlcNAc and β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc, whose commercial price today exceeds $10,000 per mg. UhgbMP-based β-mannoside synthesis processes appear as highly attractive compared with those based on mannosyltransferases, which use expensive activated sugar nucleotides as donors (56, 57), or on transmannosylation catalyzed by native or engineered mannosidases or mannanases (44, 58–62). Indeed, a two-step UhgbMP-based process would allow one to use phosphate and β-mannan as substrates to first catalyze phosphorolysis and second to reverse phosphorolysis in the presence of hydroxylated acceptors to synthesize mannosylated products. UhgbMP is thus an enzymatic tool with high potential for synthesizing molecules like *N*-glycan core oligosaccharides, of major importance for studying, and potentially also for controlling interactions between host and gut microbes.

Acknowledgments—We cordially thank Nelly Monties for technical assistance. MetaToul (Metabolomics & Fluxomics Facitilies, Toulouse, France) is gratefully acknowledged for access to NMR facilities. MetaToul is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, the SICOVAL, the Infrastructures en Biologie Sante et Agronomie (IBiSa, France), the CNRS, and the Institut National de la Recherche Agronomique.

REFERENCES

1. Dicksved, J., Halfvarson, J., Rosenquist, M., Järnerot, G., Tysk, C., Apajalahti, J., Engstrand, L., and Jansson, J. K. (2008) Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J.* **2,** 716–727
2. Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., and Dore, J. (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55,** 205–211
3. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., MetaHIT Consortium (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65
4. Erickson, A. R., Cantarel, B. L., Lamendella, R., Darzi, Y., Mongodin, E. F., Pan, C., Shah, M., Halfvarson, J., Tysk, C., Henrissat, B., Raes, J., Verberkmoes, N. C., Fraser, C. M., Hettich, R. L., Jansson, J. K. (2012) Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS ONE* **7,** e49138
5. Ottman, N., Smidt, H., de Vos, W. M., and Belzer, C. (2012) The function of our microbiota: who is out there and what do they do? *Front. Cell Infect. Microbiol.* **2,** 1–11
6. Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F. D., and Lewis, J. D. (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–108
7. Koropatkin, N. M., Cameron, E. A., and Martens, E. C. (2012) How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10,** 323–335
8. Englyst, K. N., and Englyst, H. N. (2005) Carbohydrate bioavailability. *Br. J. Nutr.* **94,** 1–11
9. Brett, C. T., and Waldron, K. W. (1996) in *Physiology and Biochemistry of Plant Cell Walls* (Black, M., and Charlwood, B. V., eds) p. 222–238, Chapman and Hall, London
10. Moran, A. P., Gupta, A., and Joshi, L. (2011) Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. *Gut* **60,** 1412–1425
11. Molinari, M. (2007) *N*-Glycan structure dictates extension of protein folding or onset of disposal. *Nat. Chem. Biol.* **3,** 313–320
12. Sheng, Y. H., Hasnain, S. Z., Florin, T. H., and McGuckin, M. A. (2012) Mucins in inflammatory bowel diseases and colorectal cancer. *J. Gastroenterol. Hepatol.* **27,** 28–38
13. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37,** D233–D238
14. Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., and Nelson, K. E. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312,** 1355–1359
15. Moreira, L. R., and Filho, E. X. (2008) An overview of mannan structure and mannan-degrading enzyme systems. *Appl. Microbiol. Biotechnol.* **79,** 165–178
16. Zhu, Y., Suits, M. D., Thompson, A. J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K. W., Xiang, Y., Siriwardena, A., Williams, S. J., Gilbert, H. J., and Davies, G. J. (2010) Mechanistic insights into a Ca^{2+}-dependent family of α-mannosidases in a human gut symbiont. *Nat. Chem. Biol.* **6,** 125–132
17. Tailford, L. E., Money, V. A., Smith, N. L., Dumon, C., Davies, G. J., and Gilbert, H. J. (2007) Mannose foraging by *Bacteroides thetaiotaomicron* structure and specificity of the β-mannosidase, BtMan2A. *J. Biol. Chem.* **282,** 11291–11299
18. Thompson, A. J., Williams, R. J., Hakki, Z., Alonzi, D. S., Wennekes, T., Gloster, T. M., Songsrirote, K., Thomas-Oates, J. E., Wrodnigg, T. M., Spreitz, J., Stütz, A. E., Butters, T. D., Williams, S. J., and Davies, G. J. (2012) Structural and mechanistic insight into *N*-glycan processing by endo-α-mannosidase. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 781–786
19. Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., and Gordon, J. I. (2003) A genomic view of the human *Bacteroides thetaiotaomicron* symbiosis. *Science* **299,** 2074–2076
20. Martens, E. C., Chiang, H. C., and Gordon, J. I. (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4,** 447–457
21. Deleted in proof
22. Deleted in proof
23. Nakai, H., Kitaoka, M., Svensson, B., and Ohtsubo, K. (2013) Recent development of phosphorylases possessing large potential for oligosaccharide synthesis. *Curr. Opin. Chem. Biol.* **17,** 1–9
24. Desmet, T., and Soetaert, W. (2012) Broadening the synthetic potential of disaccharide phosphorylases through enzyme engineering. *Process Biochem.* **47,** 11–17
25. Han, S.-E., Kwon, H.-B., Lee, S.-B., Yi, B.-Y., Murayama, I., Kitamoto, Y., and Byun, M. O. (2003) Cloning and characterization of a gene encoding trehalose phosphorylase (TP) from *Pleurotus sajor-caju*. *Protein Expr. Purif.* **30,** 194–202
26. Van der Borght, J., Chen, C., Hoflack, L., Van Renterghem, L., Desmet, T., and Soetaert, W. (2011) Enzymatic properties and substrate specificity of the trehalose phosphorylase from *Caldanaerobacter subterraneus*. *Appl. Environ. Microbiol.* **77,** 6939–6944
27. Rathore, R. S., Garg, N., Garg, S., and Kumar, A. (2009) Starch phosphorylase: role in starch metabolism and biotechnological applications. *Crit. Rev. Biotechnol.* **29,** 214–224
28. Aerts, D., Verhaeghe, T. F., Roman, B. I., Stevens, C. V., Desmet, T., and Soetaert, W. (2011) Transglucosylation potential of six sucrose phosphorylases toward different classes of acceptors. *Carbohydr. Res.* **346,** 1860–1867

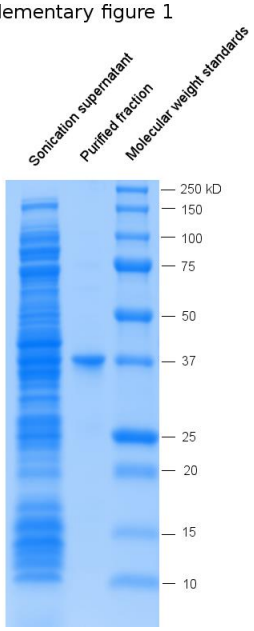
29. Nakai, H., Hachem, M. A., Petersen, B. O., Westphal, Y., Mannerstedt, K., Baumann, M. J., Dilokpimol, A., Schols, H. A., Duus, J. Ø., and Svensson, B. (2010) Efficient chemoenzymatic oligosaccharide synthesis by reverse phosphorolysis using cellobiose phosphorylase and cellodextrin phosphorylase from *Clostridium thermocellum*. *Biochimie* **92,** 1818–1826
30. Nakajima, M., and Kitaoka, M. (2008) Identification of lacto-*N*-biose I phosphorylase from *Vibrio vulnificus* CMCP6. *Appl. Environ. Microbiol.* **74,** 6333–6337
31. Hidaka, M., Honda, Y., Kitaoka, M., Nirasawa, S., Hayashi, K., Wakagi, T., Shoun, H., and Fushinobu, S. (2004) Chitobiose phosphorylase from *Vibrio proteolyticus*, a member of glycosyltransferase family 36, has a clan GH-L-like (α/α)6 barrel fold. *Structure* **12,** 937–947
32. Senoura, T., Ito, S., Taguchi, H., Higa, M., Hamada, S., Matsui, H., Ozawa, T., Jin, S., Watanabe, J., Wasaki, J., and Ito, S. (2011) New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase. *Biochem. Biophys. Res. Commun.* **408,** 701–706
33. Nihira, T., Suzuki, E., Kitaoka, M., Nishimoto, M., Ohtsubo, K., and Nakai, H. (2013) Discovery of β-1,4-D-mannosyl-*N*-acetyl-D-glucosamine phosphorylase involving in the metabolism of *N*-glycans. *J. Biol. Chem.* **288,** 27366–27374
34. Kawahara, R., Saburi, W., Odaka, R., Taguchi, H., Ito, S., Mori, H., and Matsui, H. (2012) Metabolic mechanism of mannan in a ruminal bacterium, *Ruminococcus albus*, involving two mannoside phosphorylases and cellobiose 2-epimerase discovery of a new carbohydrate phosphorylase, β-1,4-manno-oligosaccharide phosphorylase. *J. Biol. Chem.* **287,** 42389–42399
35. Berman, H., Henrick, K., Nakamura, H., and Markley, J. L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **35,** D301–D303
36. 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., and Potocki-Veronese, G. (2010) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* **20,** 1605–1612
37. Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41,** 207–234
38. Motomura, K., Hirota, R., Ohnaka, N., Okada, M., Ikeda, T., Morohoshi, T., Ohtake, H., and Kuroda, A. (2011) Overproduction of YjbB reduces the level of polyphosphate in *Escherichia coli*: a hypothetical role of YjbB in phosphate export and polyphosphate accumulation. *FEMS Microbiol. Lett.* **320,** 25–32
39. Segel, Irwin, H. (1975) *Enzyme Kinetics–Behaviour and Analysis of Rapid Equilibrium and Steady-state Enzyme Systems,* pp. 274–320, John Wiley & Sons, Inc., New York
40. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–738
41. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797
42. Huson, D. H., and Scornavacca, C. (2012) Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61,** 1061–1067
43. Süel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2002) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Mol. Biol.* **10,** 59–69
44. Dilokpimol, A., Nakai, H., Gotfredsen, C. H., Baumann, M. J., Nakai, N., Abou Hachem, M., and Svensson, B. (2011) Recombinant production and characterisation of two related GH5 endo-β-1,4-mannanases from *Aspergillus nidulans* FGSC A4 showing distinctly different transglycosylation capacity. *Biochim. Biophys. Acta* **1814,** 1720–1729
45. Dey, P. M. (1978) in *Advances in Carbohydrate Chemistry and Biochemistry* (Tipson, R. S., and Horton, D., eds) pp. 341–376, Elsevier, New York
46. Stam, M. R., Danchin, E. G., Rancurel, C., Coutinho, P. M., and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: toward improved functional annotations of α-amylase-related proteins. *Protein Eng. Des. Sel.* **19,** 555–562
47. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H., 3rd, and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* **12,** 186
48. Hu, L., Zhang, N., Yang, G., and Zhang, J. (2011) Effects of Tween-80 on the dissolution properties of daidzein solid dispersion *in vitro*. *Int. J. Chem.* **3,** 68–73
49. Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Doré, J., MetaHIT Consortium, Antolín, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., (2011) Enterotypes of the human gut microbiome. *Nature* **473,** 174–180
50. NIH HMP Working Group, Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., Bonazzi, V., McEwen, J. E., Wetterstrand, K. A., Deal, C., Baker, C. C., Di Francesco, V., Howcroft, T. K., Karp, R. W., Lunsford, R. D., Wellington, C. R., Belachew, T., Wright, M., Giblin, C., David, H., Mills, M., Salomon, R., Mullins, C., Akolkar, B., Begg, L., Davis, C., Grandison, L., Humble, M., Khalsa, J., Little, A. R., Peavy, H., Pontzer, C., Portnoy, M., Sayre, M. H., Starke-Reed, P., Zakhari, S., Read, J., Watson, B., and Guyer, M. (2009) The NIH Human Microbiome Project. *Genome Res.* **19,** 2317–2323
51. Luley-Goedl, C., and Nidetzky, B. (2010) Carbohydrate synthesis by disaccharide phosphorylases: reactions, catalytic mechanisms and application in the glycosciences. *Biotechnol. J.* **5,** 1324–1338
52. Hughes, C. V., Malki, G., Loo, C. Y., Tanner, A. C., and Ganeshkumar, N. (2003) Cloning and expression of α-D-glucosidase and *N*-acetyl-β-glucosaminidase from the periodontal pathogen, *Tannerella forsythensis* (*Bacteroides forsythus*). *Oral Microbiol. Immunol.* **18,** 309–312
53. Gloster, T. M., Turkenburg, J. P., Potts, J. R., Henrissat, B., and Davies, G. J. (2008) Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora. *Chem. Biol.* **15,** 1058–1067
54. Heinken, A., Sahoo, S., Fleming, R. M., and Thiele, I. (2013) Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* **4,** 28–40
55. Salinardi, T. C., Rubin, K. H., Black, R. M., and St-Onge, M.-P. (2010) Coffee manno-oligosaccharides, consumed as part of a free-living, weight-maintaining diet, increase the proportional reduction in body volume in overweight men. *J. Nutr.* **140,** 1943–1948
56. Revers, L., Bill, R. M., Wilson, I. B., Watt, G. M., and Flitsch, S. L. (1999) Development of recombinant, immobilised β-1,4-mannosyltransferase for use as an efficient tool in the chemoenzymatic synthesis of *N*-linked oligosaccharides. *Biochim. Biophys. Acta* **1428,** 88–98
57. Zhao, Y., and Thorson, J. S. (1999) Chemoenzymatic synthesis of the Salmonella group E1 core trisaccharide using a recombinant β-(1→4)-mannosyltransferase. *Carbohydr. Res.* **319,** 184–191
58. Ishimizu, T., Sasaki, A., Okutani, S., Maeda, M., Yamagishi, M., and Hase, S. (2004) Endo-β-mannosidase, a plant enzyme acting on *N*-glycan: purification, molecular cloning, and characterization. *J. Biol. Chem.* **279,** 38555–38562
59. Usui, T., Suzuki, M., Sato, T., Kawagishi, H., Adachi, K., and Sano, H. (1994) Enzymic synthesis of the trisaccharide core region of the carbohydrate chain of *N*-glycoprotein. *Glycoconj. J.* **11,** 105–110
60. Murata, T., and Usui, T. (1997) Preparation of oligosaccharide units library and its utilization. *Biosci. Biotechnol. Biochem.* **61,** 1059–1066
61. Sasaki, A., Ishimizu, T., and Hase, S. (2005) Substrate specificity and molecular cloning of the lily endo-β-mannosidase acting on *N*-glycan. *J. Biochem.* **137,** 87–93
62. Eneyskaya, E. V., Sundqvist, G., Golubev, A. M., Ibatullin, F. M., Ivanen, D. R., Shabalin, K. A., Brumer, H., and Kulminskaya, A. A. (2009) Transglycosylating and hydrolytic activities of the β-mannosidase from *Trichoderma reesei*. *Biochimie* **91,** 632–638

Supplemental information

Supplementary Figure 1:

SDS-PAGE gel of the sonication supernatant of recombinant *E. coli* cells producing Uhgb_MP, and of purified Uhgb_MP.

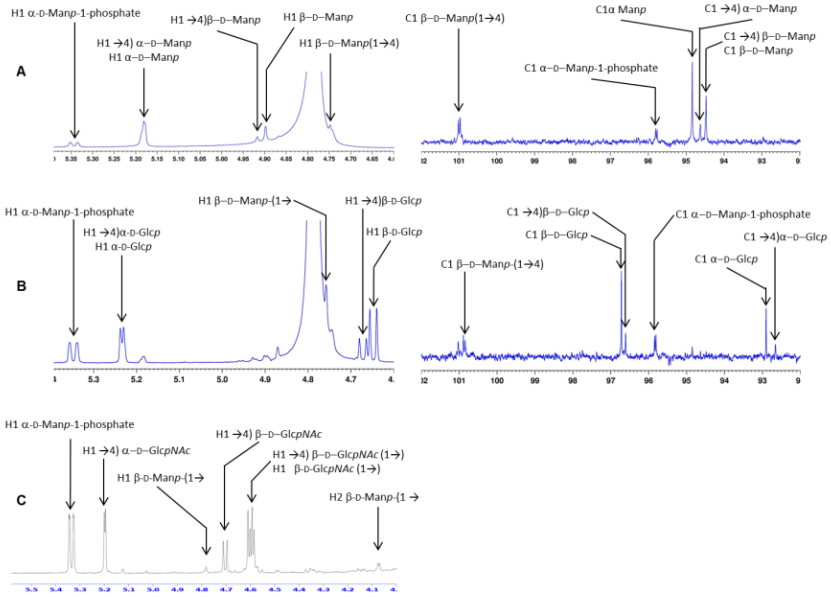
Supplementary figure 1



Supplementary Figure 2:

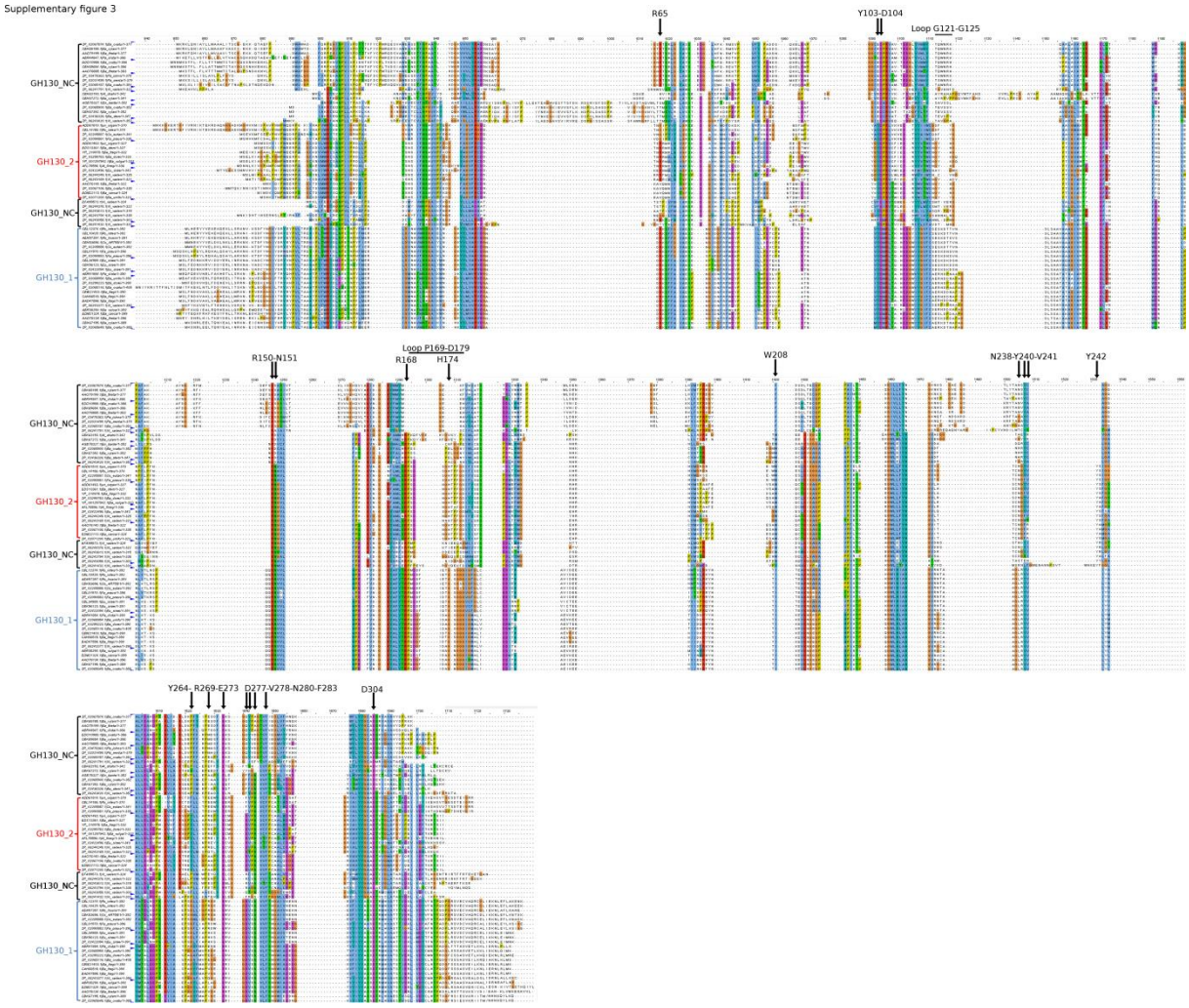
^{1}H (A) and ^{13}C (B) NMR spectra of the products synthesized by Uhgb_MP from 1) 10mM α-D-mannopyranose-1-phosphate and 10mM D-mannose, 2) 10mM α-D-mannopyranose-1-phosphate and 10mM D-glucose, 3) 10mM α-D-mannopyranose-1-phosphate and 10mM *N*,*N*'-diacetyl chitobiose.

Supplementary figure 2



Supplementary Figure 3:

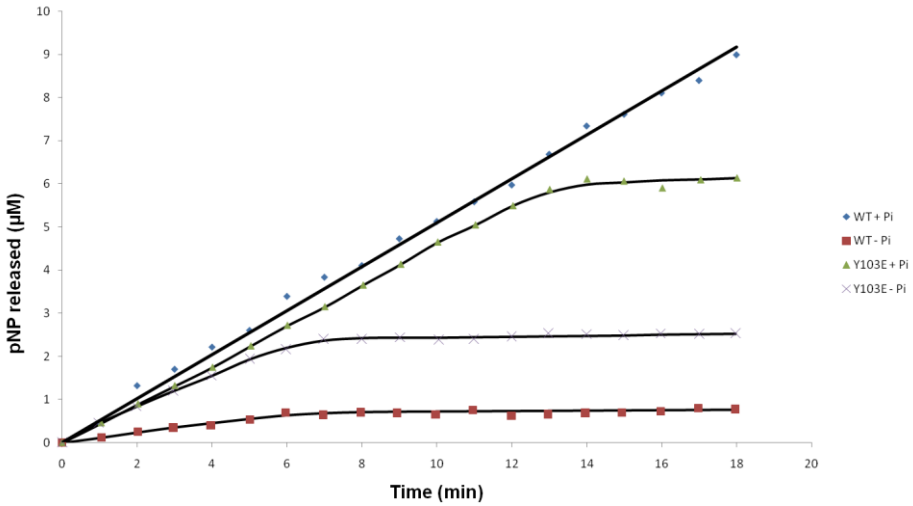
Multiple alignment of 64 GH130 protein sequences, illustrating similitudes and differences of residue conservation between the GH130_1, GH130_2 and GH130_NC sequence clusters.



Supplementary Figure 4:

*p*NP release curves from 1mM *p*NP-β-D-mannopyranoside, for wild-type Uhgb_MP and its Y103E mutant, in the absence or presence of 10mM inorganic phosphate at 37°C in 20mM Tris HCl, pH7.

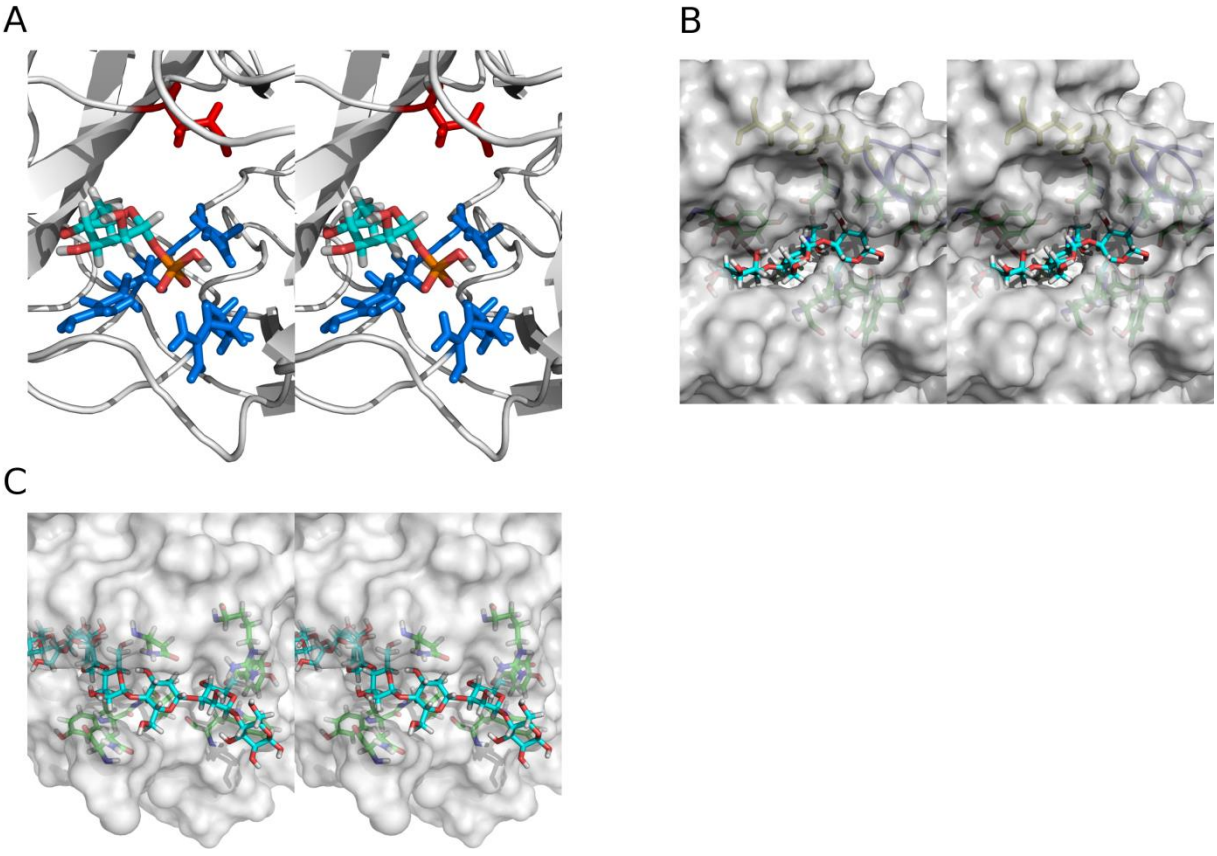
Supplementary figure 4



Supplementary Figure 5:

Stereo views of the 3D model of the docked complexes of Uhgb_MP with α-D-mannopyranose-1-phosphate (A), β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc (B) or β-1,4 linked D-mannoheptaose (C). The most probable catalytic residue D104 is shown in red, while the phosphate stabilizing residues R150, N151 and R168 are shown in blue.

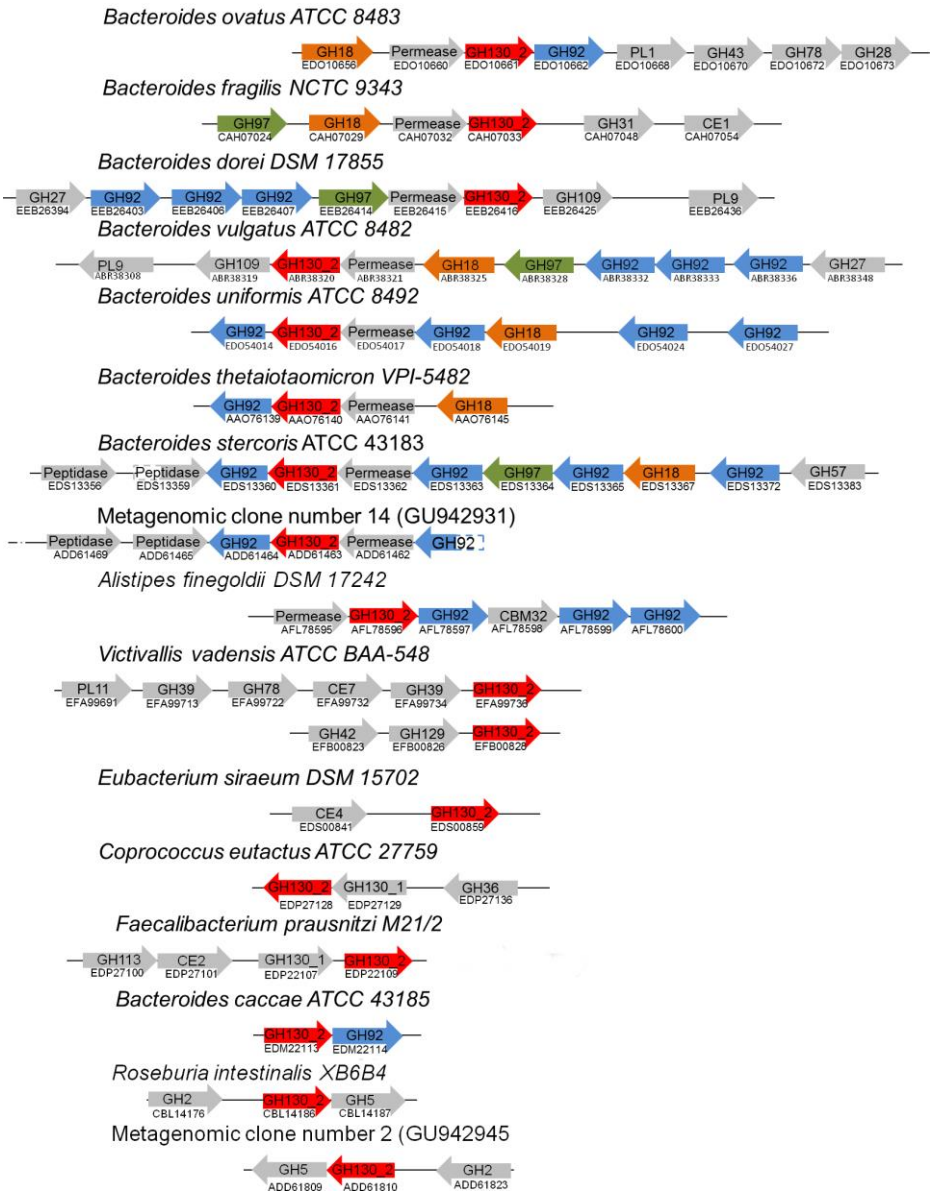
Supplementary figure 5



Supplementary Figure 6:

Comparison of genomic environments of GH130_2 encoding genes found in human gut bacterial genomes, and in two metagenomic DNA inserts of metagenomic clones number 14 (Genbank Accession number GU942931) and number 2 (Genbank Accession number GU942931). This figure only shows the CAZy genes on the same strand as the GH130_2 one, with a single exception which are the peptidase encoding genes shared by the metagenomic DNA encoding Uhgb_MP and the *B. stercoris* ATCC 43183 genome. For the metagenomic DNA encoding Uhgb_MP, the last GH92 gene is located truncated at the end of the DNA insert. However, the last 377 residues are 99% identical to EDS13363. The EDS13359 gene, encodes a 130 residue putative peptidase homologous to ADD61465, which is 171 residue long. These two proteins are 99% identical on the common part of sequence.

Supplementary figure 6



Supplementary Table 1:

Accession numbers and microbial origin of the enzyme sequences classified in the GH130_1 and the GH130_2 subfamilies.

GH130 subfamily	**Accession number**	**Organism name**
GH130_1	AEN74334.1	*Rhodothermus marinus* SG0.5JP17-172
GH130_1	ACY49318.1	*Rhodothermus marinus* DSM 4252
GH130_1	AFN73750.1	*Melioribacter roseus* P3M
GH130_1	AFC29298.1	*Paenibacillus mucilaginosus* 3016
GH130_1	AFH61477.1	*Paenibacillus mucilaginosus* K02
GH130_1	AEI40684.1	*Paenibacillus mucilaginosus* KNP414
GH130_1	ACT00812.1	*Paenibacillus sp.* JDR-2
GH130_1	ADT88764.1	*Vibrio furnissii* NCTC 11218
GH130_1	ADL52883.1	*Clostridium cellulovorans* 743B
GH130_1	AEY67874.1	*Clostridium sp.* BNL1100
GH130_1	ACL77297.1	*Clostridium cellulolyticum* H10 ATCC 35319
GH130_1	AEE53364.1	*Haliscomenobacter hydrossis* DSM 1100
GH130_1	AEI50003.1	*Runella slithyformis* DSM 19594
GH130_1	ADQ19078.1	*Leadbetterella byssophila* DSM 17132
GH130_1	AFD07545.1	*Solitalea canadensis* DSM 3403
GH130_1	ACU06175.1	*Pedobacter heparinus* DSM 2366
GH130_1	ADY53512.1	*Pedobacter saltans* DSM 12145
GH130_1	ACU62827.1	*Chitinophaga pinensis* DSM 2588
GH130_1	AEW00652.1	*Niastella koreensis* GR20-10
GH130_1	AAS19693.1	*Cellvibrio mixtus*
GH130_1	ACE86269.1	*Cellvibrio japonicus Ueda* 107
GH130_1	ABD79771.1	*Saccharophagus degradans* 2-40
GH130_1	AFV00487.1	*Simiduia agarivorans SA1* = DSM 21679
GH130_1	ACR12108.1	*Teredinibacter turnerae* T7901
GH130_1	AEE24224.1	*Glaciecola sp.* 4H-3-7+YE-5
GH130_1	ACY02068.1	*Flammeovirga yaeyamensis* MY04
GH130_1	ABR41854.1	*Parabacteroides distasonis* ATCC 8503
GH130_1	ADY36278.1	*Bacteroides salanitronis* DSM 18170
GH130_1	ADV42257.1	*Bacteroides helcogenes* P 36-108 ATCC 35417
GH130_1	CBW21403.1	*Bacteroides fragilis* 638R
GH130_1	CAH06518.1	*Bacteroides fragilis* NCTC 9343
GH130_1	BAD47598.1	*Bacteroides fragilis* YCH46
GH130_1	AEA20333.1	*Prevotella denticola* F0289
GH130_1	ADE81775.1	*Prevotella ruminicola* 23
GH130_1	ADQ79955.1	*Paludibacter propionicigenes* WB4
GH130_1	ABR38290.1	*Bacteroides vulgatus* ATCC 8482
GH130_1	CBK67195.1	*Bacteriodes xylanisolvens* XB1A

GH130_1	ACA61147.1	uncultured microorganism
GH130_1	ADB80102.1	uncultured microorganism
GH130_1	CAJ19137.1	unidentified microorganism
GH130_1	AAO78134.1	*Bacteroides thetaiotaomicron* VPI-5482
GH130_1	ADZ77504.1	*Sphingobacterium sp.* 21
GH130_1	ABQ07954.1	*Flavobacterium johnsoniae* UW101
GH130_1	ADF50716.1	*Zunongwangia profunda* SM-A87 SMA-87
GH130_1	ADF53791.1	*Zunongwangia profunda* SM-A87 SMA-87
GH130_1	EAR01803.1	*Flavobacteriales bacterium* HTCC 2170
GH130_1	AEJ60368.1	*Spirochaeta thermophila* DSM 6578
GH130_1	ADN01070.1	*Spirochaeta thermophila* DSM 6192
GH130_1	AEE17334.1	*Treponema brennaborense* DSM 12168
GH130_1	CBL12374.1	*Roseburia intestinalis* XB6B4
GH130_1	CBL10429.1	*Roseburia intestinalis* M50/1
GH130_1	AEN97397.1	*Roseburia hominis* A2-183
GH130_1	CBK83696.1	*Coprococcus sp.* ART55/1
GH130_1	CBL01970.1	*Faecalibacterium prausnitzii* SL3/3
GH130_1	ADU26466.1	*Ethanoligenens harbinense* YUAN-3
GH130_1	ADU21379.1	*Ruminococcus albus* 7
GH130_1	AGB19552.1	*Thermoanaerobacterium thermosaccharolyticum* M0795
GH130_1	AFK87330.1	*Thermoanaerobacterium saccharolyticum* JW/SL-YS485
GH130_1	ADL69421.1	*Thermoanaerobacterium thermosaccharolyticum* DSM 571
GH130_1	CBL34989.1	*Eubacterium siraeum* V10Sc8a
GH130_1	CBK96123.1	*Eubacterium siraeum* 70/3
GH130_1	ADZ85051.1	*Clostridium lentocellum* DSM 5427
GH130_1	CCI71614.1	*Paenibacillus polymyxa* M1
GH130_1	ADO59103.1	*Paenibacillus polymyxa* SC2
GH130_1	ADM72412.1	*Paenibacillus polymyxa* E681
GH130_1	AET57921.1	*Paenibacillus terrae* HPL-003
GH130_1	BAM48377.1	*Amphibacillus xylanus* NBRC 15112
GH130_1	ADU28565.1	*Bacillus cellulosilyticus* DSM 2522
GH130_1	ACB75590.1	*Opitutus terrae* PB90-1
GH130_1	ZP_08159838.1	*Ruminococcus albus* 8
GH130_1	ZP_06243377.1	*Victivallis vadensis* ATCC BAA-548
GH130_1	ZP_03299223.1	*Bacteroides dorei* DSM 17855
GH130_1	ZP_02422094.1	*Eubacterium siraeum* DSM 15702
GH130_1	ZP_02205888.1	*Coprococcus eutactus* ATCC 27759
GH130_1	ZP_02090883.1	*Faecalibacterium Prausnitzii* M21/2
GH130_1	ZP_02068954.1	*Bacteroides uniformis* ATCC 8492
GH130_1	ZP_02065648.1	*Bacteroides ovatus* ATCC 8483
GH130_1	ZP_02065116.1	*Bacteroides ovatus* ATCC 8483
GH130_1	ZP_01959662.1	*Bacteroides caccae* ATCC 43185

GH130_2	AEY67872.1	*Clostridium sp.* BNL1100
GH130_2	AFN75330.1	*Melioribacter roseus* P3M
GH130_2	ADO59098.1	*Paenibacillus polymyxa* SC2
GH130_2	CCI71609.1	*Paenibacillus polymyxa* M1
GH130_2	ADD61810.1	uncultured organism
GH130_2	CBL14186.1	*Roseburia intestinalis* XB6B4
GH130_2	AFK85719.1	*Thermoanaerobacterium saccharolyticum* JW/SL-YS485
GH130_2	ADA67643.1	*Thermotoga naphthophila* RKU-10
GH130_2	AAD36300.1	*Thermotoga maritima* MSB8
GH130_2	ACB09929.1	*Thermotoga sp.* RQ2
GH130_2	ABQ47551.1	*Thermotoga petrophila* RKU-1
GH130_2	ACM23522.1	*Thermotoga neapolitana* DSM 4359
GH130_2	ABV33993.1	*Thermotoga lettingae* TMO
GH130_2	AGB28392.1	*Prevotella dentalis* DSM 3688
GH130_2	AFL78596.1	*Alistipes finegoldii* DSM 17242
GH130_2	ADD61463.1	uncultured organism
GH130_2	EDS13361.1	*Bacteroides stercoris* ATCC 43183
GH130_2	ABX42090.1	*Clostridium phytofermentans* ISDg
GH130_2	AFG35891.1	*Fervidobacterium pennivorans* DSM 9078
GH130_2	ACJ75237.1	*Thermosipho africanus* TCF52B
GH130_2	ADL43426.1	*Caldicellulosiruptor obsidiansis* OB47
GH130_2	AEM72946.1	*Caldicellulosiruptor lactoaceticus* 6A
GH130_2	ACM61623.1	*Anaerocellum thermophilum* DSM 6725
GH130_2	ABR30874.1	*Thermosipho melanesiensis* BI429
GH130_2	AFK07300.1	*Mesotoga prima* MesG1.Ag.4.2
GH130_2	AFK07371.1	*Mesotoga prima* MesG1.Ag.4.2
GH130_2	AFL98126.1	*Ornithobacterium rhinotracheale* DSM 15997
GH130_2	AAO76140.1	*Bacteroides thetaiotaomicron* VPI-5482
GH130_2	ACB75595.1	*Opitutus terrae* PB90-1
GH130_2	ZP_08158270.1	*Ruminococcus albus* 8
GH130_2	ZP_06244248.1	*Victivallis vadensis* ATCC BAA-548
GH130_2	ZP_06243169.1	*Victivallis vadensis* ATCC BAA-548
GH130_2	ZP_03299783.1	*Bacteroides dorei* DSM 17855
GH130_2	ZP_02422496.1	*Eubacterium siraeum* DSM 15702
GH130_2	ZP_02205887.1	*Coprococcus eutactus* ATCC 27759
GH130_2	ZP_02090881.1	*Faecalibacterium prausnitzii* M21/2
GH130_2	ZP_02071200.1	*Bacteroides uniformis* ATCC 8492
GH130_2	ZP_02067106.1	*Bacteroides ovatus* ATCC 8483
GH130_2	ZP_01958898	*Bacteroides caccae* ATCC 43185
GH130_2	YP_210978.1	*Bacteroides fragilis* NCTC 9343
GH130_2	YP_004103295.1	*Ruminococcus albus* 7
GH130_2	YP_001297942.1	*Bacteroides vulgatus* ATCC 8482

Supplementary Table 2:

Prevalence analysis of 23 GH130_1, 17 GH130_2 and 25 GH130_NC genes in the fecal metagenome of 27 IBD suffering patients and 135 other healthy European subjects from the MetaHit cohort, illustrated by the mean number of TBLASTN hits per number of individuals (E-value = 0, Identity = 90%).

	GH130_1	**GH130_2**	**GH130_NC**
Total number of GH130 hits	193	124	120
Number of subjects	27	27	27
Number of GH130 hits/100 Mb	10.1	6.5	6.3
Number of GH130 hits/subject	7.2	4.6	4.4
Total number of GH130 hits	826	474	372
Number of subjects	135	135	135
Number of GH130 hits/100 Mb	9.0	5.1	4.0
Number of GH130 hits/subject	6.1	3.5	2.8

Second article

Our previous work allowed us to characterize the function of Uhgb_MP, and identify its role in mannoside foraging by gut bacteria. By using site-directed mutagenesis, we also identified the most important residues for catalysis, in particular the proton donor.

However, even if a Uhgb_MP 3D model has been generated and described in our first paper, structural data were still lacking to explain the particular Uhgb_MP specificity. Indeed, we previously analyzed the characterized representatives of the GH130 family on the basis of their specific catalytic activities, primary structures and genomic context, which allowed us to create subfamilies and to link them to their functional role *in vivo*. Nevertheless, only one member of the GH130 family (*Bf*MGP) was both structurally and functionally characterized until January 2015, limiting the understanding of the link existing between structure and substrate specificity for GH130 enzymes. The following article, which presents the analysis of the ternary and quaternary Uhgb_MP structures, was published in Acta Crystallographia section D in 2015. The crystallization trials, X-ray data acquisition and crystalline structure analysis were carried out thanks to Samuel Tranier and Lionel Mourey, who hosted me at the IPBS and collected, with and without me, X-ray data at the ESRF and at the ALBA synchrotron. I was also lucky to collaborate with Pierre Roblin and Gianluca Cioci for SAXS analysis at the SOLEIL synchrotron radiation facility.

BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Structural bases for N-glycan processing by mannoside phosphorylase

Simon Ladevèze,[a,b,c] Gianluca Cioci,[a,b,c] Pierre Roblin,[d] Lionel Mourey,[e,f] Samuel Tranier[e,f]* and Gabrielle Potocki-Véronèse[a,b,c]*

[a]Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, 31077 Toulouse, France, [b]CNRS, UMR5504, 31400 Toulouse, France, [c]INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, 31400 Toulouse, France, [d]Synchrotron SOLEIL, L'Orme des Merisiers, BP 48, Saint Aubin, 91192 Gif-sur-Yvette CEDEX, France, [e]Institut de Pharmacologie et de Biologie Structurale (IPBS), Centre National de la Recherche Scientifique (CNRS), 205 Route de Narbonne, BP 64182, 31077 Toulouse, France, and [f]Université de Toulouse, Université Paul Sabatier, IPBS, 31077 Toulouse, France. *Correspondence e-mail: samuel.tranier@ipbs.fr, veronese@insa-toulouse.fr

Received 23 January 2015
Accepted 1 April 2015

Edited by Z. S. Derewenda, University of Virginia, USA

Keywords: GH130 enzymes; N-glycans; glycoside phosphorylases; human gut microbiota.

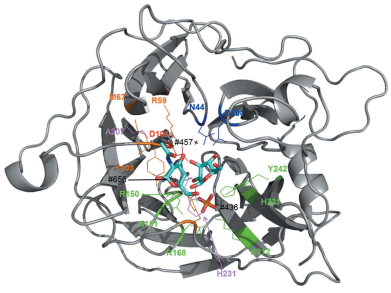
PDB references: Uhgb_MP, apo, 4udi; complex with mannose, 4udj; complex with *N*-acetylglucosamine, 4udg; complex with mannose and *N*-acetylglucosamine, 4udk

Supporting information: this article has supporting information at journals.iucr.org/d

The first crystal structure of Uhgb_MP, a β-1,4-mannopyranosyl-chitobiose phosphorylase belonging to the GH130 family which is involved in N-glycan degradation by human gut bacteria, was solved at 1.85 Å resolution in the apo form and in complex with mannose and *N*-acetylglucosamine. SAXS and crystal structure analysis revealed a hexameric structure, a specific feature of GH130 enzymes among other glycoside phosphorylases. Mapping of the −1 and +1 subsites in the presence of phosphate confirmed the conserved Asp104 as the general acid/base catalytic residue, which is in agreement with a single-step reaction mechanism involving Man O_3 assistance for proton transfer. Analysis of this structure, the first to be solved for a member of the GH130_2 subfamily, revealed Met67, Phe203 and the Gly121–Pro125 loop as the main determinants of the specificity of Uhgb_MP and its homologues towards the N-glycan core oligosaccharides and mannan, and the molecular bases of the key role played by GH130 enzymes in the catabolism of dietary fibre and host glycans.

1. Introduction

N-linked glycans are present in many living organisms, notably eukaryotes, and play a key role in major processes, including cell signalling and recognition, protein stability and activity tuning (Varki *et al.*, 2009). These oligosaccharides, which are covalently linked to the asparagine residues of glycoproteins, display relatively restricted structural diversity (Lehle *et al.*, 2006; Larkin & Imperiali, 2011). In eukaryotes, N-glycans share a common core structure composed of the β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc (Man-GlcNAc$_2$) trisaccharide, carrying decorations on the nonreducing β-linked mannosyl residue to form more complex structures (Aebi *et al.*, 2010; Nagae & Yamaguchi, 2012). Although the whole pathways of human N-glycan synthesis and maturation have been well described, little is known about their degradation, especially by bacteria or fungi, despite the fact that degradation is a key factor in microbe–host interactions (Suzuki & Harada, 2014). Until 2013, only glycoside hydrolases (GHs) had been shown to be implicated in N-glycan breakdown in the CAZy database (http://www.cazy.org/; Lombard *et al.*, 2014). Huge efforts have been made in recent years to understand exactly how this hydrolytic process takes place, particularly among gut inhabitants, as the alteration of host glycans by microbes is thought to be related to intestinal disorders, including Crohn's disease and other inflammatory



OPEN ACCESS

bowel diseases (IBDs; Martens *et al.*, 2008; Sheng *et al.*, 2012). The aim of these studies was to identify a broad consortium of enzymes acting on different parts of the N-glycan structure, such as exo-mannosidases and endo-mannosidases or endo-*N*-acetyl-β-D-glucosaminidases produced by gut commensals and pathogens (Roberts *et al.*, 2000; Burnaugh *et al.*, 2008; Renzi *et al.*, 2011), particularly by *Bacteroides* species (Martens *et al.*, 2009; Zhu *et al.*, 2010).

In 2013, the first evidence for N-glycan breakdown by phosphorolysis was published, which involved gut bacterial mannoside phosphorylases belonging to glycoside hydrolase family 130 (GH130; Nihira *et al.*, 2013; Ladevèze *et al.*, 2013). Only two enzymes, namely the mannoside phosphorylase (EC 2.4.1.–) Uhgb_MP, an enzyme produced by an uncultivated *Bacteroides* bacterium, and *Bt*1033, produced by *B. thetaiotaomicron* VPI-5482, are known to catalyze the conversion of β-D-Man*p*-1,4-β-D-Glc*p*NAc (Man-GlcNAc) or β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc (Man-GlcNAc$_2$) and inorganic phosphate into α-D-mannopyranose-1-phosphate and GlcNAc or GlcNAc$_2$, respectively (Nihira *et al.*, 2013; Ladevèze *et al.*, 2013). CAZy subfamilies are subgroups found within a family that share a more recent ancestor and that are usually more uniform in molecular function, reflecting a high degree of conservation in their active site (Aspeborg *et al.*, 2012). In the GH130 family, at least two enzyme subfamilies have been identified (Ladevèze *et al.*, 2013). Subfamily GH130_1 gathers enzymes that are highly specific for β-D-Man*p*-1,4-D-Glc, while GH130_2 contains enzymes that are much more flexible towards mannosides. Uhgb_MP and *Bt*1033 are classified in the GH130_2 subfamily, together with 40 other GH130 sequences, among which 15 originate from gut bacterial genomes. Integration of metagenomic and genomic data on the scale of the entire human gut microbiota revealed that GH130_2 enzymes, especially Uhgb_MP and *Bt*1033, probably play a critical role in alteration of the intestinal barrier, as their encoding genes are particularly prevalent in the human gut microbiome of patients suffering from IBDs (Ladevèze *et al.*, 2013). Based on genomic context analysis and on the *in silico* detection of signal peptides, the physiological role of Uhgb_MP and *Bt*1033 would be the intracellular phosphorolysis of β-D-Man*p*-1,4-D-GlcNAc, which can be internalized in the cell after extracellular hydrolysis of N-glycans by glycoside hydrolases belonging to the GH18, GH92 and possibly also the GH97 families (Ladevèze *et al.*, 2013). In addition to Uhgb_MP and *Bt*1033, subfamily GH130_2 contains only one other biochemically characterized enzyme, the *Ra*MP2 enzyme from the ruminal bacterium *Ruminococcus albus* 7. It has been suggested that this enzyme is involved in mannan catabolism in the bovine rumen, as it catalyzes the phosphorolysis of β-1,4-manno-oligosaccharides (Kawahara *et al.*, 2012). *In vitro*, these three enzymes present a relaxed substrate specificity compared with all other known mannoside phosphorylases. This property makes them extremely interesting biocatalytic tools for the synthesis of diverse manno-oligosaccharides by reverse phosphorolysis. In particular, Uhgb_MP is extremely efficient at producing N-glycan core oligosaccharides such as β-D-Man*p*-1,4-D-GlcNAc and β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc, the current commercial price of which exceeds $10 000 per milligram (Ladevèze *et al.*, 2014). Moreover, it is the only known phosphorylase to act on mannans and long manno-oligosaccharides. Uhgb_MP-based β-mannoside synthesis processes are highly attractive, thanks to its flexible specificity and because it is the only known enzyme able to produce such high added value compounds from a hemicellulose constituent as a substrate. Indeed, a one-pot reaction would allow Uhgb_MP to produce β-D-Man*p*-1,4-β-D-Glc*p*NAc directly from *N*-acetylglucosamine and mannan following two reaction steps: a first step of mannan phosphorolysis releasing α-D-Man-1-phosphate, and a second step of reverse phosphorolysis converting α-D-Man-1-phosphate and *N*-acetylglucosamine into β-D-Man*p*-1,4-β-D-Glc*p*NAc.

Currently, six GH130 enzyme structures are available in the RCSB Protein Data Bank, sharing a common five-bladed β-propeller fold. The crystal structure of *Bf*MGP, a *B. fragilis* NCTC 9343 enzyme classified with 78 other sequences into the GH130_1 subfamily, was recently solved in complex with the genuine substrates 4-*O*-β-D-mannosyl-D-glucose and phosphate and the product α-D-mannose-1-phosphate (Nakae *et al.*, 2013; PDB entries 3wat, 3was, 3wau and 4kmi). This enzyme, which is involved in the final steps of mannan catabolism in the human gut, exhibits a very narrow specificity towards β-D-Man*p*-1,4-D-Glc (Senoura *et al.*, 2011), like the other characterized members of the GH130_1 subfamily (*Ra*MP1 from the ruminal bacterium *R. albus* 7 and the *Rm*MGP protein from the marine bacterium *Rhodothermus marinus* DSM4252; Jaito *et al.*, 2014). This pioneering study on *Bf*MGP highlighted a probably unique reaction mechanism among known disaccharide phosphorylases, as the invariant residue Asp131, which is assumed to be the general acid/base, was not found close to the glycosidic O atom, which should be protonated in the catalytic reaction.

The five other three-dimensional structures of GH130 enzymes available to date are the apo forms of (i) four proteins belonging to the GH130_NC cluster, which gathers enzymes that are not classified into the GH130_1 and GH130_2 subfamilies, namely BACOVA_03624 and BACOVA_02161 from *B. ovatus* ATCC 8483 (PDB entries 3qc2 and 4onz; Joint Center for Structural Genomics, unpublished work), BDI_3141 from *Parabacteroides distasonis* ATCC 8503 (PDB entry 3taw; Joint Center for Structural Genomics, unpublished work) and BT_4094 from *B. thetaiotaomicron* VPI-5482 (PDB entry 3r67; Joint Center for Structural Genomics, unpublished work), and (ii) *Tm*1225 from *Thermotoga maritima* MSB8 (PDB entry 1vkd; Joint Center for Structural Genomics, unpublished work), which belongs to the GH130_2 subfamily. No function has yet been attributed to these five proteins, thus limiting our understanding of their structure–specificity relationships. Until now, nothing has been established regarding the molecular bases of the relaxed specificity of the enzymes classified into the GH130_2 family. In addition, no structural feature has been identified to explain the efficiency of Uhgb_MP and *Bt*1033 in binding and breaking down N-glycan core oligosaccharides.

Here, we present the first crystal structure of an N-glycan phosphorolytic enzyme, Uhgb_MP, solved by X-ray crystallography in complex with inorganic phosphate, mannose and *N*-acetylglucosamine. This study made it possible to review the previously published three-dimensional model of Uhgb_MP and provides key information to understand its catalytic mechanism. Comparative analysis of this new tertiary and quaternary structure with other GH130 structures allowed us to identify structural features specific to GH130 subfamilies that could explain their functional specificities and hence their key role in mannose foraging in the human gut. This work therefore paves the way for enzyme optimization by rational engineering to fit industrial needs as well as for the design of specific inhibitors to investigate, and potentially to control, interactions between host and gut microbes.

2. Materials and methods

2.1. Recombinant production of Uhgb_MP and enzyme purification

Uhgb_MP was produced in *Escherichia coli* BL21-AI cells (Invitrogen) after its encoding gene had been cloned into the pET-28a vector, yielding an N-terminally hexahistidine-tagged protein (detailed procedures are provided as Supporting Information). After purification by His-tag affinity chromatography and gel filtration, the enzyme was stored in 20 m*M* potassium phosphate pH 7.0, 150 m*M* NaCl (see Supporting Information).

2.2. Activity measurements

Phosphorolytic activity was assessed using two substrates, *p*NP-β-D-mannopyranose and β-D-mannopyranosyl-1,4-D-mannose. All reactions were carried out with 0.1 mg ml^{-1} purified enzyme at 37°C (the optimal temperature for Uhgb_MP) in 20 m*M* Tris–HCl pH 7.0 (the optimal pH for Uhgb_MP). For measurement of the activity in the presence of 10 m*M* inorganic phosphate and 1 m*M* *p*NP-β-D-mannopyranose, the *p*NP release rate was monitored at 405 nm using a Cary-100 UV–visible spectrophotometer (Agilent Technologies). The release rate of α-D-mannopyranose-1-phosphate from 10 m*M* inorganic phosphate and 10 m*M* β-D-mannobiose (Megazyme, Ireland) was measured by quantification of α-D-mannopyranose-1-phosphate using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) as described previously (Ladevèze *et al.*, 2013).

2.3. Size-exclusion chromatography multi-angle laser light scattering (SEC-MALLS) experiments

A 30 µl sample of gel-filtered Uhgb_MP at a concentration of 6 mg ml^{-1} in 20 m*M* potassium phosphate pH 7.0, 150 m*M* NaCl was loaded onto a Superdex 200 HR 10/300 column (GE Healthcare, Massy, France) using an Agilent 1260 Infinity LC chromatographic system (Agilent Technology) coupled to a multi-angle laser light scattering (MALLS) detection system. The protein was centrifuged for 5 min at 4°C at 10 000*g* before the sample was loaded. The column was equilibrated with a 0.1 µm filtered buffer composed of 20 m*M* potassium phosphate pH 7.0, 150 m*M* NaCl. Separation was performed at a flow rate of 0.4 ml min^{-1} at 15°C. Data were collected using a DAWN HELEOS 8+ (eight-angle) light-scattering detector and an Optilab T-rEX refractive-index detector (Wyatt Technology, Toulouse, France). The results were analyzed using the *ASTRA* v.6.0.2.9 software (Wyatt Technology).

2.4. Protein crystallization

Purified Uhgb_MP protein was concentrated using polyethersulfone Vivaspin concentrators (Vivascience, Sartorius, Göttingen, Germany). The concentration was determined by measuring the $A_{280\,\mathrm{nm}}$ using a NanoDrop instrument (Wilmington, Delaware, USA). All crystallization experiments were carried out at 12°C by the sitting-drop vapour-diffusion method using MRC 96-well microplates (Molecular Dimensions, Newmarket, England) and a Nanodrop ExtY crystallization instrument (Innovadyne Technologies, Santa Rosa, USA) to prepare 400 nl droplets. The best Uhgb_MP crystals were obtained within a week with a 1:1(*v*:*v*) ratio of protein (9–12 mg ml^{-1} in 20 m*M* potassium phosphate pH 7.0, 150 m*M* NaCl supplemented with 5 m*M* mannose and/or 5 m*M* *N*-acetylglucosamine for the co-crystallization assays) to precipitant solution [17.5–20%(*w*/*v*) polyethylene glycol 3350, 0.175–0.2 *M* ammonium chloride]. Uhgb_MP crystals grew to dimensions of 0.2 × 0.08 × 0.02 mm in a week. They diffracted to a maximum resolution of 1.80 Å, while those obtained by co-crystallization with mannose, *N*-acetylglucosamine or both diffracted to maximum resolutions of 1.94, 1.60 and 1.76 Å, respectively.

2.5. Data collection and determination of the structure

X-ray experiments were carried out at 100 K. Crystals of Uhgb_MP were soaked for a few seconds in reservoir solution supplemented with 15%(*v*/*v*) glycerol (apo) or 15%(*v*/*v*) PEG 300 (complexes) prior to flash-cooling. Apo Uhgb_MP diffraction data sets were collected on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, while those for the complexes were collected on the XALOC beamline at the ALBA Synchrotron, Cerdanyola del Vallès, Spain (Juanhuix *et al.*, 2014). The diffraction intensities were integrated and scaled using *XDS* (Kabsch, 2010) and 5% of the scaled amplitudes were randomly selected and excluded from the refinement procedure. All crystals belonged to the orthorhombic space group $P2_12_12_1$, with six molecules per asymmetric unit, giving Matthews coefficients of 2.22 and 2.11 Å^3 Da^{-1} and solvent contents of 44 and 42% for the apo forms and the three complexes, respectively. The structures were solved by the molecular-replacement method using *Phaser* (McCoy *et al.*, 2007) from the *CCP*4 software suite (Potterton *et al.*, 2003) and chain *A* of the crystal structure of *Tm*1225 from *T. maritima* MSB8 (PDB entry 1vkd) as a search model for the apo form. The final translation-function *Z*-score was 42.8 and the *R* and R_{free} values of the refined structure were 0.155 and 0.190, respectively. Once solved, the apo

Table 1
Data-collection and refinement statistics for Uhgb_MP.

Values in parentheses are for the outer resolution shell.

	Native (P_i)	Mannose + P_i	*N*-Acetylglucosamine + P_i	Mannose + *N*-acetylglucosamine + P_i
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 84.1, b = 141.2, c = 176.2, $\alpha = \beta = \gamma$ = 90	a = 83.9, b = 140.8, c = 168.7, $\alpha = \beta = \gamma$ = 90	a = 83.8, b = 140.9, c = 168.6, $\alpha = \beta = \gamma$ = 90	a = 83.7, b = 140.9, c = 168.8, $\alpha = \beta = \gamma$ = 90
No. of molecules in asymmetric unit	6	6	6	6
Matthews coefficient (Å^3 Da^{-1})	2.22	2.11	2.11	2.11
Solvent content (%)	44.66	41.82	41.74	41.75
Wavelength (Å)	0.96863	0.97949	0.97949	0.97949
Resolution range (Å)	48.16–1.80 (1.91–1.80)	75.11–1.94 (1.98–1.94)	46.68–1.60 (1.64–1.60)	45.43–1.76 (1.80–1.76)
No. of unique reflections	190253 (27574)	148300 (21410)	261650 (41727)	197652 (31069)
No. of observed reflections	927534 (101770)	1217594 (168817)	2453350 (392431)	1818257 (276166)
Completeness (%)	98.00 (88.70)	99.66 (98.85)	99.84 (98.42)	99.64 (95.24)
Multiplicity	4.88 (3.69)	8.20 (7.90)	9.37 (7.00)	9.19 (8.88)
$\langle I/\sigma(I)\rangle$	10.46 (1.63)	9.30 (3.40)	14.49 (2.07)	13.70 (2.54)
R_{merge} (%)	9.1 (70.0)	15.5 (58.9)	9.3 (103.5)	11.9 (86.5)
Refinement				
R_{work}/R_{free}	0.157/0.190	0.152/0.193	0.155/0.183	0.158/0.192
Root-mean-square deviations				
Bond lengths (Å)	0.0188	0.0181	0.0191	0.0183
Bond angles (°)	1.9218	1.8528	1.8905	1.9235
Ramachandran plot				
Favoured (%)	91.6	91.3	91.3	91.1
Allowed (%)	8.1	7.9	8.5	8.6
B factors (Å^2)				
Wilson *B*	24	18	22	22
Mean	35	20	22	21
Main chain	33	18	20	19
Side chain	37	21	24	23
Ligand/water	28/37	23/26	26/30	27/26
PDB code	4udi	4udj	4udg	4udk

structure was then used to solve the protein–ligand structures. The structures of Uhgb_MP in complex with mannose, with *N*-acetylglucosamine and with mannose and *N*-acetylglucosamine were refined to final R/R_{free} values of 0.150/0.193, 0.154/0.183 and 0.158/0.192, respectively, using *REFMAC*5 (Murshudov *et al.*, 2011). Models were built manually in α_A-weighted electron-density maps using *Coot* (Emsley & Cowtan, 2004). Water molecules were manually checked after automatic assignment and ligand molecules were manually fitted in residual maps. Refinement statistics are listed in Table 1.

2.6. SAXS measurements

Small-angle X-ray scattering (SAXS) experiments were performed on the SWING beamline at the SOLEIL synchrotron, Gif-sur-Yvette, France. The wavelength was set to 1.033 Å. A 17 × 17 cm Aviex CCD detector was positioned 1800 mm from the sample, with the direct beam off-centred. The resulting exploitable q-range was 0.006–0.6 Å^{-1}, where $q = 4\pi\sin\theta/\lambda$, considering 2θ as the scattering angle. The samples were circulated in a thermostated quartz capillary with a diameter of 1.5 mm and 10 µm wall thickness positioned inside a vacuum chamber. A 80 µl volume of sample was injected onto a size-exclusion column (Bio SEC3 300, Agilent) equilibrated in phosphate-based buffer (20 m*M* potassium phosphate pH 7.0, 150 m*M* NaCl) or Tris-based buffer (20 m*M* Tris–HCl pH 7.0, 300 m*M* NaCl supplemented with 10% glycerol) using an Agilent high-performance liquid-chromatography (HPLC) system and eluted directly into the SAXS capillary cell at a flow rate of 200 µl min^{-1} at a temperature of 10°C. Samples were separated from the pushing liquid (water) by two air volumes of 6 µl each, as described previously (David & Pérez, 2009). SAXS data were collected online throughout the elution time and a total of 149 frames, each lasting 2 s, were recorded separated by a dead time of 0.5 s between frames. The transmitted intensity was continuously measured with an accuracy of 0.1% using a diode embedded in the beam stop. For each sample, the stability of the associated radius of gyration and the global curve shape in the frames corresponding to the main elution peak were checked, and the resulting selection of curves were averaged as described previously (David & Pérez, 2009). The recorded curves were normalized to the transmitted intensity and subsequently averaged using *Foxtrot*, a dedicated in-house application. The same protocol was applied to buffer scattering. R_g values were determined by a Guinier fit of the one-dimensional curves using the *ATSAS* package (Petoukhov *et al.*, 2007). The $P(r)$ function was calculated using the *GNOM* program and the corresponding *ab initio* envelopes were calculated using the *GASBOR* program. Rigid-body SAXS modelling was performed using the *CORAL* program.

3. Results and discussion

3.1. Conformational stability optimization of Uhgb_MP

Previous work on Uhgb_MP allowed the production and purification of a recombinant form of the protein in amounts suitable for crystallization (Ladevèze *et al.*, 2013). However, owing to enzyme instability, an optimized production system was set up by subcloning the open reading frame of Uhgb_MP into the pET-28a vector (Supporting Information §S1). Subcloning into pET-28a made it possible to produce a recombinant protein with a thrombin-cleavable N-terminal hexahistidine tag and a five-residue shortened linker between the tag and the N-terminal extremity of the native enzyme. After protein purification and processing in the same buffer as previously described (Ladevèze *et al.*, 2013), the activity on *p*NP-*β*-D-mannopyranose was increased by 73% to 10.9 × 10^{-3} µmol min^{-1} mg^{-1}, indicating that the 16-amino-acid linker used in the initial construct negatively impacted on the Uhgb_MP activity. To avoid the use of Tween 80, which is not suitable for crystallization assays, we then screened for an optimized buffer composition by differential scanning fluorimetry (DSF; Supporting Information §S1). In-house-prepared 96 deep-well screens adapted from Ericsson *et al.* (2006) were used to assess the effect of buffer nature, pH and NaCl concentration on protein thermal stability. The denaturation curves revealed two fusion temperatures, T_{m1} = 65.7°C and T_{m2} = 70.8°C, in the initial Tris–HCl buffer, indicating that Uhgb_MP may adopt different oligomeric states in solution. The DSF results showed unambiguously that phosphate-based buffers (sodium and potassium phosphate) largely stabilize the Uhgb_MP structure at all of the assayed pH values (5.0, 5.5, 6.0, 6.5 and 7.0) and NaCl concentrations (136, 159, 287 and 439 m*M*). The best result was observed using 100 m*M* potassium phosphate pH 6.0 with 136 m*M* NaCl, leading to an increase in T_{m1} and T_{m2} of 6.82 ± 0.07 and 8.47 ± 0.07°C, respectively. We therefore chose to use potassium phosphate buffer supplemented with 150 m*M* NaCl to purify and store the protein produced using pET-28a::Uhgb_MP. In addition, the pH was set to 7.0 to allow sufficient separation efficiency of the protein in the affinity-purification step (T_{m1} and T_{m2} increased by 4.80 ± 0.30 and 6.45 ± 0.30°C, respectively, at this pH value). Under these optimal conditions, the protein production yield reached 90 mg pure protein per litre of culture. Finally, the *β*-1,4-D-mannobiose phosphorolytic activity of the enzyme stored in these conditions was increased tenfold compared with that of enzyme previously expressed in pDEST17 and purified in Tris buffer (Ladevèze *et al.*, 2013).

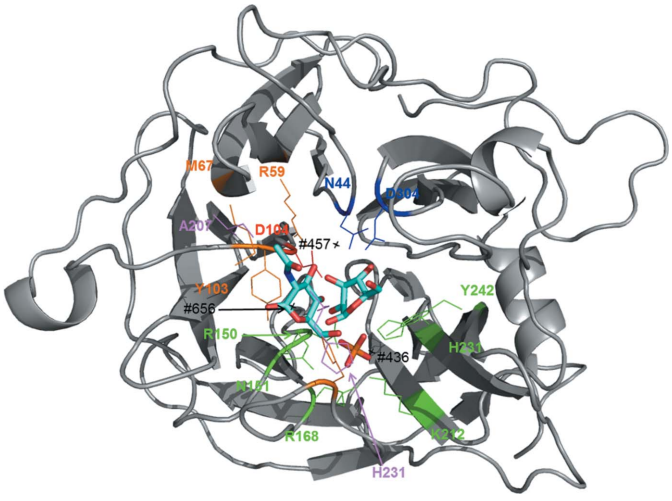


Figure 1
Monomeric Uhgb_MP fold, substrates and interacting residues. Like other GH130 enzymes, the Uhgb_MP monomer has a five-bladed *β*-propeller fold with a central catalytic furrow. The P_i, mannose and *N*-acetylglucosamine molecules present at the catalytic site are shown as sticks, while interacting residues are shown as lines. The catalytic Asp104 is shown in red, P_i-interacting residues in green, mannose-interacting residues in blue and *N*-acetylglucosamine-interacting residues in orange. The Asn44 and Asp104 side chains are shown in the *B* conformation, *i.e.* the conformation that is catalytically active. Water molecules 436, 457 and 656, which mediate interactions between the *N*-acetylglucosamine and residues Lys212 and Tyr242, Asp304 and His235, respectively, are shown as black crosses. Protein–P_i interactions: NH_2 of Arg150 is contacting P_i O^4 and O^2, while NH_2 of Asn151 contacts P_i O^4. ωNH_2 of the Arg168 side chain binds to P_i O^1 and its $\omega'NH_2$ is interacting with P_i O^2. The side-chain amine of Lys212 binds to P_i O^2 and His231 $N^{\varepsilon 2}$ is at a hydrogen-bond distance from P_i O^1. Finally, the hydroxyl of the Tyr242 side chain is in contact with P_i O^3.

3.2. Crystallographic structure of Uhgb_MP subunits

The overall structure of Uhgb_MP was determined by molecular replacement using the structure of *Tm*1225 from *T. maritima* MSB8 (PDB entry 1vkd), which shares 60% identity with Uhgb_MP, as a model. The crystal structure of

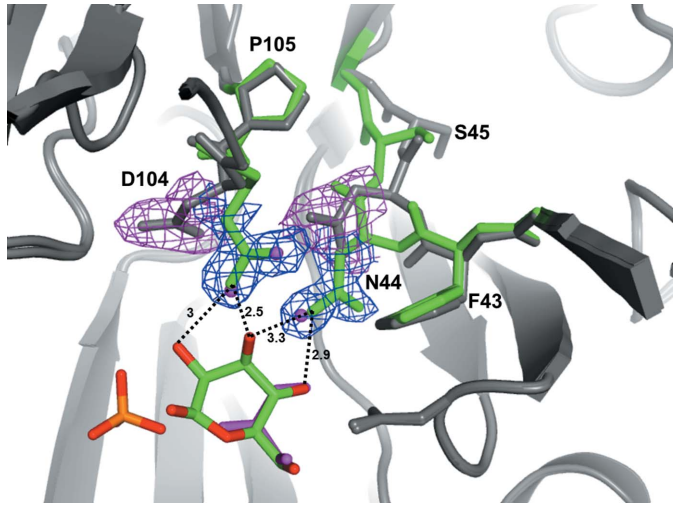


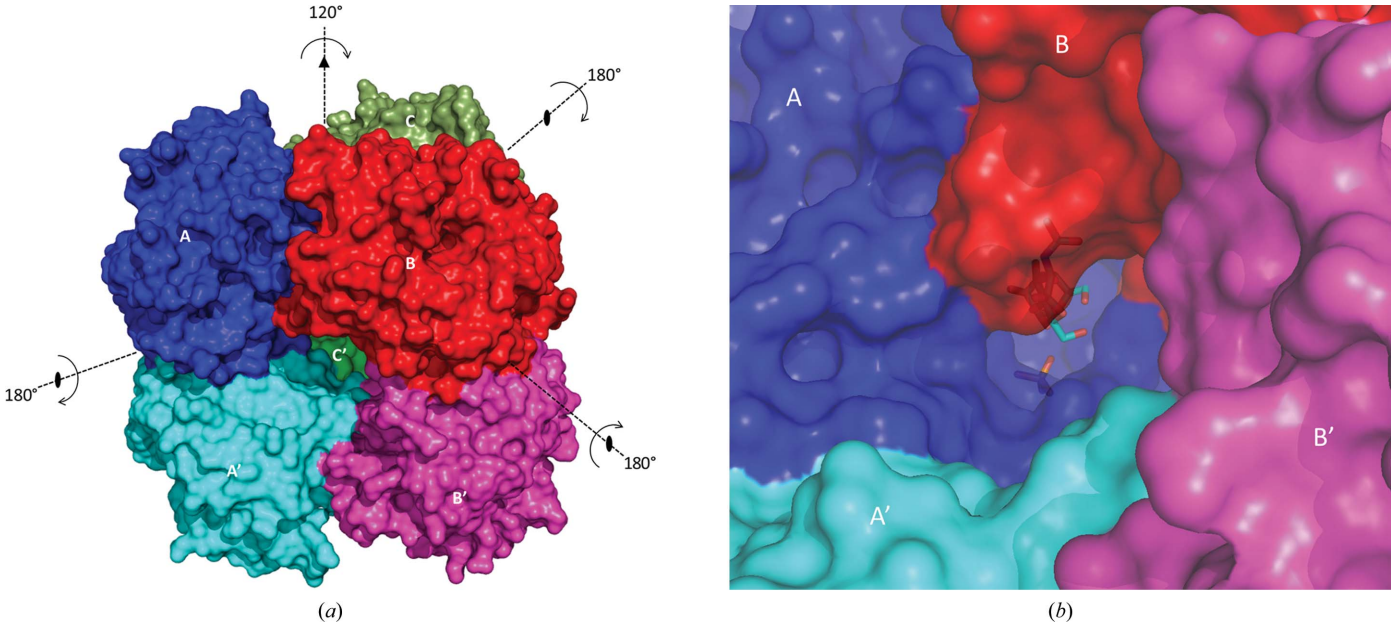
Figure 2
Alternative conformations of Asn44, Ser45 and the catalytic Asp104 upon mannose binding in the −1 subsite. Superposition of the apo structure of Uhgb_MP (PDB entry 4udi) and Uhgb_MP complexed with mannose (PDB entry 4udj), illustrating the movement of the catalytic residues when mannose is bound at the −1 subsite. The backbone of the apo form is shown in grey (*A* conformation), while the backbone of the complexed, catalytically active form (*B* conformation) is in green. Water, phosphate and glycerol molecules in the apo form are shown. $2F_o - F_c$ electron-density maps are shown (contoured at 1.0σ) for the catalytic residues in the apo and mannose-bound forms. Interatomic distances are labelled in Å.

Uhgb_MP revealed a homohexameric organization. The homohexamer consisted of a trimer of dimers with *D*3 symmetry, with six molecules per asymmetric unit. The apo structure was refined to 1.80 Å resolution, while the complexes with mannose, with *N*-acetylglucosamine and with the two combined were refined to 1.94, 1.60 and 1.76 Å, respectively (Table 1). The electron-density maps did not enable construction of the N-terminal extremity of the polypeptide chains. The N-terminal hexahistidine tag and the following 6–8 first residues have thus been omitted from the final model. The overall fold of each Uhgb_MP protomer consists of a five-bladed β-propeller (Fig. 1). The catalytic centre is located in the central cleft as previously hypothesized (Ladevèze *et al.*, 2013), as a phosphate ion (P_i) and mannose and *N*-acetylglucosamine were observed in the central furrow. P_i is deeply buried in the catalytic site, strongly stabilized by hydrogen bonds and ionic interactions with the surrounding residues (Fig. 1). Compared with other glycoside phosphorylases, P_i appeared to be quite strongly bound, since the P_i dissociation constant values previously determined for the ternary enzyme–P_i–mannobiose and enzyme–P_i–β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc complexes (0.64 and 0.13 m*M*, respectively; Ladevèze *et al.*, 2013) are more than 200 times lower than that determined for *Ra*MP1 (belonging to the GH130_1 subfamily), the only other GH130 enzyme for which a P_i dissociation constant has been determined. In the apo structure, a molecule of glycerol, which was used as a cryoprotectant, occupied the −1 subsite. This glycerol molecule was hydrogen-bonded to the side-chain carboxylate moiety of Asp304 ($O^{\delta1}$ to O_3 and $O^{\delta2}$ to O_1), while its O_2 atom was hydrogen-bonded to the P_i, mimicking the interactions between mannose and the surrounding amino acids that were observed in the protein–ligand complexes. Interestingly, in the Uhgb_MP structures where binding of mannose in the −1 subsite occurred (PDB entries 4udj and 4udk in Table 1), the mannose ring was found in a stressed $B_{2,5}$ boat conformation stabilized by hydrogen bonding to Asp304 ($O^{\delta1}$ to the C_6 hydroxyl and $O^{\delta2}$ to the C_4 hydroxyl); the mutation of this critical residue to asparagine abolishes 96% of the activity (Ladevèze *et al.*, 2013). This unusual conformation of mannose was present in all six chains (Supplementary Fig. S1) and was previously observed in the *Bf*MGP structures (PDB entry 3was, for which *in cristallo* activity was observed, and PDB entry 3wat). It should be noted that this $B_{2,5}$ boat conformation is less unstable in mannose compared with other monosaccharides owing to the pseudo-equatorial position of the C_2–OH, which is in an *anti* configuration to the ring O atom, thus bending the C_3–OH towards the β-glycosidic O atom, in a *syn* axial position. The binding of mannose in the −1 subsite induced a large conformational movement in the active site. Indeed, in the apo form, where a glycerol molecule was observed in place of mannose, the amino moiety of Asn44 interacts with the C_3 hydroxyl of glycerol through a water molecule (conformation *A* in Fig. 2). Upon mannose binding (conformation *B*), the peptide bond between Phe43 and Asn44 flips in order to allow direct interaction of the Asn44 side chain with the C_3 and C_4 hydroxyls of the sugar. The side chain of the catalytic residue Asp104, mutation of which to asparagine completely abolished the activity (Ladevèze *et al.*, 2013), is also moved towards mannose in a position that is occupied by two water molecules in the apo form. In this *B* conformation, Asn44 stabilizes the catalytic Asp104 through hydrogen bonding, thereby imposing selection of the rotamer facing the mannose C_3 hydroxyl, which acts as a proton relay during catalysis (Fig. 2). This is the first time that such a concerted movement upon substrate binding in the −1 subsite has been reported for a glycoside phosphorylase. It must be noted that the *B* conformation of the catalytic Asp104 is probably the one that is active since it has also been observed for *Bf*MGP, which was demonstrated to be catalytically active *in cristallo* (Nakae *et al.*, 2013). The *A* conformation of the catalytic residue Asp104 of Uhgb_MP was observed in the structure of the GH130_2 *Tm*1225 protein in the apo form (PDB entry 1vkd). In the Uhgb_MP structures, the *B* conformation was only observed when mannose was bound in the −1 subsite, and is therefore independent of the presence of *N*-acetylglucosamine in the +1 subsite. Indeed, the *A* conformation was observed in the complex with *N*-acetylglucosamine alone (PDB entry 4udg, with glycerol in the −1 subsite and *N*-acetylglucosamine in the +1 subsite), while the *B* conformation was observed in the complex with mannose and *N*-acetylglucosamine (PDB entry 4udk, with mannose in the −1 subsite and *N*-acetylglucosamine in the +1 subsite). In the *A* and *B* conformations, *N*-acetylglucosamine bound in the +1 subsite was found in a $^4C_1$ relaxed chair conformation, stacked with Tyr103, and bound through hydrogen bonding to the C_6 hydroxyl group, the His174 imidazole ring, the Lys212 side chain and the Tyr242 hydroxyl group *via* a water molecule. The C_3 hydroxyl interacts with the Arg59 side-chain amine moiety, as well as with Asp304 $O^{\delta1}$ through a water molecule. The *N*-acetyl moiety is also involved in binding through hydrogen bonding between its NH group and the S atom of Met67 and between its carbonyl moiety and the His235 carbonyl *via* a water molecule. The hydrophobic methyl moiety faces the side chain of Ala207 and Met67, with these residues forming a hydrophobic pocket (Fig. 1).

These data enabled us to revise our previous Uhgb_MP model, which was built using the atomic coordinates of the *Tm*1225 protein from *T. maritima* MSB8 (PDB entry 1vkd) as a structural template, considering a monomeric form of the enzyme and using geometrical constraints provided by a classical inverting GH-like single-displacement mechanism. In this model, we previously hypothesized a +1 subsite formed by the Tyr103, Asp304, His174, Tyr240 and Phe283 residues, which are specifically conserved in the GH130_2 family, while the +2 subsite would be delineated by Tyr242, Pro279, Asn280 and Asp304. In this configuration, the exit of the catalytic tunnel would be orientated towards the inside of the oligomeric structure, thereby reducing access to the catalytic site. In addition, the conserved His235–Tyr240 loop from a cognate monomer would block the furrow that we have hypothesized. These new data emphasize the importance of taking into account the quaternary organization when modelling oligomeric enzymes, using SAXS data when possible to define the

low-resolution envelope and avoiding restraining the possible docking modes of substrates too much to envisage original reaction mechanisms. Here, thanks to the high-resolution crystallographic structures of hexameric Uhgb_MP, which is catalytically active (as detailed in the next section), in complex with mannose and *N*-acetylglucosamine, we propose a revised active-site topology in which the oligosaccharide chain rotates by 180°, inverting the −1 and +1 subsite positions. This new orientation allows the substrate to enter from the open side of the tunnel, and is in complete accordance with the orientation found in the crystal structure of *Bf*MGP in complex with its substrates (Nakae *et al.*, 2013).

As previously shown by kinetic analysis of Uhgb_MP, phosphorolysis of the N-glycan oligosaccharide core follows a mixed-type sequential random Bi-Bi mechanism (Ladevèze *et al.*, 2013). However, the order of substrate binding was not determined. Functional and structural data now lead us to suggest that the phosphorolytic catalytic mechanism is composed of a first step in which the inorganic phosphate is conveyed to the catalytic centre, followed by entry of the substrate to be phosphorolyzed. Indeed, the P_i binding site is located deeper in the catalytic site than the glycoside substrate, meaning that P_i could not bind after the disaccharide. Thus, mannose binding in the −1 subsite would induce a flip of the Phe43–Asn44 peptide bond from conformation *A* to conformation *B*, thus maintaining the side chain of Asp104 in a catalytically competent configuration. In reverse phosphorolysis, entry of mannose-1-phosphate would be the first step, followed by conformational changes of Phe43, Asn44, Ser45 and Asp104. Entrance of the acceptor would lead to the reverse phosphorolysis reaction. Regarding the phosphorolytic reaction itself, our data show that when the mannosyl moiety was present in the catalytic site, no water molecule was located where it could relay the proton from the catalytic Asp104 to the interosidic O atom. In addition, comparison between the apo and complexed forms showed that the change in Asp104 from conformation *A* to conformation *B* did not allow the catalytic aspartic acid to be at a hydrogen-bonding distance from the interosidic O atom, indicating that this residue is not directly involved in proton transfer to the interosidic O atom, as seen in the *Bf*MGP structures. In the latter case, Nakae and coworkers suggested a catalytic mechanism different from that of known inverting glycoside phosphorylases (GPs), involving the assistance of C_3—OH to relay proton transfer, because, like us, they did not observe a water molecule at the catalytic site in any of their structures (Supplementary Fig. S2). The stressed $B_{2,5}$ boat conformation of mannose bound in the −1 subsite and the *B* configuration of Asp104 which is only stabilized when mannose present in the −1 subsite was compatible with such a mechanism. The first step of phosphorolysis would thus be (i) nucleophilic attack of P_i on C_1 of the mannosyl moiety bound in the −1 subsite, (ii) a two-step protonation through Asp104 $O^{\delta2}$–Man O_3–GlcNac O_4, with the Asp104 side-chain carboxylic acid being located at 2.5 Å from the C_3–OH (Supplementary Fig. S2). However, even if the catalytic mechanism of Uhgb_MP and *Bf*MGP appears to be identical, clear differences in the substrate specificity of GH130



(*a*) (*b*)

Figure 3
Uhgb_MP homohexameric structure. (*a*) Uhgb_MP is a hexameric structure formed by a trimer of dimers. Individual monomers are shown in a single colour and are labelled *A*, *B*, *C*, *A*′, *B*′ and *C*′ for clarity. Each pair of dimers is coloured in pale/dark blue, green and red. (*b*) Close-up of the catalytic tunnel of a single Uhgb_MP protomer in the hexameric structure. The quaternary-structure assembly imposes structural constraints on active-site accessibility. The inorganic phosphate, the mannose and the *N*-acetylglucosamine molecules present in the catalytic site of protomer *A* are shown as sticks deeply buried in the catalytic site, which is accessible by a tunnel whose sides are formed by different protomers. The extremity of the tunnel is located at the centre of the plane formed by four Uhgb_MP molecules, in this case *A*, *A*′, *B* and *B*′.

subfamilies have been identified owing to structural motifs located in other parts of the protein, as further detailed below.

3.3. Uhgb_MP quaternary structure

In the Uhgb_MP crystal structure, each subunit of the homohexamer is roughly globular in shape. There is a large cavity inside the homohexamer and large holes at the centre of each of the three lateral planes formed by the homohexamer assembly (Fig. 3*a*). No discretely bound solvent molecules were found in these cavities, which are 15 Å in diameter at their narrowest point, indicating that these holes are large enough to enable substrate access to the active site of each protomer. The association of the surrounding subunits in the homohexamer caps the furrow of each protomer, giving rise to a funnel whose entrance is orientated towards the lateral aperture (Fig. 3*b*). The catalytic site is therefore deeply buried, with the phosphate ion and the −1 subsite located at the bottom of the tunnel.

Each dimer is formed of a large buried surface area of 1300 Å^2 involving two Uhgb_MP molecules linked by twofold symmetry (Fig. 4*a*). The interactions promoting dimer association involve the side chains of the residues at the interface, such as His123 and His196 stacking, and several hydrogen bonds or salt bridges involving side-chain atoms, such as between Arg195 and Gln142, Glu142 and Arg193, Asp93 and Arg195, and Glu189 and Arg193. Other polar interactions involve main-chain atoms of Ala144 and His194, Tyr122 and His196, and Trp191 and a symmetry mate. The homohexamer is formed by the association of three dimers arranged around a pseudo-threefold axis. Each dimer is related to its neighbours through symmetrical interactions involving each of the two protomers (covering an interaction surface of 840 Å^2 each; Fig. 4*b*). These interactions involve T-shaped stacking between the imidazole groups of His174 and His235 and hydrogen-bond interactions between side chains between Asn40 and Tyr264 and between Ser64 and Pro263. Main-chain carbonyl and amino groups are also involved in the assembly. More precisely, the Thr63 carbonyl is in contact with Tyr264 NH, while the side chains of Asn238 and Asn280 interact with the carbonyl group of Gly276 and the amide N atom of Tyr240, respectively. Finally, the side chain of Asn238 makes contact with the Pro279 carbonyl moiety.

SEC-MALLS and SAXS analysis confirmed the hexameric organization of Uhgb_MP in solution. The protein apparent molecular mass determined by SEC-MALLS was 240 kDa (n = 6.16; Supplementary Fig. S3). Guinier analysis of the SAXS data revealed that in phosphate and Tris–glycerol buffers, the radius of gyration (R_g) was considerably larger than the theoretical R_g, indicating protein aggregation. Based on data collected in Tris–glycerol buffer in the presence of 1 mM TCEP as reducing agent, an R_g value of 37.9 ± 0.08 Å was obtained, which is in good agreement with the theoretical R_g calculated from the apo crystal structure (∼37 Å). The pair distribution function $P(r)$ revealed a compact particle with a D_{max} of ∼110 Å that closely matches the largest dimension of

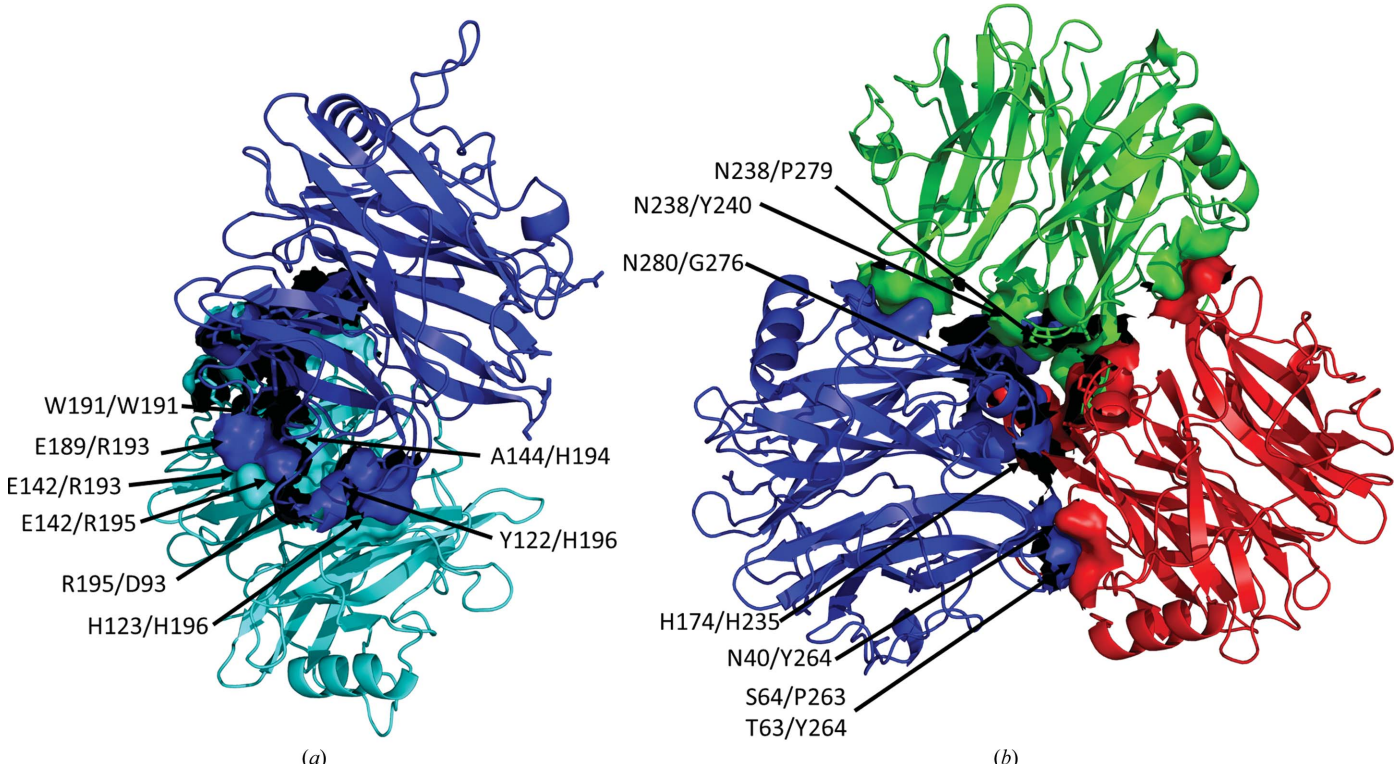


Figure 4
Interaction surfaces between the different Uhgb_MP protomers. (*a*) The interaction surfaces between two Uhgb_MP protomers involved in dimer formation. (*b*) The interaction surfaces between Uhgb_MP dimers to form the hexamer. Only the interactions involving the three upper monomers from each dimer are shown here, in order to clarify the view, as these surfaces are symmetrical in the lower monomers.

the hexamer (Supplementary Fig. S4). The *ab initio* envelope confirmed the compact shape of Uhgb_MP in solution with a trimer-of-dimers organization superimposable with the crystallographic structure. SAXS-based rigid-body modelling was attempted by taking into account the flexibility of the N-terminal residues missing from the crystal structure. The fit showed that the theoretical curve closely matched the experimental data, thus confirming the general hexameric organization of Uhgb_MP in solution (Supplementary Fig. S4).

Considering that (i) we never observed the existence of Uhgb_MP monomers, either in solution or in crystals, (ii) the quaternary structures deduced from both SAXS and crystallographic data are superimposable and (iii) the intracellular P_i concentration in bacteria is around 10 m*M* (Motomura *et al.*, 2011), meaning that P_i binds to Uhgb_MP *in vivo* as in the crystals, we conclude that hexamerization is required for enzyme activity and that the crystal structure presented here is certainly the most probable organization under physiological conditions.

All other known GPs belonging to the GH3, GH13, GH65, GH94, GH112, GT4 and GT35 families crystallize and act as homodimers. In contrast, it is difficult to find general features that control the oligomerization of GH130 enzymes, even for those belonging to the same subfamily. Indeed, all of the data that we gathered on functionally or structurally characterized GH130 enzymes showed that no single subfamily contained homogenous oligomerization profiles. Enzymes belonging to the GH130_NC group (the BACOVA_03624, BACOVA_02161, BT4094 and BDI_3141 proteins) crystallized as monomers, while the two functionally characterized proteins Teth514_1788 and Teth514_01789 have been shown to be dimeric and monomeric in solution, respectively (Chiku *et al.*, 2014). Various oligomeric forms (in solution or crystals) have been found in the GH130_1 (hexameric *Bf*MGP, pentameric *Rm*MGP and dimeric *Ra*MP1) and GH130_2 (hexameric Uhgb_MP, dimeric *Tm*1225, hexameric *Ra*MP2 and tetrameric *Bt*1033) subfamilies.

Moreover, the presence of the A*xxx*A*xxx*A motif in the *Bf*MGP N-terminal helix, which was thought to mediate oligomerization (Nakae *et al.*, 2013), was not found in *Ra*MP1 or *Rm*MGP, demonstrating that this particular motif is not the only element that is able to promote the formation of oligomers in enzymes belonging to the same GH130 subfamily. Finally, no particular secondary-structure element appeared to mediate interactions between the different Uhgb_MP protomers, which were only associated by surface interactions, without any involvement of secondary-structure elements, in contrast to what was observed for *Bf*MGP (the only GP of known structure with a similar homohexameric conformation). Indeed, the *Bf*MGP loop Thr42–Met68, which was found in contact with the N-terminal tab helix and is thought to help contact the cognate protomer, is completely absent in Uhgb_MP, and more generally in all GH130_2 sequences. Taken together, these data highlight the structural originality of GH130 enzymes among glycoside phosphorylases. Nevertheless, the solutions of many other GH130 structures will be required to be able to highlight possible structural markers of oligomerization.

3.4. Molecular bases of specificity towards mannosides

Uhgb_MP is the first member of the GH130_2 subfamily to be characterized both functionally and structurally. *Tm*1225 has only been structurally characterized, and no structural data is available for the functionally characterized GH130_2 members *Ra*MP2 and *Bt*1033. In contrast to the known enzymes classified into the GH130_1 subfamily (including *Bf*MGP, the only crystallized member of this subfamily; Senoura *et al.*, 2011; Nakae *et al.*, 2013), which exhibit a narrow

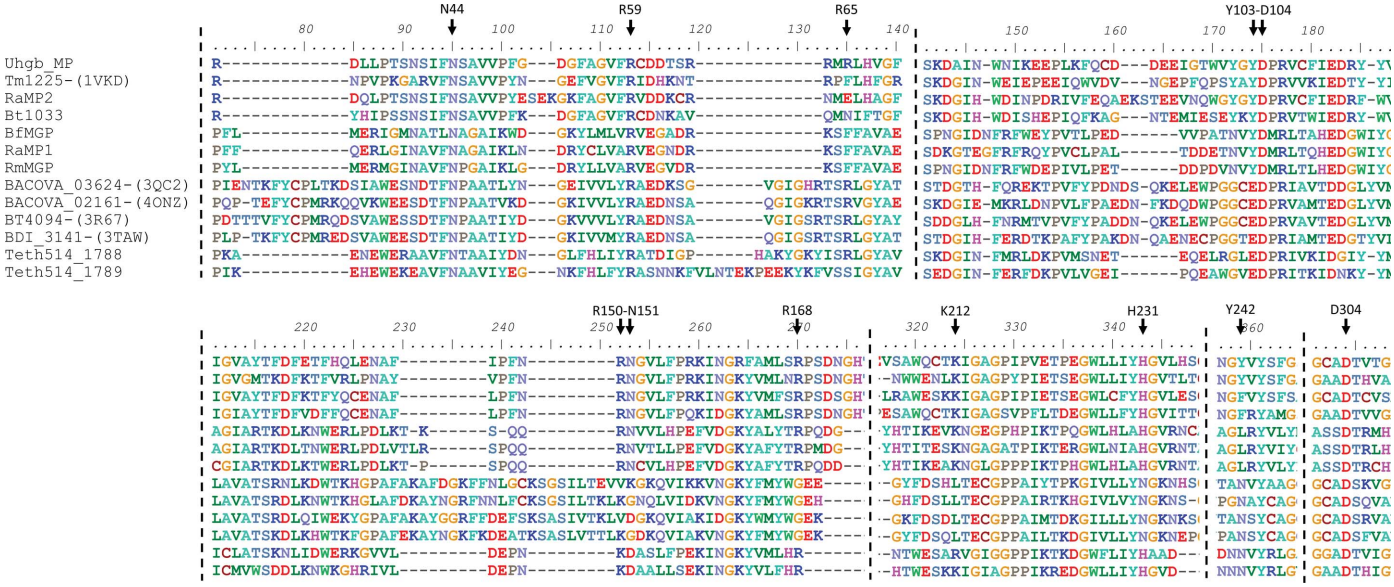


Figure 5
Sequence alignment of characterized GH130 enzymes. This alignment highlights the conservation of the catalytic and substrate-interacting residues among characterized GH130 enzymes, as well as family-specific loops.

specificity towards β-D-Man*p*-1,4-D-Glc, GH130_2 enzymes present a highly relaxed substrate specificity. Moreover, the Uhgb_MP structure is the first structure of an inverting glycoside phosphorylase that is active on a polysaccharide. Indeed, the only enzymes of known structure that are able to phosphorolyze polysaccharides are the retaining α-maltosyl phosphate:α-1,4-D-glucan-4-α-D-maltosyltransferases belonging to GH13 and glycogen or starch phosphorylases belonging to the GT35 family (Egloff *et al.*, 2001; Mirza *et al.*, 2006).

Uhgb_MP structures were compared with the six structures available for GH130 enzymes in order to identify any structural features that could explain the enzyme specificity, in particular towards N-glycan oligosaccharides, long manno-oligosaccharides and mannans. The residues involved in the catalytic machinery and in substrate binding are highly conserved in both the sequences and the three-dimensional structures, with the notable exception of the side chains of Asp104, Asn44 and Ser45 in the *A* configuration (Fig. 5 and Fig. 6), with overall r.m.s.d. values after C^{α} superposition of 2.1, 2.1, 2.1, 2.0, 2.0 and 1.0 Å for *Bf*MGP (25% identity with Uhgb_MP), BACOVA_03624 (23% identity), BACOVA_02161 (22% identity), *Bt*4094 (23% identity), BDI_3141 (25% identity) and *Tm*1225 (61% identity), respectively (Fig. 6). In addition, the P_i and glycosyl moieties in the −1 and +1 subsites superimposed perfectly with those present in the structure of *Bf*MGP in complex with inorganic phosphate, mannose and glucose. The electron density of mannose was separated from that of *N*-acetylglucosamine, in contrast to what would have been observed for the disaccharide β-D-Man*p*-1,4-D-Glc*p*NAc, because of the impossible superimposition of mannose O_1 and *N*-acetylglucosamine O_4. However, some structural features that are conserved in the subfamily explain the differences in substrate specificities observed between subfamilies (Fig. 6).

The most significant structural changes were identified in the Uhgb_MP Gly121–Pro125 loop, which is 11 residues longer in GH130_1 enzymes compared with those belonging to the GH130_2 and GH130_NC clusters (Fig. 6). This longer loop appeared at the extremity of the catalytic tunnel and, in the case of *Bf*MGP, actually filled it. Therefore, the accommodation of longer substrates than disaccharide would be impossible for GH130_1 enzymes, which is in accordance with the biochemical data published to date (Senoura *et al.*, 2011; Kawahara *et al.*, 2012; Jaito *et al.*, 2014). In GH130_2 enzymes the shorter loop would enable the entry of longer substrates, such as long manno-oligosaccharides or even mannans for Uhgb_MP, into the large cavity formed inside the quaternary structure between the three lateral planes of the homohexamer. Moreover, the +1 subsite flexibility of the GH130_2 members, which are able to accommodate *N*-acetylglucosamine and the C_2 epimer of glucose, would be explained by the location of the Uhgb_MP Arg65 residue. Indeed, the arginine side chain is at a distance of 6.43 Å from the O_2 atom compared with 2.96 Å for the side chain of the corresponding *Bf*MGP residue, Arg94, which would be responsible for the specificity of the GH130_1 members for β-D-Man*p*-1,4-D-Glc through hydrogen bonding to O_2 of glucose.

In addition, the stronger specificity of GH130_2 towards *N*-acetylglucosamine at the +1 subsite compared with glucose or mannose is explained by hydrophobic interactions of the *N*-acetyl-methyl moiety with Met67, which is not conserved in the GH130_1 subfamily,

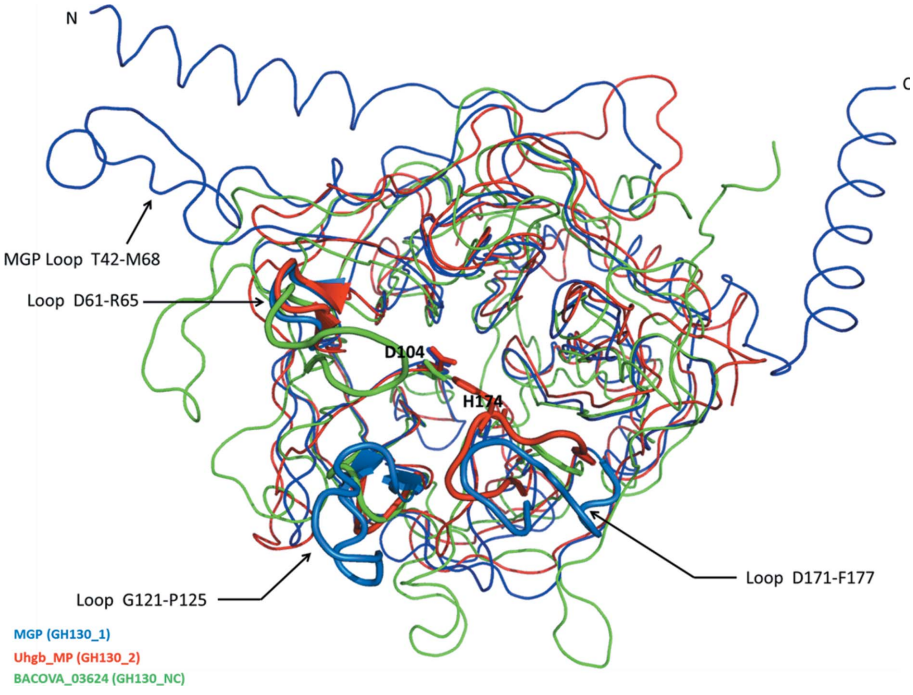


Figure 6
Superposition of GH130 structures. Superposition of Uhgb_MP (GH130_2; red) with mannosyl-glucose phosphorylase (*Bf*MGP) from *B. fragilis* NCTC 9343 (GH130_1; PDB entry 4kmi; blue), and BACOVA_03624 from *Bacteroides ovatus* ATCC 8483 (GH130_NC; PDB entry 3qc2; green), illustrating the structural differences between GH130 subfamilies. The catalytic Asp104 is shown in the *B* conformation. With the exception of the *Bf*MGP loop Thr42–Met68, which does not exist in the Uhgb_MP structure, the loops are numbered with respect to the Uhgb_MP sequence. The 11-residue longer Gly121–Pro125 loop, which is specific to GH130_1, fills the entrance to the Uhgb_MP tunnel. In place of the Asp61–Arg65 loop, an extension is observed for GH130_NC, capping the catalytic site. These two loops, which are specific to GH130_1 and GH130_NC, respectively, may explain the inability of enzymes belonging to these subfamilies to phosphorolyze long substrates. Loop Asp171–Phe177 (the so-called 'lid loop' in *Bf*MGP) is shorter in GH130_NC enzymes than in GH130_1 and GH130_2, thus allowing access to the active site. In addition, in GH130_1 enzymes loop Asp171–Phe177 is very mobile because of the GSGGG motif located at its base, which is locked close to the catalytic site only when a substrate is bound, as shown for *Bf*MGP structures. In contrast, in Uhgb_MP the loop is not so mobile and holds His174, which is conserved in GH130_2 and which has already been shown to be involved in the +1 subsite. The *Bf*MGP loop Thr42–Met68 in contact with the N-terminal helix involved in oligomerization is completely absent in Uhgb_MP even when both proteins are assembled as homohexamers.

while being present in half of the GH130_2 members, especially those acting on β-D-Man*p*-1,4-D-GlcNAc (*Bt*1033 and *Ra*MP2).

Moreover, in the structures containing *N*-acetylglucosamine, the Phe203 side chain of a cognate monomer was found close to Met67; these two residues form a hydrophobic pocket that interacts with the methyl group of the *N*-acetyl moiety. On the contrary, in the apo form and in the structure containing mannose alone in the −1 subsite, the Phe203 side chain was rather found rotated towards the exit of the catalytic tunnel. Therefore, Phe203, which is not conserved in the GH130_1 subfamily, while being present in 25% of GH130_2 members, would thus be a specific feature of GH130 enzymes that are able to phosphorolyze β-D-Man*p*-1,4-β-D-Glc*p*NAc.

The C_3 stereochemistry at the +1 subsite also appears to be critical since all pyranoside inhibitors of Uhgb_MP (allose, L-rhamnose and altrose) share an inversion of configuration at this position compared with that of mannose. This effect is probably owing to the proximity of Tyr103 (or the equivalent Tyr130 in *Bf*MGP), thus implying a steric constraint that would select an equatorial hydroxyl at this position in the case of a $^4C_1$ chair, which is the case for *N*-acetylglucosamine in the structure of the corresponding complex. We previously observed that a Y103E mutation strongly destabilizes Uhgb_MP, as is the case for the wild-type enzyme without phosphate. The Y103E mutation also significantly increases the ratio between hydrolysis and phosphorolysis, with the glutamic acid playing the role of the second catalytic acid required for hydrolysis (Ladevèze *et al.*, 2013). The role of Tyr103 in stabilizing the active-site conformation in the presence of phosphate is owing to hydrogen-bonding interactions between its lateral chain and that of Arg150, which interacts with phosphate (Fig. 1). The Y103E mutation would decrease hydrogen-bonding interactions, while positioning the glutamic acid at a hydrogen-bonding distance (less than 4 Å) from the interosidic O atom to allow hydrolysis to occur.

3.5. Significance

In this paper, we present the first structure of a phosphorolytic enzyme involved in N-glycan degradation in its apo form and in complex with mannose and *N*-acetylglucosamine. As previously highlighted by the integration of biochemical, genomic and metagenomic data, Uhgb_MP and GH130 enzymes more generally can be considered as interesting targets to study interactions between host and gut microbes, especially since GH130_2 sequences are overrepresented in the metagenomes of IBD patients. Further studies will be needed to confirm the physiological role of these enzymes and their potential involvement in damage to the intestinal barrier, such as metabolomic and transcriptomic analyses of the gut bacteria that produce them in the presence of N-glycans as a carbon source or inoculated in model animals with and without inhibitors. This three-dimensional structure paves the way for such studies through the design of specific GH130_2 inhibitors that could mimic substrate binding.

In addition, analysis of tertiary and quaternary structures led to the identification of structural features involved in the accommodation of long oligosaccharides and polysaccharides. This is a key feature that is unique to Uhgb_MP and is of great biotechnological interest for the conversion of hemicellulose into compounds with high added value. Identification of the structural determinants of the strong specificity of Uhgb_MP and other GH130_2 enzymes towards Man-GlcNAc also paves the way for the rational engineering of GH130 enzymes optimized for manno-oligosaccharide synthesis and diversification. Functional investigations of structurally characterized enzymes classified in the different GH130 sequence clusters would significantly advance our knowledge of the molecular bases of substrate specificities and improve our understanding of their role in key catabolic pathways, especially in the mammalian gut.

4. Related literature

The following references are cited in the Supporting Information for this article: Studier (2005).

Acknowledgements

This work was supported by the French Ministry of Higher Education and Research and by the French National Institute for Agricultural Research (INRA, 'Meta-omics of Microbial Ecosystems' research program). The equipment used for protein purification (ICEO facility), biophysical (DSF, SEC-MALLS) and crystallographic experiments are part of the Integrated Screening Platform of Toulouse (PICT, IBiSA). We thank Dr Valérie Guillet for technical assistance with SEC-MALLS. We also thank the European Synchrotron Radiation Facility (ESRF), Grenoble, France, in particular the staff of beamline ID-23-1. Experiments were also performed on the XALOC beamline at the ALBA Synchrotron (Barcelona, Spain) with the collaboration of the ALBA staff (Dr Jordi Juanhuix). Author contributions: Uhgb_MP production and purification, SL; crystallographic studies and X-ray data collection, SL, ST, GC and LM; SAXS experiments, PR, SL and GC; DSF and SEC-MALLS experiments, SL and ST. Experiments were designed by SL, ST and GPV. The manuscript was written primarily by SL and GPV with contributions from ST, GC and PR. SL, GC and PR prepared the figures. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement No. 283570).

References

Aebi, M., Bernasconi, R., Clerc, S. & Molinari, M. (2010). *Trends Biochem. Sci.* **35**, 74–82.
Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H. & Henrissat, B. (2012). *BMC Evol. Biol.* **12**, 186.
Burnaugh, A. M., Frantz, L. J. & King, S. J. (2008). *J. Bacteriol.* **190**, 221–230.
Chiku, K., Nihira, T., Suzuki, E., Nishimoto, M., Kitaoka, M., Ohtsubo, K. & Nakai, H. (2014). *PLoS One*, **9**, e114882.
David, G. & Pérez, J. (2009). *J. Appl. Cryst.* **42**, 892–900.

Egloff, M. P., Uppenberg, J., Haalck, L. & van Tilbeurgh, H. (2001). *Structure*, **9**, 689–697.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.
Jaito, N., Saburi, W., Odaka, R., Kido, Y., Hamura, K., Nishimoto, M., Kitaoka, M., Matsui, H. & Mori, H. (2014). *Biosci. Biotechnol. Biochem.* **78**, 263–270.
Juanhuix, J., Gil-Ortiz, F., Cuní, G., Colldelram, C., Nicolás, J., Lidón, J., Boter, E., Ruget, C., Ferrer, S. & Benach, J. (2014). *J. Synchrotron Radiat.* **21**, 679–689.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kawahara, R., Saburi, W., Odaka, R., Taguchi, H., Ito, S., Mori, H. & Matsui, H. (2012). *J. Biol. Chem.* **287**, 42389–42399.
Ladevèze, S., Tarquis, L., Cecchini, D. A., Bercovici, J., André, I., Topham, C. M., Morel, S., Laville, E., Monsan, P., Lombard, V., Henrissat, B. & Potocki-Véronèse, G. (2013). *J. Biol. Chem.* **288**, 32370–32383.
Ladevèze, S., Tarquis, L., Henrissat, B., Monsan, P., Laville, E. & Potocki-Veronese, G. (2014). International Patent WO/2015/014973.
Larkin, A. & Imperiali, B. (2011). *Biochemistry*, **50**, 4411–4426.
Lehle, L., Strahl, S. & Tanner, W. (2006). *Angew. Chem. Int. Ed.* **45**, 6802–6818.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.
Martens, E. C., Chiang, H. C. & Gordon, J. I. (2008). *Cell Host Microbe*, **4**, 447–457.
Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. (2009). *J. Biol. Chem.* **284**, 24673–24677.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mirza, O., Skov, L. K., Sprogøe, D., van den Broek, L. A. M., Beldman, G., Kastrup, J. S. & Gajhede, M. (2006). *J. Biol. Chem.* **281**, 35576–35584.
Motomura, K., Hirota, R., Ohnaka, N., Okada, M., Ikeda, T., Morohoshi, T., Ohtake, H. & Kuroda, A. (2011). *FEMS Microbiol. Lett.* **320**, 25–32.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nagae, M. & Yamaguchi, Y. (2012). *Int. J. Mol. Sci.* **13**, 8398–8429.
Nakae, S., Ito, S., Higa, M., Senoura, T., Wasaki, J., Hijikata, A., Shionyu, M., Ito, S. & Shirai, T. (2013). *J. Mol. Biol.* **425**, 4468–4478.
Nihira, T., Suzuki, E., Kitaoka, M., Nishimoto, M., Ohtsubo, K. & Nakai, H. (2013). *J. Biol. Chem.* **288**, 27366–27374.
Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). *J. Appl. Cryst.* **40**, s223–s228.
Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.
Renzi, F., Manfredi, P., Mally, M., Moes, S., Jenö, P. & Cornelis, G. R. (2011). *PLoS Pathog.* **7**, e1002118.
Roberts, G., Tarelli, E., Homer, K. A., Philpott-Howard, J. & Beighton, D. (2000). *J. Bacteriol.* **182**, 882–890.
Senoura, T., Ito, S., Taguchi, H., Higa, M., Hamada, S., Matsui, H., Ozawa, T., Jin, S., Watanabe, J., Wasaki, J. & Ito, S. (2011). *Biochem. Biophys. Res. Commun.* **408**, 701–706.
Sheng, Y. H., Hasnain, S. Z., Florin, T. H. J. & McGuckin, M. A. (2012). *J. Gastroenterol. Hepatol.* **27**, 28–38.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Suzuki, T. & Harada, Y. (2014). *Biochem. Biophys. Res. Commun.* **453**, 213–219.
Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W. & Etzler, M. E. (2009). Editors. *Essentials of Glycobiology*, 2nd ed. New York: Cold Spring Harbor Laboratory Press.
Zhu, Y., Suits, M. D. L., Thompson, A. J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K. W., Xiang, Y., Siriwardena, A., Williams, S. J., Gilbert, H. J. & Davies, G. J. (2010). *Nature Chem. Biol.* **6**, 125–132.

BIOLOGICAL
CRYSTALLOGRAPHY

Volume 71 (2015)

Supporting information for article:

Structural bases for N-glycan processing by mannoside-phosphorylase

Simon Ladevèze, Gianluca Cioci, Pierre Roblin, Lionel Mourey, Samuel Tranier and Gabrielle Potocki-Véronèse

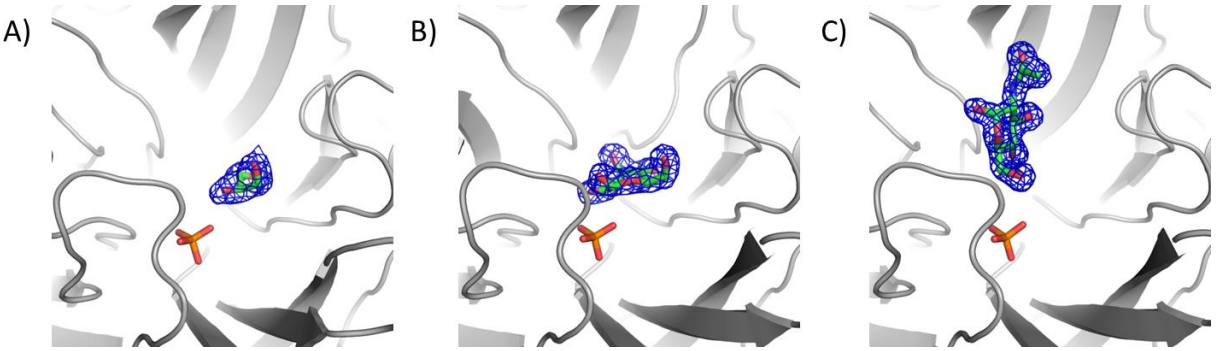


Figure S1 2Fo-Fc electron density maps contoured at 1,0 σ around the ligands. A) *apo* form, with a glycerol bound in the -1 sub-site; B) mannose bound in the -1 sub-site; C) *N*-acetylglucosamine bound in the +1 sub-site.

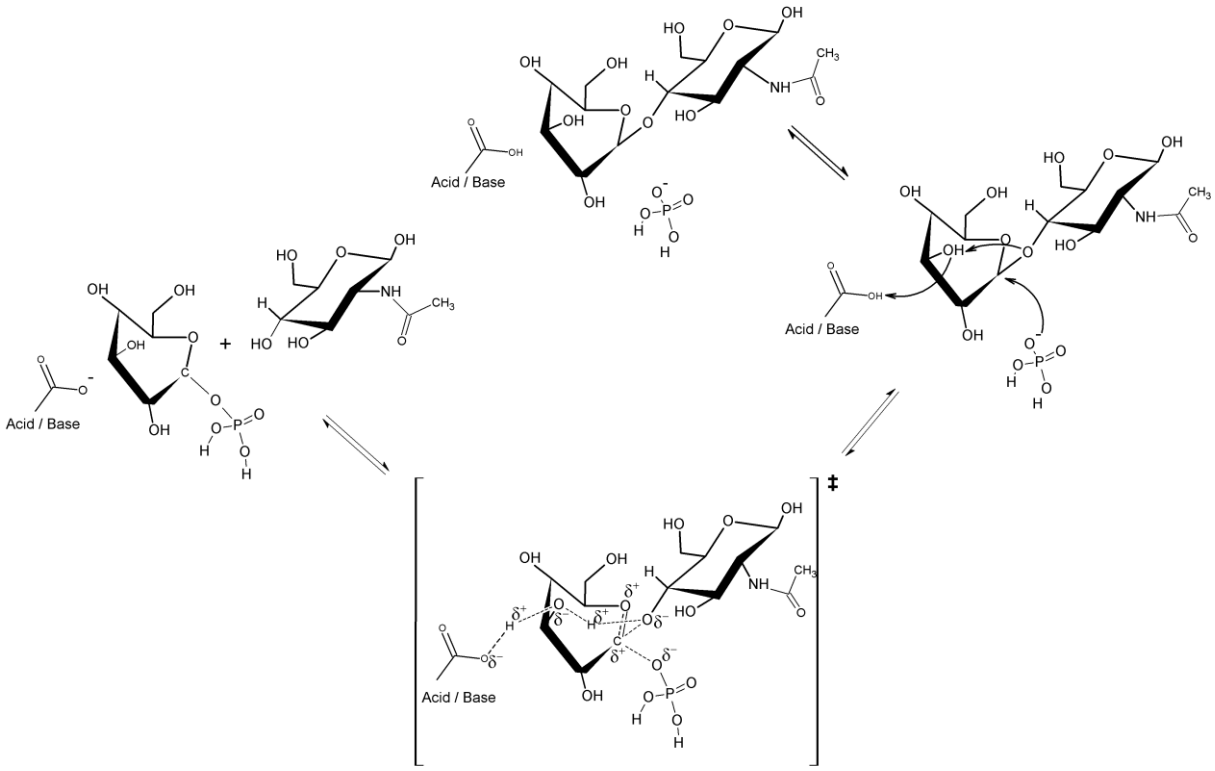


Figure S2 Proposed catalytic mechanism of β-D-Man*p*-1,4-D-Glc*p*NAc phosphorolysis by UhgbMP. This mechanism is similar to the one described by Nakae et al. for the GH130_1 *Bf*MGP protein. It involves the assistance of the C_3-OH of the mannosyl residue in the -1 sub-site to relay the proton from the catalytic D104 to the *N*-acetylglucosamine O_4.

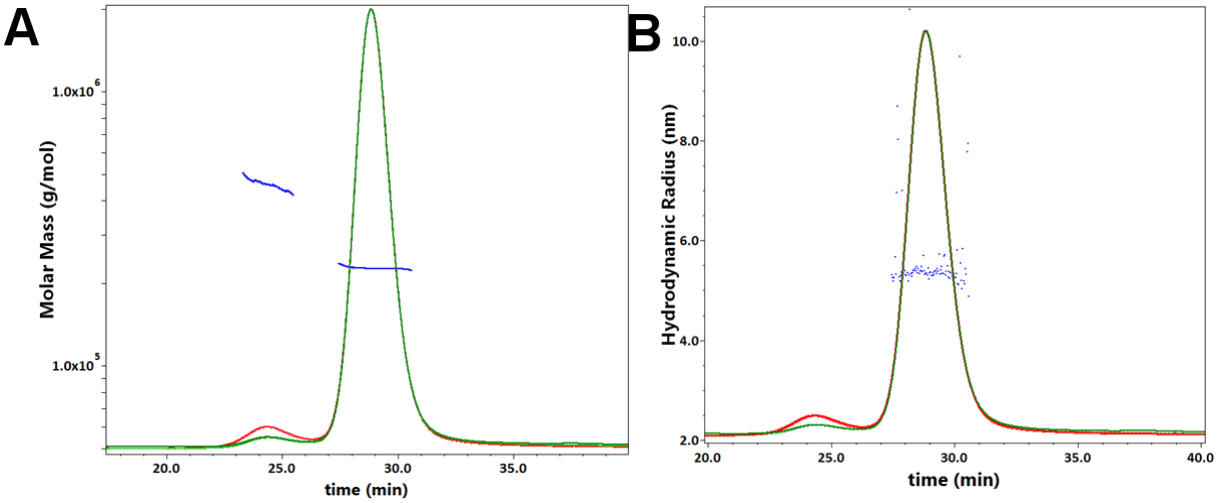


Figure S3 Size Exclusion Chromatography Multi-Angle Laser Light Scattering (SEC-MALLS) profile of Uhgb_MP. Protein apparent molecular mass was 240 kDa (n= 6.16). The main peak corresponding to the hexameric form accounted for 94% of the total mass fraction (Rh of 5.4 nm), with the presence of a small peak (3% of the total mass) that could correspond to a higher oligomeric form, possibly a dodecamer. A. Profile of the SEC-MALLS experiment displaying the light scattering signal (red curve) and the differential refractive index signal (green curve) versus elution time (min) for Uhgb_MP. The blue curve represents the molar mass ($g.mol^{-1}$) calculated across the elution peak according to SLS measurement. The average molar mass ($2.277x10^5$ ±0.025% $g.mol^{-1}$) is indicated by the dashed black line. B. Profile of the SEC-MALLS experiment showing the light scattering signal (red curve) and the differential refractive index signal (green curve) versus elution time (min) for Uhgb_MP. The blue curve represents the hydrodynamic radius (nm) calculated across the elution peak according to SLS measurement.

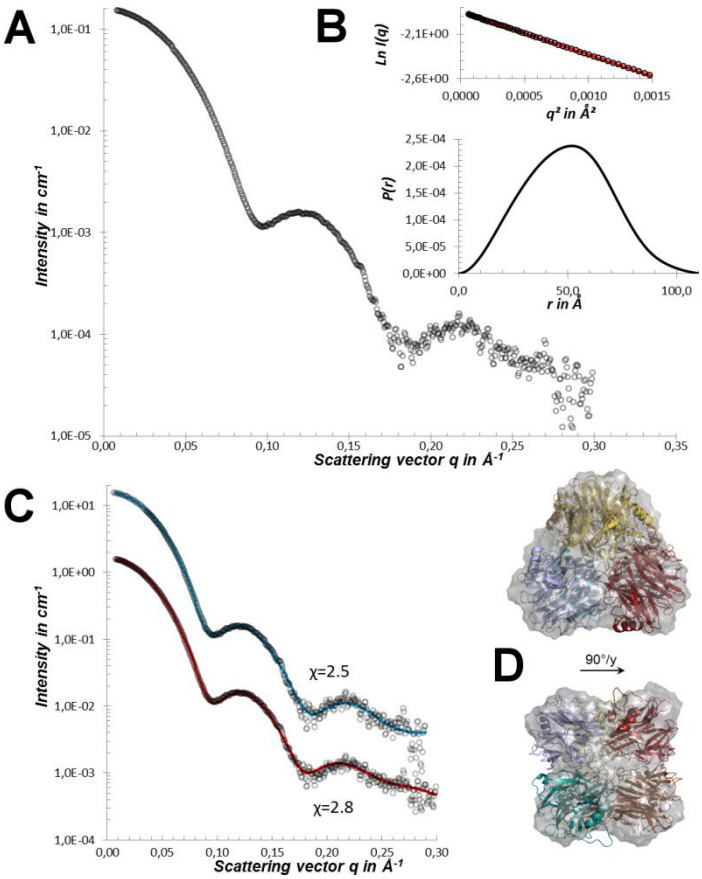


Figure S4 SAXS analysis. A. SAXS curve of Uhgb_MP in Tris-glycerol buffer supplemented with 1mM TCEP. B. Guinier fit and P(r) function. C. Fits of the ab-initio envelope (blue curve) and the crystal structure with modelled N-terms (red curve). D. Two orthogonal views of the model (shown as ribbon) superposed to the ab-initio envelop.

S1. Supplemental experimental procedures

S1.1. Uhgb_MP production and purification

The Uhgb_MP encoding gene was subcloned from pDEST17 vector (Ladevèze et al. 2013) into pET28a between NdeI and BamHI restriction sites, yielding a 6-histidine N-terminal thrombin cleavable tagged protein. Primers forward 5' TACGCTAGCCATATGAGTATGAGTAGCAAAGTT 3' and reverse 5' TACGCTGGATCCTCAGATGATGCTTGTA 3' were used to amplify the gene, using the Phusion® High-Fidelity DNA Polymerase (New England Biolabs), and to introduce NdeI and BamHI restriction

sites on both ends of the PCR product. They were subsequently digested, purified and ligated together. After *E. coli* TOP10 transformation, the construct was extracted and checked by Sanger sequencing.

To allow heterologous Uhgb_MP production, *E.coli* BL21-AI cells (Invitrogen) transformed with the Uhgb_MP::pET28a plasmid were cultured at 20°C for 24 hours in 200mL ZYM-5052 auto-induction medium (Studier 2005) supplemented with 50 µg/mL Kanamycin, inoculated at OD_{600} nm of 0.1. Next, two purification protocols were designed, protocol 1 to prepare the protein in 20 mM Tris-HCl pH 7.0, 300 mM NaCl for DSF experiments and first SAXS trials, and protocol 2, to prepare it in 20 mM potassium phosphate pH 7.0, 150 mM NaCl, for protein crystallization, SEC-MALLS and further SAXS experiments. Protocol 1: After culture, the cells were harvested and resuspended at an OD_{600} nm of 80 in lysis buffer 1 (50 mM Tris-HCl, pH 7.0, 300 mM NaCl, 10% (w/v) glycerol, 40 mM imidazole), supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche), 0.75 mg/ml lysozyme, and lysed by sonication. After centrifugation at 15,000 g for 30 min, soluble lysate was filtered through 0.2µm disposable filters (Sartorius) before loading onto a Histrap HP 1ml column equilibrated with lysis buffer 1 and connected to an ÄKTA Express (GE Healthcare) device. The column was washed with 15 volumes of affinity buffer 1 (20 mM Tris-HCl, pH 7.0, 300 mM NaCl, 10% (w/v) glycerol) supplemented with 200 mM imidazole, and the protein was then eluted in affinity buffer 1 supplemented with 500 mM imidazole. Fractions containing pure protein were identified by SDS-PAGE, pooled, and concentrated to 9–12 mg/ml. The concentrated pool was then applied onto a HiLoad 16/60 Superdex 200 (Amersham Biosciences) column and eluted using gel filtration buffer 1 (20 mM Tris-HCl pH 7.0, 300 mM NaCl). Uhgb_MP was purified to homogeneity with a yield of 70 mg/L of culture.

Protocol 2: After culture, the cells were harvested and resuspended in lysis buffer 2 (50 mM potassium phosphate pH 7.0, 150 mM NaCl, 10% (w/v) glycerol, 40 mM imidazole) supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche) and 0.75 mg/ml lysozyme to an OD_{600} nm of 80. The subsequent purification steps were the same as in Protocol 1, but affinity buffer 1 was replaced withby affinity buffer 2 (20 mM potassium phosphate pH 7.0, 150 mM NaCl, 10% (w/v) glycerol), and gel filtration buffer 1 was replaced with gel filtration buffer 2 (20 mM potassium phosphate pH 7.0, 150 mM NaCl).

S1.2. Differential Scanning Fluorimetry

Two in-house 96-well plates were designed to monitor pH stability in the range 4.5-9.0 using various common buffer species (citrate, acetate, phosphate, MES, PIPES, HEPES, TRIS, Imidazole, Bicine, and CHES) at a final concentration of 100 mM and increasing NaCl concentrations ranging 136 to 439 mM NaCl. Samples were loaded into a 96-well PCR plate (Bio-Rad) at a final volume of 11 µl. The concentration of protein in each well was 11 µM (0.4 $mg.ml^{-1}$), and 5X SYPRO Orange (Invitrogen) was added. DSF experiments were carried out using a CFX96 real-time PCR system (Bio-Rad) set to use the 480nm/500nm excitation and 560nm/580nm emission channels. The samples were heated from 20°C to 89.9°C. A single fluorescence measurement was taken every 0.3°C and each measurement lasted 6.5s. *Tm* was given by the inflection point of the curve of relative fluorescence unit (rfu) as a function of temperature (rfu=$f(T)$). The degree of thermal shift (ΔTm) was calculated as follows: $\Delta Tm = Tm(x) - Tm_0$, $Tm(x)$ being the *Tm* measured in each condition, and Tm_0 being the *Tm* of reference, measured in the purification buffer.

Ladevèze S, Tarquis L, Cecchini DA, Bercovici J, André I, Topham CM, Morel S, Laville E, Monsan P, Lombard V, Henrissat B, Potocki-Véronèse G (2013) J Biol Chem 288:32370–32383.

Studier FW (2005). Protein Expr Purif 41:207–234.

Third article

As explained in the first chapter of the 'Results' section, the over representation of GH130 sequences in the gut metagenome of IBD suffering patients, and especially those belonging to the GH130_2 subfamily, may indicate that these enzymes may be somehow involved in these diseases, by participating to the breakdown of the epithelium mucus layer. Designing inhibitors that would specifically inhibit host glycan degrading enzymes would therefore appear as a seducing strategy for IBD treatment. Various strategies can be adopted for enzyme inhibitor design. Among them, multivalency has been demonstrated to be very efficient for the specific inhibition of glycosidases involved in various human pathologies. However, the mechanism by which these molecules can reduce the activity of these enzymes remains to be elucidated. The following article describes the effects and limits of multivalent functionalization of iminosugars derivatives for inhibition of a panel of CAZymes including glycoside hydrolases, glycosyltransferases and glycoside phosphorylases (namely Uhgb_MP). The results of kinetic analyses and biophysical experiments also allowed us to propose a physical interaction model. Unexpected inhibition/activation concentration dependent effects of ultravalent compounds are also discussed. This work, to which I had the chance to contribute by testing the effect of multivalent imminosugars on Uhgb_MP activity and aggregation behavior, was directed by Sebastien Gouin (CNRS Nantes). The following paper was published in Bioconjugate Chemistry in 2015.

Note: In order to reduce the document length, a reduced version of this article's supplemental information is provided, without purification procedures, NRM and infrared spectra. The complete version of the supplemental information is accessible at the following address: http://pubs.acs.org.gate1.inist.fr/doi/suppl/10.1021/acs.bioconjchem.5b00081.

Bioconjugate Chemistry

Article

pubs.acs.org/bc

Polymeric Iminosugars Improve the Activity of Carbohydrate-Processing Enzymes

Yoan Brissonnet,[†] Simon Ladevèze,[‡] David Tezé,[†] Emeline Fabre,[§] David Deniaud,[†] Franck Daligault,[⊥] Charles Tellier,[⊥] Sergej Šesták,[∥] Magali Remaud-Simeon,[‡] Gabrielle Potocki-Veronese,[‡] and Sébastien G. Gouin*[,†]

[†]LUNAM Université, CEISAM, Chimie Et Interdisciplinarité, Synthèse, Analyse, Modélisation, UMR CNRS 6230, UFR des Sciences et des Techniques, 2, rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, France

[‡]Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, Institut National des Sciences Appliquées, CNRS UMR5504, Institut National de Recherche Agronomique, UMR792, 135 Ave. de Rangueil, 31077 Toulouse, France

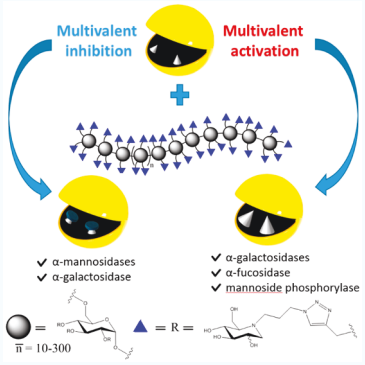
[§]Université Lille 1, Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576, 59650 Villeneuve d'Ascq, France

[⊥]UFIP, UMR CNRS 6286, Faculté des Sciences et Techniques, Université de Nantes, 2 rue de la Houssinière, F-44322 Nantes, France

[∥]Institute of Chemistry, Center for Glycomics, Slovak Academy of Sciences, Dúbravska cesta 9, 845 38 Bratislava, Slovakia

Supporting Information

ABSTRACT: Multivalent iminosugars have recently emerged as powerful tools to inhibit the activities of specific glycosidases. In this work, biocompatible dextrans were coated with iminosugars to form linear and ramified polymers with unprecedently high valencies (from 20 to 900) to probe the evolution of the multivalent inhibition as a function of ligand valency. This study led to the discovery that polyvalent iminosugars can also significantly enhance, not only inhibit, the enzymatic activity of specific glycoside-hydrolase, as observed on two galactosidases, a fucosidase, and a bacterial mannoside phosphorylase for which an impressive 70-fold activation was even reached. The concept of glycosidase activation is largely unexplored, with a unique recent example of small-molecules activators of a bacterial O-GlcNAc hydrolase. The possibility of using these polymers as "artificial enzyme effectors" may therefore open up new perspectives in therapeutics and biocatalysis.



INTRODUCTION

Multivalency is a well-established and powerful approach to develop strong and selective inhibitors of carbohydrate-binding proteins called lectins.[1−3] The so-called "glycoclusters", in which ligands are displayed in multiple copies on a synthetic core, mimic the glycocalyx displayed at the surface of cells and can lead to strong affinity enhancements compared to monovalent references. The simplicity and efficiency of the concept has led many groups to choose the multivalent way rather than improving the potency of a monovalent lead in a "lock and key" approach.

Until recently, the multivalent approach has been most successful with lectins, with only a few studies reported on carbohydrate-processing proteins. A striking example in the field was the development of multi- and polyvalent inhibitors of influenza neuraminidase, which showed outstanding ability to inhibit influenza virus replication in vivo.[4,5] To estimate the potential of the multivalent approach in glycosidase inhibition, we previously designed a set of mono-, di-, and trivalent inhibitors of the deoxynojirimycin (DNJ) iminosugar, a broad specificity glycosidase inhibitor.[6] The inhibitory activity of these compounds measured on nine glycosidases demonstrated that the multivalent strategy cannot be considered a general mode of inhibition of these enzymes as it is with lectins. However, a small but significant multivalent effect was detected for the first time on Jack bean α-mannosidase (JbMan). Much lower inhibitory activities and higher multivalent effects were then observed on the same enzyme with DNJ compounds of higher valencies.[7,8] Interestingly, multivalent effects were also seen on therapeutically relevant carbohydrate-processing enzymes including (i) β-glucocerebrosidase for which the deficiency in hydrolyzing glucosylceramide can lead to Gaucher disease,[9] the most common form of lysosomal storage disorders, (ii) the bacterial heptosyl-transferase WaaC incorporating heptosyl subunits in the outer membrane of Gram-negative bacteria,[10] (iii) human liver glycogen phosphorylase, a target for the treatment of noninsulin dependent diabetes mellitus,[11] and (iv) ManIIb as a representative of Golgi α-mannosidase II, a potential target of anticancer therapy.[12] Multivalent iminosugars were also shown to be able to correct the cell-deficiency of

Received: February 6, 2015
Revised: March 3, 2015
Published: March 5, 2015

DOI: 10.1021/acs.bioconjchem.5b00081
Bioconjugate Chem. 2015, 26, 766−772

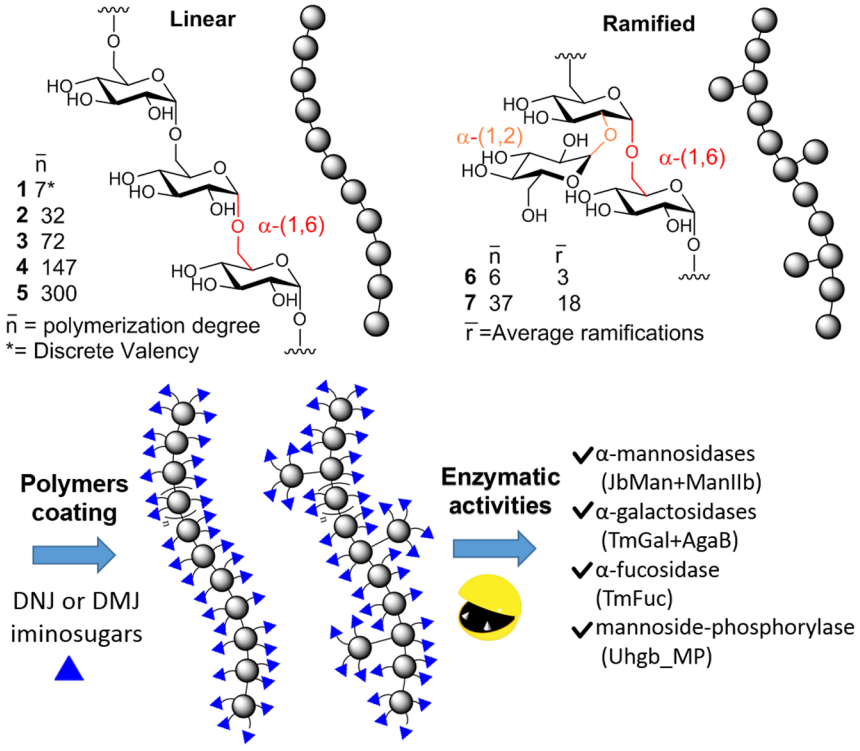


Figure 1. Schematic representation of the work. Linear (**1**−**5**) and ramified (**6**,**7**) dextran scaffolds were coated with DNJ and DMJ iminosugars. The effect on the enzymatic activity of the resulting biopolymers was assessed against six glycoside-processing enzymes.

a transmembrane regulator causing cystic fibrosis.[13] Taken together, these results highlight that the multivalent approach in inhibiting carbohydrate-processing enzymes is a rapidly expanding field,[14,15] with real pharmaceutical potential. The fact that all targeted enzymes do not respond positively to multivalency limits the potential applications, but is very beneficial in terms of selectivity.

Although the possibility of improving inhibitory activities is now well established, the crucial question of the multivalent binding mode(s) operating with this class of molecules is far from being answered. We,[12] and others,[16,17] addressed this question using the JbMan model and iminosugar or sp^2-iminosugar conjugates,[18] showing that the corresponding multivalent displays could promote the formation of large aggregates[12] and interact in enzyme subsites,[16] respectively. Recently, the importance of the density of iminosugar ligands was also studied with micellar glycopeptides.[19]

In this work, we were interested in studying the potential limit of the multivalent effect in the inhibition of glycoside-processing enzymes. As glycoclusters designed for lectins often show a plateau of affinity after exceeding a discrete valency, it is of interest to determine whether the same interaction profile also occurs with the enzymes targeted in this study. Importantly, this led to the discovery that polymeric iminosugars can also function as significant activators of specific glycoside hydrolases.

RESULTS AND DISCUSSION

Hydrophilic polymers, with an unprecedentedly high average number of DNJ or deoxymannojirimycin (DMJ) iminosugars ranging from 20 to 900, were designed from commercial dextrans. Inhibitory activities were assessed on a panel of five retaining glycosidases from the GH 29 (α-L-fucosidase from *Thermotoga maritima*, TmFuc), GH 36 (α-D-galactosidases from *T. maritima*, TmGal, and *Geobacillus stearothermophilus*, AgaB), and GH38 families (α-D-mannosidases from *Canavalia ensiformis*, JbMan, and *Drosophilia melanogaster*, ManIIb) and an inverting *N*-glycan phosphorylase (Uhgb_MP) belonging to the GH130 family. Multivalent effects were quantified as a function of the valency of the ligands for the model JbMan, the anticancer target Golgi α-mannosidase (ManIIb), and for TmGal.

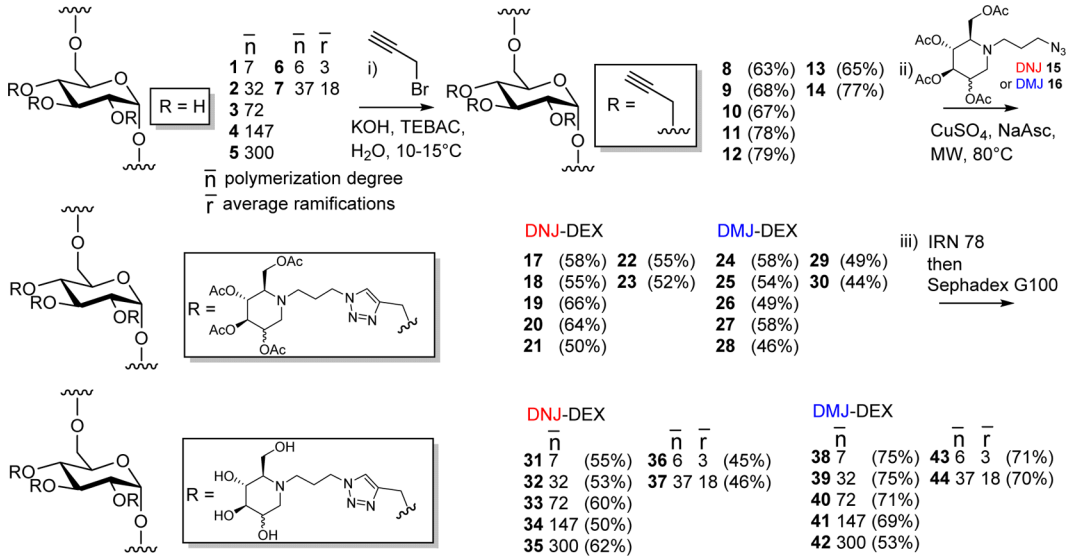
We selected various commercial dextrans products of different molecular weights for the design of the polymeric iminosugars. Dextrans are highly hydrophilic and nontoxic scaffolds. They are homogeneous polymers of glucosyl units composed of approximately 95% α-(1,6) linkages.[20,21] The three available hydroxyls of each glucose unit can be functionalized and further loaded by iminosugars leading to highly dense DNJ and DMJ-coated dextrans (DNJ-DEX and DMJ-DEX). As the spatial presentation of DNJ in low valency ligands is a critical factor for enzyme affinity and selectivity,[12] the set of linear dextrans was completed with two ramified analogues bearing ~33% of α-1,2 linked glucosyl units (Figure 1).

To produce scaffolds **6** and **7**, linear α-1,6 dextrans from *Leuconostoc mesenteroides* NRRL B-512F were incubated with sucrose and the α-1,2 branching sucrase, GBD-CD2, a truncated enzyme derived from the DSR-E dextransucrase from *L. mesenteroides* NRRL B-1299.[22] Their MW mass was calculated from the size of their backbone and their degree of branching, evaluated by NMR. Purification was achieved by preparative ion exchange followed by C18 chromatography.[23]

Dextrans were first functionalized by propargylic groups following a previously described protocol for arabinogalactans using a biphasic mixture of solubilized dextrans in an aqueous basic solution and propargyl bromide in toluene.[24] A catalytic amount of triethylbenzylammonium chloride (TEBAC) was

DOI: 10.1021/acs.bioconjchem.5b00081
Bioconjugate Chem. 2015, 26, 766−772

Scheme 1. Chemical Synthesis of the Linear and Ramified Ultravalent DNJ-DEX 31−37 and DMJ-DEX 38−44



shown to improve significantly the reaction yields of the linear **8−12** and ramified **13−14** alkynyl-dextrans (Scheme 1). Relative integration of the ^{1}H NMR signals for the sugar anomeric and propargylic protons indicated a full propargylation of the dextrans. This was further confirmed by infrared spectra showing the disappearance of the large 2700−3600 cm^{-1}(υ_{O-H}) band and new characteristic ($\upsilon_{\equiv C-H}$) and ($\upsilon_{C\equiv C}$) bands at around 3290 and 2120 cm^{-1}, respectively.

Azidopropyl-DMJ **16** was obtained in nine steps (Figure S1) using a similar procedure to that described for the DNJ analogue **15**.[12] Both iminosugars were grafted onto alkynyl-dextrans **8−14** by a microwave-assisted copper-catalyzed azide−alkyne cycloaddition to give multivalent iminosugars **17−30**. The complete conversion of the alkynes was evidenced by ^{1}H and ^{13}C NMR and the unique formation of the 1,4 cycloadduct was shown by the large $\Delta(\delta C\text{-}4-\delta C\text{-}5)$ values (>20 ppm).[25] Hydroxyl groups were deprotected with the basic Amberlite resin IRN 78 and the crude dextrans **31−44** were purified by size-exclusion chromatography on a Sephadex G100 column.

In addition, adequate monovalent references were synthesized to assess the level of potential multivalent effects (Figure 2). A tetravalent derivative based on a similar scaffold was also designed and included in the assays (Figure 2).

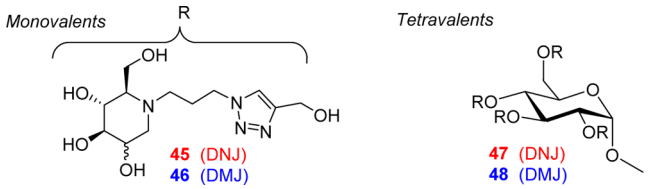


Figure 2. Structure of the monovalent DNJ **45** and DMJ **46** references and the tetravalent DNJ **47** and DMJ **48**.

Inhibitory activities were first evaluated on JbMan, for which the first and highest multivalent effects were reported with iminosugars.[6−8]

The inhibitory activities (K_{ic} values) reported in Table 1 are expressed in total concentration of iminosugars (valency-corrected) and not in polymer molecules concentrations (K_i values), for a fair assessment of the multivalent effect. As an example, the best compound **42**, with an average DMJ valency of 902, displayed a K_i value of 1.5 nM and a valency-corrected K_{ic} value of 1.4 μM (1.5 × 902) (Table 1). We also calculated the β value, a factor frequently used to quantify the multivalent effect observed with a glycocluster against a specific lectin. Similarly, the β factor represents here the improved inhibitory potency per iminosugar, and is calculated by dividing the K_i value of monovalent references **45** or **46** by the K_{ic} value of the ultravalent DNJ-DEX or DMJ-DEX, respectively. As an example, the β factor for **36** is calculated as follows (394:25 = 16).

The results showed that a significant multivalent effect occurred for the whole set of DNJ- and DMJ-DEX. In all cases, the linear DMJ-DEX were better inhibitors than their DNJ

Table 1. Inhibitory Activities against JbMan[a]

$\tilde{V}_{Im}$	DNJ-Cpd	K_{ic} (μM)	β	DMJ-Cpd	K_{ic} (μM)	β
1	**45**	394 ±28	1	**46**	167 ±17	1
4	**47**	196 ±64	2.0	**48**	132 ±48	1
23	**31**	6.0 ±1.5	66	**38**	2 ±0.5	79
29	**36**	25 ±1.8	16	**43**	4 ±1.2	40
98	**32**	5.7 ±0.3	69	**39**	2.2 ±0.1	74
167	**37**	32 ±1.8	12	**44**	3.4 ±0.1	49
218	**33**	12 ±2.4	34	**40**	2.0 ±0.4	83
443	**34**	26 ±3.6	15	**41**	2.3 ±0.3	71
902	**35**	11 ±2.6	36	**42**	1.4 ±0.1	121

[a]$\tilde{V}_{im}$ = average valency of iminosugars. K_{ic} = inhibitory activity (μM) calculated per iminosugar, average of three measurements. β = affinity enhancement for each iminosugar calculated by dividing K_i values of the adequate monovalent reference (**45** for DNJ compounds and **46** for DMJ) by the valency-corrected K_{ic} value of the ultravalent compound. Ramified compounds in gray cells.

DOI: 10.1021/acs.bioconjchem.5b00081
Bioconjugate Chem. 2015, 26, 766−772

analogues and showed higher multivalent effects in the valency range considered (Figure 3). Ramified polymers **36**, **37**, **43**, and

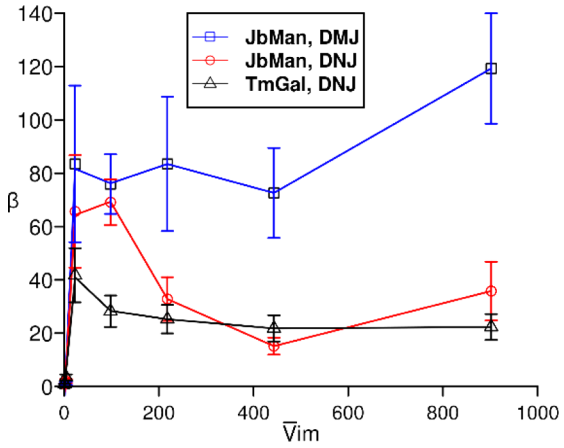


Figure 3. Evolution of the multivalent inhibitory effect β in function of the average iminosugar valency.

44 showed significantly lower β values than the linear ones, indicating the detrimental role of the spatial presentation of the ligands. β values rapidly reached a peak for valencies around 20 (Figure 3), then a subsequent decrease was observed for the DNJ-DEX series. The increased β value of +65% for the longest DMJ-DEX **42** compared to **38** can be seen as moderate considering the high valency gap between these two compounds (from 23 to 902 DMJ units). The maximal β value of 121 observed here for **42** can also be compared with published β values of the most potent ligands, with lower and discrete valencies, such as a tetravalent DNJ based porphyrine (β = 200),[12] and a tetradecavalent DNJ-based cyclodextrin (β = 610).[8] The higher β values obtained with these low valency compounds and the moderate improvement observed here at high DNJ or DMJ loading suggest that focusing on discrete compounds with adequate iminosugar topologies is more relevant for JbMan inhibition than overincreasing the ligand valency.

The *Drosophilia melanogaster* enzyme homologous to the human ManIIb was used for the assay. Results obtained with DMJ-DEX are presented in Table 2. They confirm that ManIIb is a potentially interesting target for multivalent inhibition.[12] However, this class of compounds only showed a significant multivalent effect at the highest valencies.

DNJ-DEX and DMJ-DEX were then assessed against TmGal which is a monomeric type I α-galactosidase belonging to the GH36 family. The DMJ derivatives did not show significant inhibition in preliminary assays and were not considered further. K_{ic} and β values for DNJ-DEX are reported in Table 2 and Figure 3. Although the whole set of DNJ compounds were competitive binders against JbMan, an uncompetitive binding mode prevailed with TmGal, as evidenced by the Lineweaver–Burk representations (Figures S8, S9).

Interestingly, the DNJ-DEX **31**–**35** showed significant multivalent effects against TmGal (Table 2), which is therefore a newly identified glycosidase target sensitive to multivalency. Consistent with the observations made with JbMan, the ramified compounds **36** and **37** showed much lower β values than linear DNJ-DEX. The inhibition profile of DNJ-DEX was also very similar to that observed on JbMan, with a peak at low valency, followed by a plateau of inhibition (Figure 3).

Table 2. Inhibitory Activities against ManIIb and TmGal[a]

	ManIIb			TmGal		
$\tilde{V}_{Im}$	DMJ Cpd	K_{ic} (μM)	β	DNJ Cpd	K_{ic} (μM)	β
1	**46**	435 ±68	1	**45**	960 ±150	1
4	**48**	62 ±11	1.8	**47**	276 ±44	3.5
23	**38**	306 ±84	1.4	**31**	23 ±2.0	42
29	**43**	--	--	**36**	45 ±3.6	21
98	**39**	321 ±96	1.4	**32**	34 ±1.8	28
167	**44**	141 ±48	3.1	**37**	159 ±27	6.0
218	**40**	--	--	**33**	38 ±2.1	25
443	**41**	66 ±12	6.6	**34**	44 ±3.0	22
902	**42**	51 ±18	8.5	**35**	43 ±2.6	22

[a] $\tilde{V}_{im}$ = average valency of iminosugars. K_{ic} = inhibitory activity (μM) calculated per iminosugar, average of three measurements. β = affinity enhancement for each iminosugar calculated by dividing K_i values of the adequate monovalent reference (**45** for DNJ compounds and **46** for DMJ) by the valency-corrected K_{ic} value of the polyvalent compound. Ramified compounds in gray cells. – = not evaluated.

Interestingly, the tetravalent **47** and ultravalent **31**–**35** showed dramatically different behavior at low concentrations (below 8 μM), switching from being inhibitors to significant activators of the enzymatic activity (Figure S12). This activation phenomenon was not observed with the monovalent DNJ derivatives. A similar scenario, but at a lower level, was observed with the whole set of DNJ-DEX and DMJ-DEX **31**–**42** against TmFuc, which belongs to the GH29 family and forms a hexameric complex in solution.[26] In comparison, the monovalent derivatives **45** and **46** showed only a slight inhibition.

This enhancing effect was more prevalent with our next target, the α-galactosidase B from *G. stearothermophilus* (AgaB). This enzyme is a tetrameric α-galactosidase belonging to the GH36 family. High activity enhancements were observed with the whole set of ultravalent compounds, with tetravalent DNJ **47** showing a specific behavior being either an enzyme activator or an inhibitor depending on the concentration.

To quantify the enhancing or inhibitory effect better, we calculated the variation in the initial turnover number (k_{cat}), which represents the number of substrate molecules converted to product by AgaB per unit of time. The activation effect of different concentrations of iminosugars on the activity of AgaB was tested at a fixed concentration of *p*NP-Gal as substrate (Figure 4), where F represents the relative activity compared to the rate in the absence of iminosugar (activation if F > 100%, inhibition if F < 100%).

The F values obtained for tetravalent **47** and **48** and ramified compounds **37** and **44** with ∼167 DNJ and DMJ units, respectively, are presented in Figure 4. The results clearly showed different inhibition and activation profiles depending on both the valency and nature of the iminosugar. Tetravalent DNJ-based **47** activated AgaB below 80 μM but switched to being an inhibitor at higher concentration, consistent with what was observed against TmGal.

The DMJ analogue **48** showed a pure AgaB activation with an F value of 140% at 250 μM. Ultravalent compounds **37** and **44** were potent activators in the concentration range tested. **37**

DOI: 10.1021/acs.bioconjchem.5b00081
Bioconjugate Chem. 2015, 26, 766–772

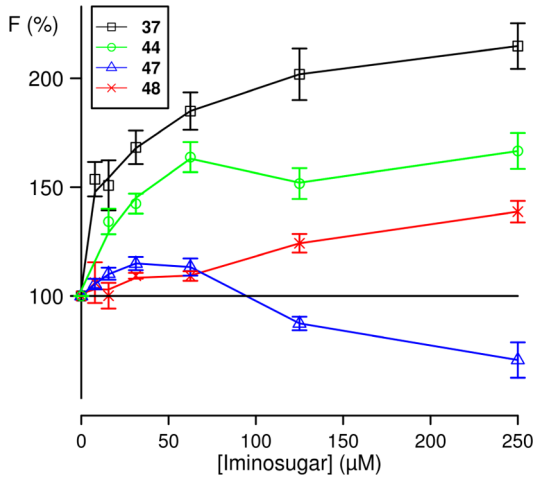


Figure 4. Activation of AgaB promoted by compounds **37**, **44**, **47**, and **48**. Average of three measurements per value.

was the most potent of the series, reaching 220% of the initial AgaB activity. Such an activation phenomenon was never reported with multivalent iminosugars.

Changing the concentration of a specific ultravalent compound raises the unexpected possibility of fine-tuning the enzymatic activity. This is possible because some ultravalent compounds have a specific activation/inhibition profile depending on the concentration as illustrated by compound **37** in Figure 5. For example, AgaB and TmGal would be activated in

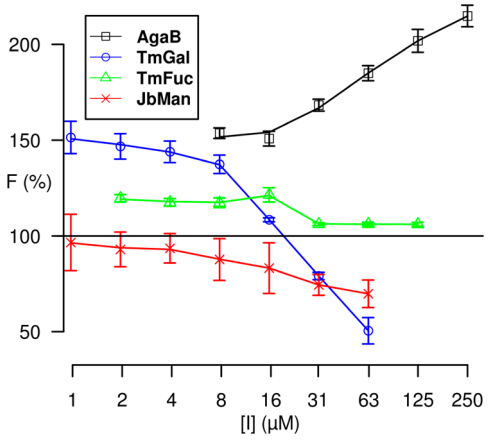


Figure 5. Activation or inhibition of AgaB, TmGal, TmFuc, and JbMan, promoted by compound **37**, average of three measurements per value.

the presence of low concentrations of **37** (8 μM and below). In contrast, higher concentrations of **37** (20 μM and above) will activate AgaB while inhibiting TmGal. As AgaB and TmGal are two enzymes from the same family (GH36), these results illustrate that ultravalent iminosugars are potential modulators able to discriminate between related glycosidases.

The iminosugars were then tested on another biologically relevant target, the mannoside-phosphorylase (EC 2.4.1.-) Uhgb_MP. Belonging to the GH130 family, this enzyme produced by an uncultivated *Bacteroides* human gut bacterium, is involved in *N*-glycan degradation.[27] Uhgb_MP and its homologues are encoded by highly abundant genes in the gut microbiome, especially that of patients suffering from inflammatory bowel diseases. They could thus participate in the alteration of the intestinal barrier, through human *N*-glycan degradation. **44** was a very strong activator of Uhgb_MP up to 500 μM. A 70-fold activation was even reached at this concentration when phosphorolysis activity was quantified by using *p*NP-β-D-mannopyranose as substrate (Figure 6). At 3

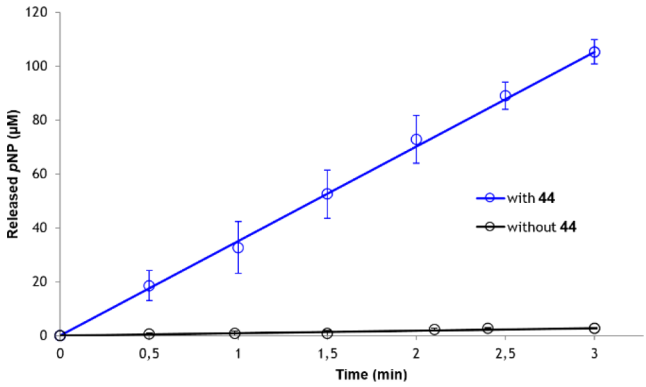


Figure 6. Activation of Uhgb_MP promoted by compound **44** at 500 μM, average of three measurements per value.

mM **44**, an inhibition was observed, due to enzyme aggregation and precipitation, which was clearly visible in the reaction medium. Uhgb_MP aggregation was confirmed by DLS analysis, which revealed particles increasing in size with increasing **44** concentration. At 500 μM particle size of 1990 nm was observed (Figure S13).

CONCLUSION

In summary, we have developed a set of polymeric DNJ and DMJ iminosugars to explore the multivalent effect on the inhibition of carbohydrate-processing enzymes of various structures and functions, at high iminosugar valencies ranging from 20 to 900.

DNJ- or DMJ-based polymers were shown to reach high β-values rapidly at low valencies against JbMan, the most studied glycosidase target for multivalency, and TmGal, a new glycosidase identified here as sensitive to multivalency, which extend the multivalent concept to the inhibition of α-galactosidases. At higher valencies, a drop in β-values was observed, followed at best by a moderate increase, or by a plateau of inhibition. These results suggest a threshold limit in the multivalent enzyme inhibition, as often observed with lectins. However, it should be noted that the multivalent effect limits observed here apply strictly only to dextran conjugates and may not be extrapolated to other iminosugar-coated polymers with different spatial presentation of the ligands.

Importantly, the polymeric DNJ and DMJ were shown to promote enzyme activity significantly on several targets. High activity enhancements were observed on a galactosidase (AgaB, F = 220%) and a mannoside-phosphorylase (Uhgb_MP, ×70-fold). We will next investigate the mechanism(s) leading to the enhanced enzymatic activity to decipher if DNJ- and DMJ-DEX are potential allosteric effectors or if a trans-glycosylation (or reverse-phosphorylation for Uhgb_MP) process occurs. Enzyme activations have been previously described with cyclodextrin-conjugates stabilizing the protein conformation through multiple hydrogen bonds and electrostatic interactions.[28] We do not favor this hypothesis here, as no signs of self-decreasing catalytic efficiency were seen during the experiments.

DOI: 10.1021/acs.bioconjchem.5b00081
Bioconjugate Chem. 2015, 26, 766−772

Glycosidase activation remain an unexplored area and as far as we know this work is only the second example describing glycoside hydrolase activators. Small-molecule effectors of a bacterial glycoside hydrolase (O-GlcNAc hydrolase) have been recently published.[29] The fact that multivalent iminosugars can function as "artificial enzyme effectors" is an interesting finding that may open up new biological perspectives in therapy and biocatalysis where fine-tuning glycoside-hydrolase activity is required.

■ ASSOCIATED CONTENT

Supporting Information

Synthesis conditions, protocols, inhibitory activities. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sebastien.gouin@univ-nantes.fr. Phone: 00-33-2-51125406. Fax: 00-33-2-51125402.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was carried out with financial support from the Centre National de la Recherche Scientifique and the Ministère Délégué à l'Enseignement Supérieur et à la Recherche. We cordially thank Angeline Rizzo for technical assistance regarding Uhgb_MP inhibition and activation.

■ REFERENCES

(1) Lundquist, J. J., and Toone, E. J. (2002) The cluster glycoside effect. *Chem. Rev. 102*, 555−578.

(2) Deniaud, D., Julienne, K., and Gouin, S. G. (2011) Insights in the rational design of synthetic multivalent glycoconjugates as lectin ligands. *Org. Biomol. Chem. 9*, 966−979.

(3) Bernardi, A., Jiménez-Barbero, J., Casnati, A., De Castro, C., Darbre, T., Fieschi, F., Finne, J., Funken, H., Jaeger, K.-E., Lahmann, M., et al. (2013) Multivalent glycoconjugates as anti-pathogenic agents. *Chem. Soc. Rev. 42*, 4709−4727.

(4) Macdonald, S. J. F., Cameron, R., Demaine, D. A., Fenton, R. J., Foster, G., Gower, D., Hamblin, J. N., Hamilton, S., Hart, G. J., Hill, A. P., et al. (2005) Dimeric zanamivir conjugates with various linking groups are potent, long-lasting inhibitors of influenza neuraminidase including H5N1 avian influenza. *J. Med. Chem. 48*, 2964−2971.

(5) Honda, T., Yoshida, S., Arai, M., Masuda, T., and Yamashita, M. (2002) Synthesis and anti-influenza virus activity of 7-O-alkylated derivatives related to zanamivir. *Bioorg. Med. Chem. Lett. 12*, 1929−1932.

(6) Diot, J., García-Moreno, M. I., Gouin, S. G., Ortiz Mellet, C., Haupt, K., and Kovensky, J. (2009) Multivalent iminosugars to modulate affinity and selectivity for glycosidases. *Org. Biomol. Chem. 7*, 357−363.

(7) Compain, P., Decroocq, C., Lehl, J., Holler, M., Hazelard, D., Mena Barragán, T., Ortiz-Mellet, C., and Nierengarten, J.-F. (2010) Glycosidase inhibition with fullerene iminosugar balls: a dramatic multivalent effect. *Angew. Chem., Int. Ed. 49*, 5753−5756.

(8) Decroocq, C., Rodríguez-Lucena, D., Russo, V., Mena Barragán, T., Ortiz-Mellet, C., and Compain, P. (2011) The multivalent effect in glycosidase inhibition: probing the influence of architectural parameters with cyclodextrin-based iminosugar click clusters. *Chem.—Eur. J. 17*, 13825−13831.

(9) Decroocq, C., Rodríguez-Lucena, D., Ikeda, K., Asano, N., and Compain, P. (2012) Cyclodextrin-based iminosugar click clusters: the first examples of multivalent pharmacological chaperones for the treatment of lysosomal storage disorders. *ChemBioChem 13*, 661−664.

(10) Durka, M., Buffet, K., Iehl, J., Holler, M., Nierengarten, J.-F., and Vincent, S. P. (2012) The inhibition of liposaccharide heptosyltransferase WaaC with multivalent glycosylated fullerenes: a new mode of glycosyltransferase inhibition. *Chem.—Eur. J. 18*, 641−651.

(11) Rath, V. L., Ammirati, M., Danley, D. E., Ekstrom, J. L., Gibbs, E. M., Hynes, T. R., Mathiowetz, A. M., McPherson, R. K., Olson, T. V., Treadway, J. L., and Hoover, D. J. (2000) Human liver glycogen phosphorylase inhibitors bind at a new allosteric site. *Chem. Biol. 7*, 677−682.

(12) Brissonnet, Y., Ortiz Mellet, C., Morandat, S., Garcia Moreno, M. I., Deniaud, D., Matthews, S. E., Vidal, S., Šesták, S., El Kirat, K., and Gouin, S. G. (2013) Topological effects and binding modes operating with multivalent iminosugar-based glycoclusters and mannosidases. *J. Am. Chem. Soc. 135*, 18427−18435.

(13) Compain, P., Decroocq, C., Joosten, A., de Sousa, J., Rodríguez-Lucena, D., Butters, T. D., Bertrand, J., Clément, R., Boinot, C., Becq, F., and Norez, C. (2013) Rescue of functional CFTR channels in cystic fibrosis: a dramatic multivalent effect using iminosugar cluster-based correctors. *ChemBioChem 14*, 2050−2058.

(14) For a review see: Gouin, S. G. (2014) Multivalent inhibitors for carbohydrate-processing enzymes: beyond the "lock-and-key" concept. *Chem.—Eur. J. 20*, 11616−11628.

(15) For a review see: Compain, P., and Bodlenner, A. (2014) The multivalent effect in glycosidase inhibition: a new, rapidly emerging topic in glycoscience. *ChemBioChem 15*, 1239−1251.

(16) Rísquez-Cuadro, R., García Fernández, J. M., Nierengarten, J.-F., and Ortiz Mellet, C. (2013) Fullerene-sp2-iminosugar balls as multimodal ligands for lectins and glycosidases: a mechanistic hypothesis for the inhibitory multivalent effect. *Chem.—Eur. J. 19*, 16791−16803.

(17) Decroocq, C., Joosten, A., Sergent, R., Mena Barragán, T., Ortiz Mellet, C., and Compain, P. (2013) The multivalent effect in glycosidase inhibition: probing the influence of valency, peripheral ligand structure, and topology with cyclodextrin-based iminosugar click clusters. *ChemBioChem 15*, 2038−2049.

(18) Sánchez-Fernández, E. M., Rísquez-Cuadro, R., Ortiz Mellet, C., García Fernández, J. M., Nieto, P. M., and Angulo, J. (2012) sp^2-Iminosugar *O*-, *S*-, and *N*-glycosides as conformational mimics of α-linked disaccharides; implications for glycosidase inhibition. *Chem.—Eur. J. 18*, 8527−8539.

(19) Bonduelle, C., Huang, J., Mena-Barragan, T., Ortiz Mellet, C., Decroocq, C., Etame, E., Heise, A., Compain, P., and Lecommandoux, S. (2014) Iminosugar-based glycopolypeptides: glycosidase inhibition with bioinspired glycoprotein analogue micellar self-assemblies. *Chem. Commun. 50*, 3350−3352.

(20) Rankin, J. C., and Jeanes, A. (1954) Evaluation of the periodate oxidation method for structural analysis of dextrans. *J. Am. Chem. Soc. 76*, 4435−4441.

(21) Van Cleve, J. W., Schaefer, W. C., and Rist, C. E. (1956) The structure of NRRL B-512 dextran. methylation studies. *J. Am. Chem. Soc. 78*, 4435−4438.

(22) Fabre, E., Bozonnet, S., Arcache, A., Willemot, R. M., Vignon, M., Monsan, P., and Remaud-Simeon, M. (2005) Role of the two catalytic domains of DSR-E dextransucrase and their involvement in the formation of highly alpha-1,2 branched dextran. *J. Bacteriol. 187*, 296−303.

(23) Brison, Y., Laguerre, S., Lefoulon, F., Morel, S., Monties, N., Potocki-Véronèse, G., Monsan, P., and Remaud-Simeon, M. (2013) Branching pattern of gluco-oligosaccharides and 1.5 kDa dextran grafted by the α-1,2 branching sucrase GBD-CD2. *Carbohydr. Polym. 94*, 567−576.

(24) Grischenko, L. A., Parshina, L. N., Kanitskaya, L. V., Larina, L. I., Novikova, L. N., and Trofimov, B. A. (2013) Propargylation of arabinogalactan with propargyl halides–a facile route to new functionalized biopolymers. *Carbohydr. Res. 376*, 7−14.

DOI: 10.1021/acs.bioconjchem.5b00081
Bioconjugate Chem. 2015, 26, 766−772

(25) Rodios, N. A. (1984) 13C NMR spectra of 1-(α-aroyloxyarylideneamino)-1,2,3-triazoles. Identification of 4,5-unsymmetrically substituted derivatives. *J. Heterocycl. Chem. 21*, 1169−1173.
(26) Sulzenbacher, G., Bignon, C., Nishimura, T., Tarling, C. A., Withers, S. G., Henrissat, B., and Bourne, Y. (2004) Crystal structure of Thermotoga maritima alpha-L-fucosidase. Insights into the catalytic mechanism and the molecular basis for fucosidosis. *J. Biol. Chem. 279*, 13119−13128.
(27) Ladevèze, S., Tarquis, L., Cecchini, D., Bercovici, J., André, I., Topham, C., Morel, S., Laville, E., Monsan, P., Lombard, V., Henrissat, B., and Potocki-Véronèse, G. (2013) Role of glycoside phosphorylases in mannose foraging by human gut bacteria. *J. Biol. Chem. 288*, 32370−32383.
(28) Villalonga, R., Cao, R., and Fragoso, A. (2007) Supramolecular chemistry of cyclodextrins in enzyme technology. *Chem. Rev. 107*, 3088−3116.
(29) Darby, J. F., Landström, J., Roth, C., He, Y., Davies, G. J., and Hubbard, R. E. (2014) Discovery of selective small-molecule activators of a bacterial glycoside hydrolase. *Angew. Chem., Int. Ed. 53*, 13419−13423.

DOI: 10.1021/acs.bioconjchem.5b00081
Bioconjugate Chem. 2015, 26, 766−772

Polymeric iminosugars improve the activity of carbohydrate-processing enzymes

Yoan Brissonnet,[a] Simon Ladévèze,[b] David Tezé,[a] Emeline Fabre,[c] David Deniaud,[a] Franck Daligault, [d] Charles Tellier,[d] Sergej Šesták,[e] Magali Remaud-Simeon,[b] Gabrielle Potocki-Veronese[b] and Sébastien G. Gouin*[a]

[a] Dr Y. Brissonnet, Dr D. Tezé, Dr D. Deniaud, Dr S. Gouin
LUNAM Université, CEISAM, Chimie Et Interdisciplinarité, Synthèse, Analyse, Modélisation, UMR CNRS 6230,UFR des Sciences et des Techniques,2 rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, France
E-mail: sebastien.gouin@univ-nantes.fr

[b] Dr S. Ladévèze, Dr M. Remaud-Simeon, Dr G. Potocki-Veronese
Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, Institut National des Sciences Appliquées, CNRS UMR5504, Institut National de Recherche Agronomique, UMR792, 135 Ave. de Rangueil, 31077 Toulouse, France

[c] Dr E. Fabre
Université Lille 1, Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576, 59650 Villeneuve d'Ascq, France

[d] Dr F. Daligault, Prof. C. Tellier
UFIP, UMR CNRS 6286, Faculté des Sciences et Techniques, Université de Nantes, 2 rue de la Houssinière, F-44322 Nantes, France

[e] Dr S. Šesták
Institute of Chemistry, Center for Glycomics, Slovak Academy of Sciences, Dúbravska cesta 9, 845 38 Bratislava, Slovakia,

Contents

Materials NMR spectra were recorded at room temperature with a Bruker Avance 300 Ultra Shield or eBruker Avance III 400 spectrometer and chemical shifts are reported in parts per million relative to tetramethylsilane or a residual solvent peak peak ($CHCl_3$: ^1H: δ=7.26, ^{13}C: δ=77.2; DMSO-d6: ^1H: δ=2.54, ^{13}C: δ=40.4). Peak multiplicity is reported as: singlet (s), doublet (d), triplet (t), quartet (q), multiplet (m), and broad (br). High resolution mass spectra HRMS where obtained by Electrospray Ionisation (ESI) on a Micromass-Waters Q-TOF Ultima Global or with a Bruker Autoflex III SmartBeam spectrometer (MALDI). Low-resolution mass spectra (MS) were recorded with a Thermo electron DSQ spectrometer. All reagents were purchased from Acros Organics or Aldrich and were used without further purification. Column chromatography was conducted on silica gel Kieselgel SI60 (40-63 μm) from Merck. Reactions requiring anhydrous conditions were performed under argon. Dichloromethane was distilled from calcium hydride under nitrogen prior to use. Microwave experiments were conducted in sealed vials in commercial microwave reactors especially designed for synthetic chemistry. (MultiSYNTH, Milestone). The instrument features a special shaking system that ensures high homogeneity of the reaction mixtures. Optical rotations were measured on a 343 PERKIN ELMER at 20°C in a 1cm cell in the stated solvent; $[\alpha]_D$ values are given in 10^{-1} deg.cm^2 g^{-1} (concentration c given as g/100 mL).

General procedure for propargylation of dextrans (Method A):

A solution of propargyl bromide (0.76 g, 4.92 mmol) in toluene (0.6 mL) was added dropwise to a stirred solution of dextran (0.1 g, 0.615 mmol) in 0.3 mL water, KOH-H_2O (0.64 g, 8.6 mmol) and TEBAC (8.4 mg, 36.9 μmol) were added and the mixture was stirred for 20 min at 10–15 °C. The reaction was further carried out for 48 h at rt. Then acetone (1 mL) was added, the organic layer was separated and precipitated in ethanol (75 mL). The precipitate was filtered on millipore, sequentially washed with water (25 mL) and ethanol (25 mL) then dried under vacuum.

General procedure for 1,3-dipolar cycloadditions (Method B):

Copper sulfate (0.75 eq per alkyne function) and sodium ascorbate (1.5 eq per alkyne function) were added to a solution of the alkyne dextran (mmol of monosaccharide unit) and the azido-derivative (1.4 eq per alkyne function) in dioxane-H_2O (5 mL, 4–1). The mixture was irradiated at 80 °C for 2 x 45 min in a sealed vessel, with addition of 0.25 eq of copper sulfate and 0.5 eq of sodium ascorbate for the second run. Residual copper was removed with EDTA (1,5 eq per alkyne function) for 2 h, and the mixture was poured into a NH_4Cl satd. solution (20 mL) and extracted with ethyl acetate (20 mL). The organic layer was dried ($MgSO_4$), filtered and the solvent removed under reduced pressure. The crude product was dissolved in a small amount of CH_2Cl_2 and the product was precipitated with Et_2O (50mL). The precipitate was collected by filtration, washed with Et_2O (50mL) and precipitate twice from Et_2O.

Exemple:

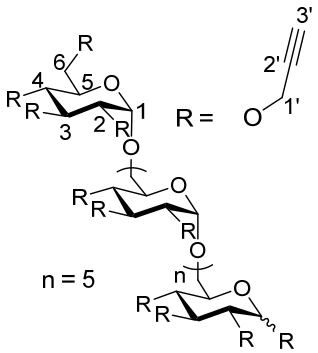
Copper sulfate (40 mg, 0.16 mmol) and sodium ascorbate (65 mg, 0.32 mmol) were added to a solution of **8** (20mg, 72.4 μmol of monosaccharide unit) and **15** (126 mg, 0.30 mmol) in dioxane-H_2O (5 mL, 4–1). The mixture was irradiated at 80 °C for 2 x 45 min in a sealed vessel, with addition of copper sulfate sulfate (13.3 mg, 53.3 μmol) and sodium ascorbate sulfate (21.7 mg, 0.11 mmol) for the second run. Residual copper was removed with EDTA (94 mg, 0.32 mmol) for 2 h, and the mixture was poured into a NH_4Cl satd. solution (20 mL) and extracted with ethyl acetate (20 mL). The organic layer was dried ($MgSO_4$), filtered and the solvent removed under reduced pressure. The crude product was dissolved in a small amount of CH_2Cl_2 and the product was precipitated with Et_2O (30mL). The precipitate was collected by filtration,

washed with Et_2O (50mL) and precipitate twice from Et_2O, to afford **17** (64 mg, 58%) as a yellow solid.

General procedure for acetates deprotection (Method C):

Compound **17** (60 mg, 39.5 μmol of monosaccharide unit) was dissolved in MeOH/H_2O (1:1, 5 mL). Amberlite resin IRN 78 1.25 meq/mL (4 g) was added to the solution, and the mixture was stirred for 48 h at rt. The resin was filtered off and washed with methanol and water. The solvent was evaporated under reduced pressure. The crude product was purified by Sephadex G100 gel filtration, affording **31** (22 mg, 55%) as a pale yellow solid.

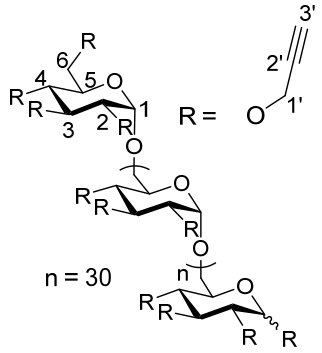
Compound 8:



Obtained following method A. This reaction afforded the compound **8** (107 mg, 63%) as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 5.15-4.93 (m broad, 1H, H-1), 4.54-4.19 (m broad, 6H, H-2, 3, 4, 5, 6), 3.94-3.65 (m broad, 4H, H-1'), 3.62-3.44 (m broad, 2H, H-1'), 2.70-2.35 (m broad, 3.2H, H-3'); FT-IR: (cm^{-1}) 3307 (ν_{C-H}), 2932 (ν_{CH2}), 2120 ($\nu_{C\equiv C}$), 1355 (ν_{C-O}), 1085 (ν_{C-O-C}).

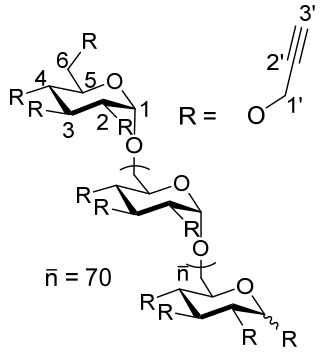
Compound 9:



Obtained following method A. This reaction afforded the compound **9** (116 mg, 68%) as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 5.17-4.94 (m broad, 1H, H-1), 4.58-4.14 (m broad, 6H, H-2, 3, 4, 5, 6), 3.96-3.67 (m broad, 4H, H-1'), 3.64-3.42 (m broad, 2H, H-1'), 2.71-2.42 (m broad, 2.8H, H-3'); FT-IR: (cm^{-1}) 3306 (ν_{C-H}), 2927, 2856 (ν_{CH2}), 2120 ($\nu_{C\equiv C}$), 1358 (ν_{C-O}), 1084 (ν_{C-O-C}).

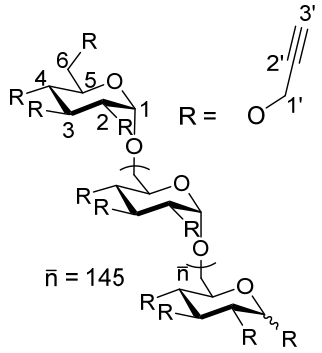
Compound 10:



Obtained following method A. This reaction afforded the compound **10** (114 mg, 67%) as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 5.22-4.90 (m broad, 1H, H-1), 4.66-4.12 (m broad, 6H, H-2, 3, 4, 5, 6), 4.00-3.64 (m broad, 4H, H-1'), 3.64-3.3.37 (m broad, 2H, H-1'), 2.77-2.41 (m broad, 2.94H, H-3'); FT-IR: (cm^{-1}) 3290 (ν_{C-H}), 2929 (ν_{CH2}), 2118 ($\nu_{C\equiv C}$), 1358 (ν_{C-O}), 1084 (ν_{C-O-C}).

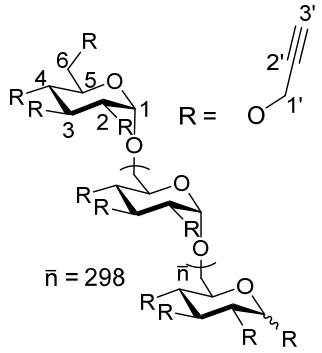
Compound 11:



Obtained following method A. This reaction afforded the compound **11** (133 mg, 78%) as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 5.14-4.92 (m broad, 1H, H-1), 4.55-4.19 (m broad, 6H, H-2, 3, 4, 5, 6), 3.94-3.64 (m broad, 4H, H-1'), 3.63-3.37 (m broad, 2H, H-1'), 2.63-2.40 (m broad, 2.92H, H-3'); FT-IR: (cm^{-1}) 3291 (ν_{C-H}), 2932 (ν_{CH2}), 2118 ($\nu_{C\equiv C}$), 1357 (ν_{C-O}), 1085 (ν_{C-O-C}).

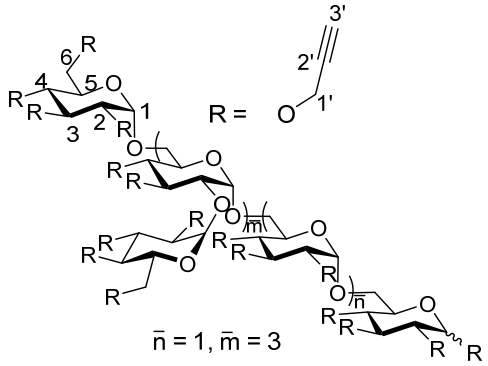
Compound 12:



Obtained following method A. This reaction afforded the compound **12** (135 mg, 79%) as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 5.14-4.92 (m broad, 1H, H-1), 4.55-4.19 (m broad, 6H, H-2, 3, 4, 5, 6), 3.94-3.64 (m broad, 4H, H-1'), 3.63-3.37 (m broad, 2H, H-1'), 2.63-2.40 (m broad, 2.92H, H-3'); FT-IR: (cm^{-1}) 3291 (ν_{C-H}), 2932 (ν_{CH2}), 2118 ($\nu_{C\equiv C}$), 1357 (ν_{C-O}), 1085 (ν_{C-O-C}).

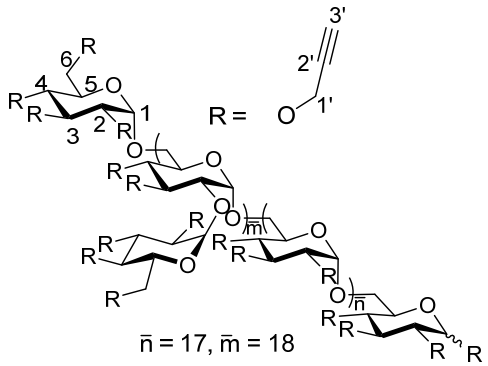
Compound 13:



Obtained following method A. This reaction afforded the compound **13** (111 mg, 65%) as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 5.22-4.93 (m broad, 1H, H-1), 4.60-4.11 (m broad, 6H, H-2, 3, 4, 5, 6), 3.94-3.31 (m broad, 6H, H-1'), 2.70-2.35 (m broad, 3.17H, H-3'); FT-IR: (cm^{-1}) 3293 (ν_{C-H}), 2930 (ν_{CH2}), 2118 ($\nu_{C\equiv C}$), 1357 (ν_{C-O}), 1082 (ν_{C-O-C}), 638 (ν_{C-H}).

Compound 14:



Obtained following method A. This reaction afforded the compound **14** (131 mg, 77%) as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 5.22-4.91 (m broad, 1H, H-1), 4.60-4.09 (m broad, 6H, H-2, 3, 4, 5, 6), 3.99-3.37 (m broad, 6H, H-1'), 2.73-2.37 (m broad, 3.2H, H-3'); FT-IR: (cm^{-1}) 3293 (ν_{C-H}), 2932 (ν_{CH2}), 2119 ($\nu_{C\equiv C}$), 1358 (ν_{C-O}), 1083 (ν_{C-O-C}), 668 (ν_{C-H}).

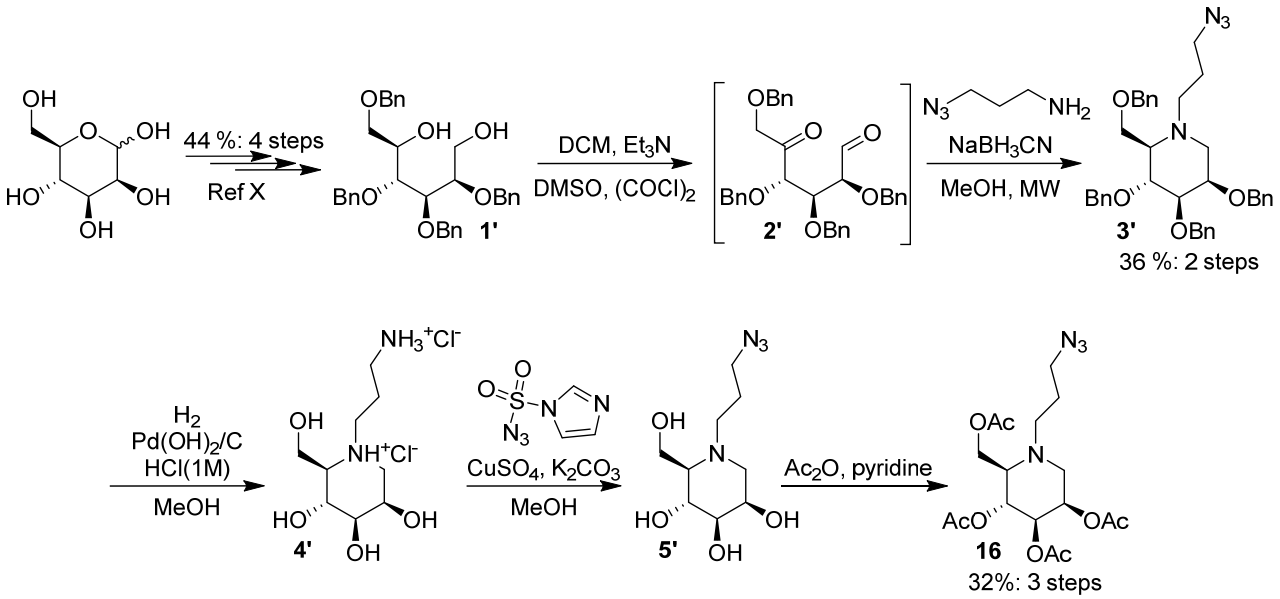
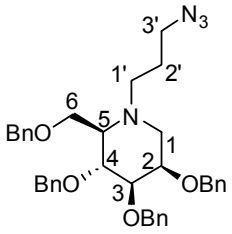


Figure S1: Chemical synthesis of the iminosugar **16**

X : Y. Jiang, Z. Fang, Q.Zheng, H. Jia, J. Cheng, B. Zheng, *Synthesis* **2009**, *16*, 2756-2760.

Compound 3'

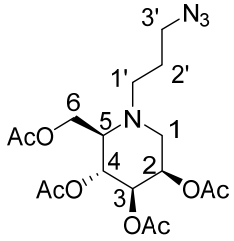


To a solution of DMSO (196 µL, 2.8 mmol) in CH_2Cl_2 (5 mL) at –78 °C was added dropwise under N_2, a solution of oxalyle chloride at 2M in CH_2Cl_2 (1.1 mL, 2.2 mmol) and the mixture was stirred for 45 min at –78 °C. Then a solution of diol **1'** (300 mg, 0.55 mmol) in CH_2Cl_2 (3 mL) was added dropwise; with the temperature of the mixture kept below –78 °C during the addition. The mixture was stirred for 2 h at –78 °C. A solution of Et_3N (0.9 mL, 6.62 mmol) in CH_2Cl_2 (2.0 mL) was then added dropwise and the mixture was warmed to rt. After 30 min, the solvent was removed under vacuum, and the crude residue was dissolved in MeOH (3 mL). 3-azido-1-propylamine (220 mg, 2.2 mmol) was added dropwise at 0°C. The mixture was stirred at 0°C for 30 min. Then, sodium cyanoborohydride (114 mg, 2.2 mmol) was added and the mixture was irradiated for 1 h at 50 °C (300 W) in a sealed vessel. After evaporation of the organic solvent under reduced pressure, the residue was dissolved in DCM (15 mL) and the organic layer was washed with $NaHCO_3$ satd. (2 x 20 mL), dried over $MgSO_4$ and filtered. The solvent was removed under reduced pressure and the residue was purified by flash

chromatography (Cyclohexane/EtOAc: 95/5) to afford **3'** (120 mg, 36% on two steps) as a colorless oil.

$[\alpha]_D^{20}$ = -12 (*c* = 1, $CHCl_3$); ^1H NMR (400 MHz, $CDCl_3$)) δ(ppm): 7.37-7.18 (m, 20H, 20 × Har), 4.73 (d, 1H, *J* = 11.4 Hz, CH2Ph), 4.68-4.38 (m, 7H, CH2Ph), 3.84-3.76 (m, 2H, H-2, H-4), 3.74-3.65 (m, 2H, 2×H-6), 3.58 (dd, 2H, *J2,3* =3. 2 Hz *J3,4* =6.8 Hz, H-3), 3.25 (2H, t, *J3',2'*=6.8 Hz, H-3'), 2.97 (1H, dd, *J1a,1b* =12.2 Hz, *J1a,2* =6.2 Hz, H-1a), 2.88-2.77 (m, 1H, H-1'a),2.75-2.69 (m, 1H, H-5), 2.69-2.60 (m, 1H, H-1'b), 2.34 (dd, 1H, *J1a,1b* =12.2, *J1b,2* =2.2 Hz, H-1b), 1.76-1.59 (m, 2H, 2 × H-2'); ^{13}C NMR (100 MHz, $CDCl_3$) δ(ppm): 138.9, 138.8, 138.7,138.5 (Car), 128.5, 128.4, 128.0, 127.9, 127.8, 127.7,127.6 (CHar), 80.2 (C-3), 76.2 (C-4), 73.8, 73.2 (CH2Ph), 72.4 (C-2), 72.2, 71.3 (CH2Ph), 68.2 (C-6), 62.8 (C-5), 50.8 (C-1'), 50.3 (C-1), 49.7 (C-3'), 26.0 (C-2'); ES-HRMS $[M+H]^+$ m/z 607.3279 requires for $C_{37}H_{42}N_4O_4$; found 607.3277; FT-IR: (ATR, cm^{-1}) 3088 (ν_{C-H}), 3063 (ν_{C-H}), 3030 (ν_{C-H}), 2923 (ν_{C-H}), 2866 (ν_{C-H}), 2094 (ν_{N3}), 1636 ($\nu_{C=C}$), 1493($\nu_{C=C}$), 1453 ($\nu_{C=C}$), 1207(ν_{C-C-N}), 1097 (ν_{C-O-C}), 1070, 1028, 736 (ν_{C-H}), 698 (ν_{C-H}), 669 (ν_{C-H}).

Compound 16

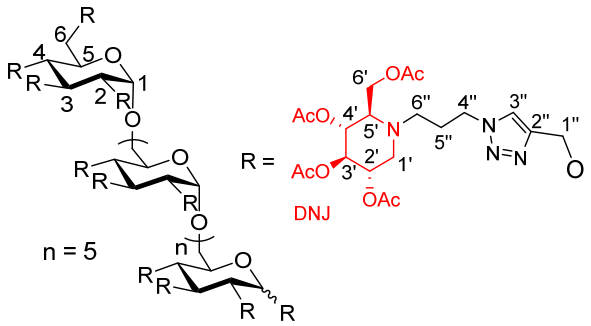


Catalytic hydrogenation of **3'** (2.8 g, 4.66 mmol) was carried out with $Pd(OH)_2$ (3.3 g, 4.66 mmol) in MeOH-1M aq. HCl solution (50 mL, 3:1) at 1 atm for 15 h. Then, the catalyst was filtered over celite, the solvents were removed under reduced pressure. Imidazole-1-sulfonyl azide (1.21 mg, 7.00 mmol) was added to a crude mixture of **4'** (1.2 g, 4.66 mmol) with K_2CO_3 (2.57 g, 18.64 mmol) and $CuSO_4.5H_2O$ (12 mg, 46.6 µmol) in MeOH (45 mL) and the mixture was stirred overnight at rt. The solvent was evaporated under reduced pressure and the crude mixture was dissolved in pyridine-acetic anhydride (1:1, 50 mL) and stirred for 24 h. Water (25 mL) was slowly added at 0°C to quench the reaction. The aqueous phase was extracted with CH_2Cl_2 (2 x 25 mL). The combined organic layers were washed with an aqueous solution of HCl (1M, 2 x 25 mL) and a saturated aqueous $NaHCO_3$ solution (1 x 25 mL). The mixture was dried over $MgSO_4$, filtered and concentrated under reduced pressure. The residue was purified

by flash chromatography (Cyclohexane/EtOAc: 7/3) affording **16** as a yellow oil (620 mg, 32 % for three steps).

$[\alpha]_D^{20}$ = -42 (c = 1, $CHCl_3$); ^1H NMR (300 MHz, $CDCl_3$) δ(ppm): 5.27-5.23 (m, 1H, H-2), 5.20 (t, 1H, H-4), 5.02 (dd, *J* = 3.5 Hz, *J* = 7.8 Hz, 1H, H-3), 4.32-4.18 (m, 2H, H-6), 3.40-3.27 (m, 2H, H-3'), 3.05 (dd, *J* = 5.9 Hz, *J* = 13.1 Hz, 1H, H-1_a), 2.92-2.82 (m, 1H, H-$1'_a$), 2.71-2.63 (m, 1H, H-$1'_b$), 2.60 (dd, *J* = 2.4 Hz, *J* = 13.1 Hz, 1H, H-1_b), 2.09, 2.08, 2.06, 2.03 (4s, 12H, C=OC*H3*), 1.74-1.65 (m, 2H, H-2'); ^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm): 170.9, 170.5, 170.0, 169.8 (C=O), 71.1 (C-3), 68.3 (C-4), 67.0 (C-2), 61.5 (C-5), 60.2 (C-6), 49.9 (C-1), 49.8 (C-1'), 49.1 (C-3'), 25.9 (C-2'), 21.1, 21.0, 20.9, 20.8 (CH_3); ES-HRMS $[M+H]^+$ m/z 415.1823 requires for $C_{17}H_{27}N_4O_8$; found 415.1823; FT-IR: (ATR, cm^{-1}) 2940 (ν_{C-H}), 2821 (ν_{C-H}), 2099 (ν_{N3}), 1743 ($\nu_{C=O}$), 1457, 1372 ($\nu_{C=C}$), 1231 (ν_{C-C-N}), 1050 (ν_{C-O-C}), 737, 668 (ν_{C-H}).

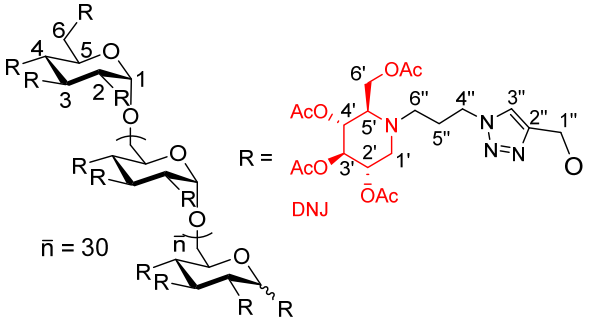
Compound 17:



Obtained following method B. This reaction afforded the compound **17** (64 mg, 58%) as a yellow solid.

^1H NMR (500 MHz, $CDCl_3$) *δ*(ppm): 8.38-7.56 (m broad, 3H, H-3''), 5.20-4.51 (m broad, 16H, H-1,2',3',4',1''), 4.48-4.26 (m broad, 6H, H-4''), 4.21-4.04 (m broad, 6H, H-6'), 3.89-3.48 (m broad, 6H, H-2,3,4,5,6), 3.28-3.11 (m broad, 3H, H-$1'_A$), 2.98-2.78 (m broad, 3H, H-$6''_A$), 2.75-2.48 (m broad, 6H, H-5',$6''_B$), 2.39-2.23 (m broad, 3H, H-$1'_B$), 2.18-1.80 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) *δ*(ppm): 171.7, 170.3, 170.0, 169.8 (C=O), 145.1 (C-2''), 123.9 (C-3''), 97.1 (C-1), 81.3 (C-3), 80.1 (C-2), 76.4 (C-4), 74.6 (C-3'), 71.3 (C-5), 69.5 (C-4'), 69.2 (C-2'), 66.0-64.0 (C-6,1''), 62.0 (C-5'), 59.8 (C-6'), 52.9 (C-1'), 48.7 (C-6''), 48.1 (C-4''), 26.9 (C-5''), 21.0, 20.8 (CH_3).

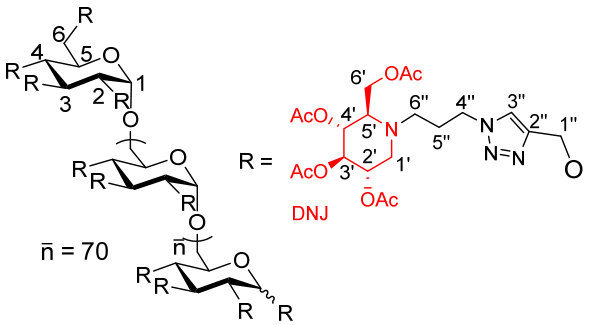
Compound 18:



Obtained following method B. This reaction afforded the compound **18** (61 mg, 55%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.20-7.56 (m broad, 3H, H-3''), 5.13-4.59 (m broad, 16H, H-1,2',3',4',1''), 4.50-4.27 (m broad, 6H, H-4''), 4.17-4.06 (m broad, 6H, H-6'), 3.91-3.55 (m broad, 6H, H-2,3,4,5,6), 3.27-3.11 (m broad, 3H, H-1'$_A$), 2.95-2.79 (m broad, 3H, H-6''$_A$), 2.75-2.49 (m broad, 6H, H-5',6''$_B$), 2.41-2.24 (m broad, 3H, H-1'$_B$), 2.19-1.83 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 171.6, 170.2, 170.0, 169.7 (C=O), 145.1 (C-2''), 123.7 (C-3''), 96.7 (C-1), 79.9 (C-3), 79.0 (C-2), 76.4 (C-4), 74.4 (C-3'), 70.6 (C-5), 69.4 (C-4'), 69.2 (C-2'), 66.1-64.1 (C-6,1''), 62.0 (C-5'), 59.8 (C-6'), 52.8 (C-1'), 48.6 (C-6''), 48.0 (C-4''), 27.0 (C-5''), 21.1, 20.9, 20.8 (CH_3).

Compound 19:

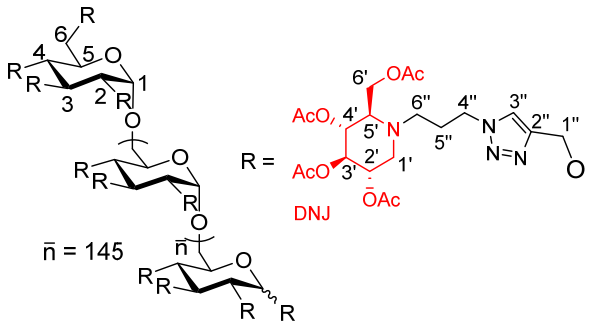


Obtained following method B. This reaction afforded the compound **19** (73 mg, 66%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.30-7.51 (m broad, 3H, H-3''), 5.08-4.56 (m broad, 16H, H-1,2',3',4',1''), 4.45-4.27 (m broad, 6H, H-4''), 4.18-4.04 (m broad, 6H, H-6'), 3.86-3.30 (m broad, 6H, H-2,3,4,5,6), 3.26-3.10 (m broad, 3H, H-1'$_A$), 2.96-2.81 (m broad, 3H, H-6''$_A$), 2.75-2.49 (m broad, 6H, H-5',6''$_B$), 2.41-2.23 (m broad, 3H, H-1'$_B$), 2.16-1.86 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 171.7, 170.3, 170.0, 169.8 (C=O), 145.1

(C-2’’), 123.9 (C-3’’), 97.1 (C-1), 81.3 (C-3), 79.9 (C-2), 76.4 (C-4), 74.6 (C-3’), 70.9 (C-5), 69.4 (C-4’), 69.2(C-2’), 66.0-64.0 (C-6,1’’), 62.0 (C-5’), 59.8 (C-6’), 52.8 (C-1’), 48.7 (C-6’’), 48.0 (C-4’’), 26.9 (C-5’’), 20.9, 20.8 (CH_3).

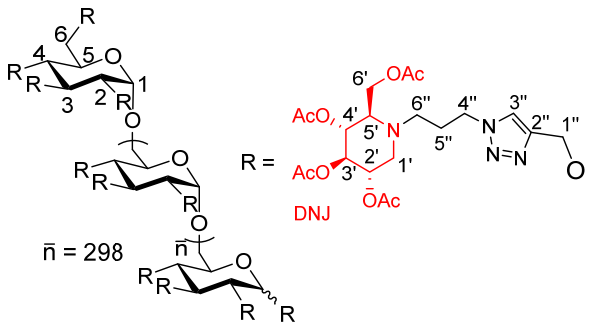
Compound 20:



Obtained following method B. This reaction afforded the compound **20** (71 mg, 64%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.27-7.68 (m broad, 3H, H-3’’), 5.16-4.55 (m broad, 16H, H-1,2’,3’,4’,1’’), 4.53-4.22 (m broad, 6H, H-4’’), 4.20-3.96 (m broad, 6H, H-6’), 3.85-3.29 (m broad, 6H, H-2,3,4,5,6), 3.27-3.07 (m broad, 3H, H-1’$_A$), 2.98-2.77 (m broad, 3H, H-6’’$_A$), 2.75-2.48 (m broad, 6H, H-5’,6’’$_B$), 2.40-2.22 (m broad, 3H, H-1’$_B$), 2.16-1.60 (m broad, 42H, H-5’’,O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 171.7, 170.3, 170.0, 169.8 (C=O), 145.0 (C-2’’), 125.4 (C-3’’), 97.6 (C-1), 81.3 (C-3), 79.9 (C-2), 76.4 (C-4), 74.6 (C-3’), 70.6 (C-5), 69.5 (C-4’), 69.2(C-2’), 66.3-64.8 (C-6,1’’), 62.0 (C-5’), 59.8 (C-6’), 52.8 (C-1’), 48.7 (C-6’’), 48.0 (C-4’’), 26.9 (C-5’’), 20.9, 20.8 (CH_3).

Compound 21:

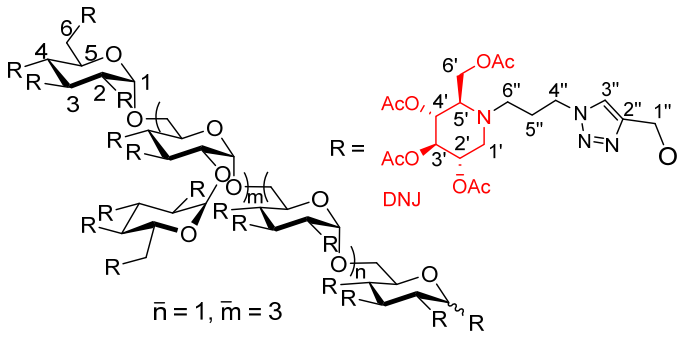


Obtained following method B. This reaction afforded the compound **21** (55 mg, 50%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.20-7.65 (m broad, 3H, H-3’’), 5.16-4.55 (m broad, 16H,

H-1,2',3',4',1''), 4.52-4.27 (m broad, 6H, H-4''), 4.24-4.03 (m broad, 6H, H-6'), 3.90-3.36 (m broad, 6H, H-2,3,4,5,6), 3.28-3.04 (m broad, 3H, H-1'$_A$), 2.97-2.79 (m broad, 3H, H-6''$_A$), 2.76-2.48 (m broad, 6H, H-5',6''$_B$), 2.41-2.21 (m broad, 3H, H-1'$_B$), 2.15-1.72 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 171.8, 170.3, 169.8 (C=O), 145.2 (C-2''), 123.4 (C-3''), 96.8 (C-1), 81.5 (C-3), 79.7 (C-2), 76.4 (C-4), 74.5 (C-3'), 70.8 (C-5), 69.5 (C-4'), 69.2(C-2'), 65.9-63.7 (C-6,1''), 62.1 (C-5'), 59.8 (C-6'), 52.8 (C-1'), 48.6 (C-6''), 48.1 (C-4''), 26.9 (C-5''), 21.0, 20.8 (CH_3).

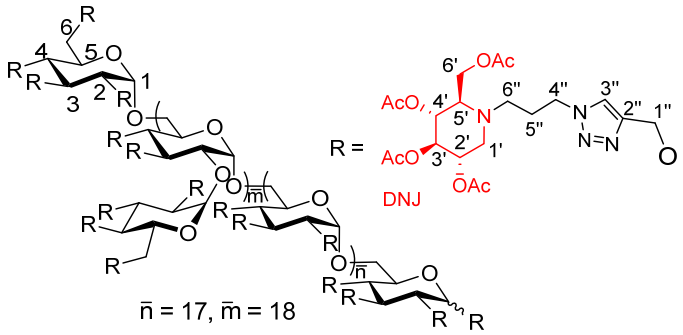
Compound 22:



Obtained following method B. This reaction afforded the compound **22** (61 mg, 55%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.32-7.56 (m broad, 3H, H-3''), 5.22-4.55 (m broad, 16H, H-1,2',3',4',1''), 4.55-4.26 (m broad, 6H, H-4''), 4.22-4.04 (m broad, 6H, H-6'), 3.89-3.48 (m broad, 6H, H-2,3,4,5,6), 3.28-3.09 (m broad, 3H, H-1'$_A$), 3.00-2.81 (m broad, 3H, H-6''$_A$), 2.77-2.53 (m broad, 6H, H-5',6''$_B$), 2.43-2.27 (m broad, 3H, H-1'$_B$), 2.20-1.81 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 171.7, 170.3, 170.0, 169.8 (C=O), 145.0 (C-2''), 124.2 (C-3''), 96.7 (C-1), 81.2 (C-3), 79.1 (C-2), 76.4 (C-4), 74.6 (C-3'), 70.2 (C-5), 69.5 (C-4'), 69.2 (C-2'), 66.0 (C-6,1''), 62.0 (C-5'), 59.8 (C-6'), 52.9 (C-1'), 48.7 (C-6''), 48.0 (C-4''), 26.9 (C-5''), 21.1, 21.0, 20.8 (CH_3).

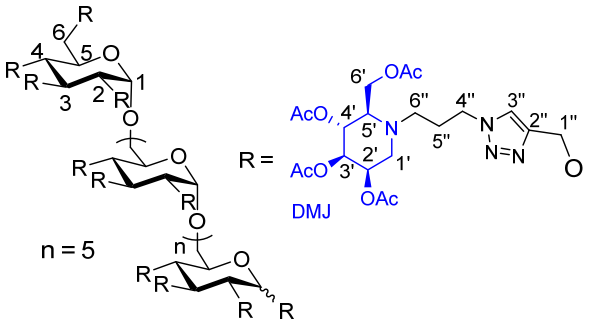
Compound 23:



Obtained following method B. This reaction afforded the compound **23** (57 mg, 52%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.33-7.67 (m broad, 3H, H-3''), 5.20-4.59 (m broad, 16H, H-1,2',3',4',1''), 4.52-4.25 (m broad, 6H, H-4''), 4.22-4.01 (m broad, 6H, H-6'), 3.88-3.39 (m broad, 6H, H-2,3,4,5,6), 3.25-3.12 (m broad, 3H, H-1'$_A$), 3.01-2.97 (m broad, 3H, H-6''$_A$), 2.74-2.55 (m broad, 6H, H-5',6''$_B$), 2.41-2.25 (m broad, 3H, H-1'$_B$), 2.14-1.86 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 171.7, 170.3, 170.0, 169.8 (C=O), 144.9 (C-2''), 124.2 (C-3''), 95.2 (C-1), 81.3-80.0 (C-3,2), 76.3 (C-4), 74.5 (C-3'), 69.4 (C-5, 4'), 69.2 (C-2'), 66.3-66.0 (C-6,1''), 62.1 (C-5'), 59.8 (C-6'), 52.9 (C-1'), 48.6 (C-6''), 48.0 (C-4''), 26.8 (C-5''), 20.9, 20.8, (CH_3).

Compound 24:

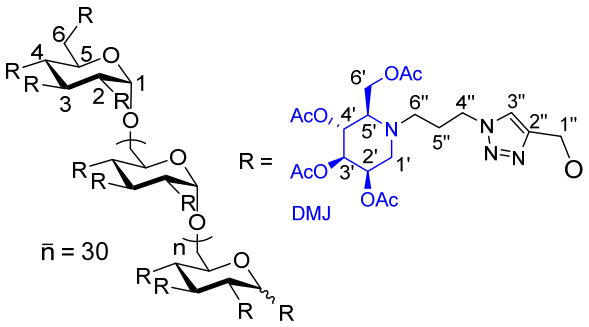


Obtained following method B. This reaction afforded the compound **24** (32 mg, 58%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.25-7.70 (m broad, 3H, H-3''), 5.35-5.10 (m broad, 7H, H-1,2',4'), 5.07-4.94 (m broad, 3H, H-3'), 4.94-4.60 (m broad, 6H, H-1''), 4.52-4.29 (m broad, 6H, H-4''), 4.29-4.14 (m broad, 6H, H-6'), 3.87-3.41 (m broad, 6H, H-2, 3, 4, 5, 6), 3.15-3.00 (m broad, 3H, H-1'$_A$), 2.95-2.77 (m broad, 6H, H-1'$_B$,5'), 2.76-2.55 (m broad, 6H, H-6''), 2.23-1.76 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 170.8, 170.4, 170.0,

169.8 (C=O), 145.1 (C-2’’), 123.8 (C-3’’), 97.1 (C-1), 81.5 (C-3), 80.2 (C-2), 76.7 (C-4), 71.4 (C-3’), 70.5 (C-5,6), 68.2 (C-4’), 67.0 (C-2’), 66.5-65.2 (C-1’’), 61.3 (C-5’), 60.1 (C-6’), 50.0 (C-1’,6’’), 47.9 (C-4’’), 27.3 (C-5’’), 21.2, 21.0, 20.9 (CH_3).

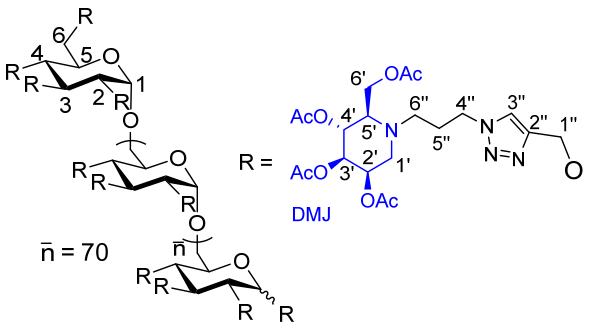
Compound 25:



Obtained following method B. This reaction afforded the compound **25** (30 mg, 54%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.26-7.65 (m broad, 3H, H-3’’), 5.38-5.14 (m broad, 7H, H-1,2’,4’), 5.09-4.98 (m broad, 3H, H-3’), 4.92-4.65 (m broad, 6H, H-1’’), 4.53-4.34 (m broad, 6H, H-4’’), 4.32-4.15 (m broad, 6H, H-6’), 3.86-3.47 (m broad, 6H, H-2, 3, 4, 5, 6), 3.18-3.02 (m broad, 3H, H-1’$_A$), 2.99-2.79 (m broad, 6H, H-1’$_B$,5’), 2.77-2.52 (m broad, 6H, H-6’’), 2.38-1.78 (m broad, 42H, H-5’’,O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 170.9, 170.5, 170.0, 169.9 (C=O), 145.2 (C-2’’), 124.2 (C-3’’), 97.3 (C-1), 81.5 (C-3), 80.1 (C-2), 76.6 (C-4), 71.1 (C-3’), 70.7 (C-5,6), 68.0 (C-4’), 67.0 (C-2’), 66.3-65.5 (C-1’’), 61.4 (C-5’), 60.0 (C-6’), 50.0 (C-1’,6’’), 47.9 (C-4’’), 27.2 (C-5’’), 21.2, 21.0, 20.9 (CH_3).

Compound 26:

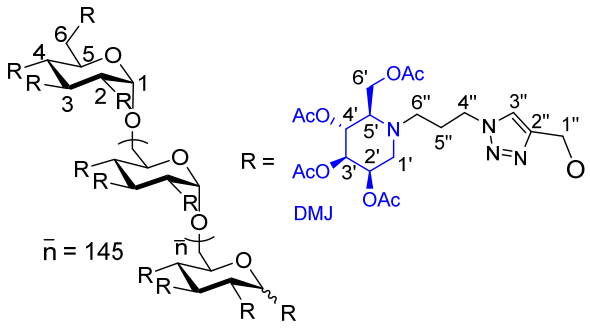


Obtained following method B. This reaction afforded the compound **26** (27 mg, 49%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.19-7.72 (m broad, 3H, H-3’’), 5.33-5.13 (m broad, 7H,

H-1,2',4'), 5.07-4.97 (m broad, 3H, H-3'), 4.89-4.60 (m broad, 6H, H-1''), 4.52-4.31 (m broad, 6H, H-4''), 4.30-4.15 (m broad, 6H, H-6'), 3.87-3.41 (m broad, 6H, H-2, 3, 4, 5, 6), 3.17-3.02 (m broad, 3H, H-1'$_A$), 2.95-2.77 (m broad, 6H, H-1'$_B$,5'), 2.77-2.53 (m broad, 6H, H-6''), 2.24-1.77 (m broad, 42H, H-5'',O=CH$_3$); ^{13}C NMR (125 MHz, CDCl$_3$) δ(ppm): 170.8, 170.5, 170.0, 169.8 (C=O), 145.0 (C-2''), 123.9 (C-3''), 97.3 (C-1), 81.2 (C-3), 80.1 (C-2), 76.5 (C-4), 71.2 (C-3'), 70.7 (C-5,6), 68.1 (C-4'), 67.0 (C-2'), 66.3-6402 (C-1''), 61.3 (C-5'), 60.1 (C-6'), 49.7 (C-1',6''), 47.9 (C-4''), 27.3 (C-5''), 21.2, 21.0, 20.9 (CH$_3$).

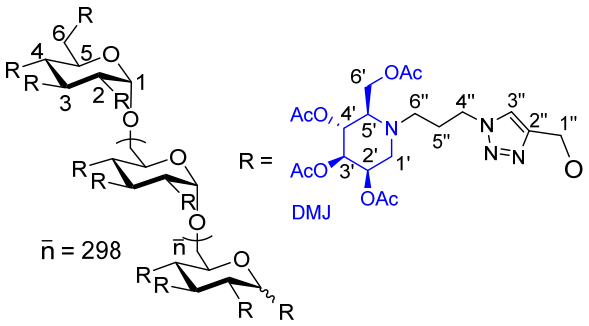
Compound 27:



Obtained following method B. This reaction afforded the compound **27** (32 mg, 58%) as a yellow solid.

^{1}H NMR (500 MHz, CDCl$_3$) δ(ppm): 8.21-7.68 (m broad, 3H, H-3''), 5.27-5.10 (m broad, 7H, H-1,2',4'), 5.04-4.93 (m broad, 3H, H-3'), 4.89-4.53 (m broad, 6H, H-1''), 4.50-4.28 (m broad, 6H, H-4''), 4.27-4.13 (m broad, 6H, H-6'), 3.80-3.35 (m broad, 6H, H-2, 3, 4, 5, 6), 3.14-2.98 (m broad, 3H, H-1'$_A$), 2.91-2.74 (m broad, 6H, H-1'$_B$,5'), 2.74-2.56 (m broad, 6H, H-6''), 2.13-1.85 (m broad, 42H, H-5'',O=CH$_3$); ^{13}C NMR (125 MHz, CDCl$_3$) δ(ppm): 170.8, 170.4, 169.9, 169.8 (C=O), 145.0 (C-2''), 123.9 (C-3''), 97.7 (C-1), 81.4 (C-3), 79.9 (C-2), 76.5 (C-4), 71.3 (C-3'), 70.7 (C-5,6), 68.2 (C-4'), 67.0 (C-2'), 66.4 (C-1''), 61.3 (C-5'), 60.1 (C-6'), 49.6 (C-1',6''), 47.9 (C-4''), 27.2 (C-5''), 21.1, 21.0, 20.8 (CH$_3$).

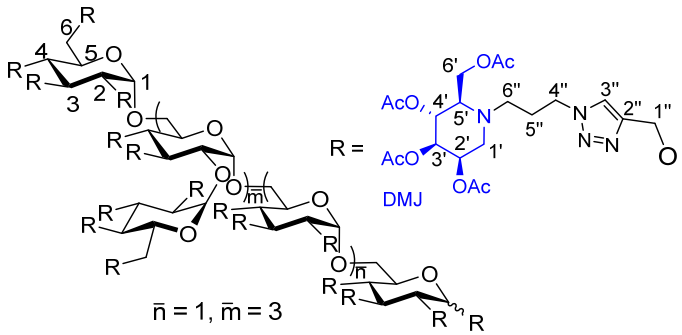
Compound 28:



Obtained following method B. This reaction afforded the compound **28** (25 mg, 46%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.10-7.70 (m broad, 3H, H-3''), 5.29-5.14 (m broad, 7H, H-1,2',4'), 5.07-4.97 (m broad, 3H, H-3'), 4.91-4.63 (m broad, 6H, H-1''), 4.50-4.31 (m broad, 6H, H-4''), 4.30-4.16 (m broad, 6H, H-6'), 3.84-3.41 (m broad, 6H, H-2, 3, 4, 5, 6), 3.14-3.02 (m broad, 3H, H-1'$_A$), 2.92-2.76 (m broad, 6H, H-1'$_B$,5'), 2.77-2.59 (m broad, 6H, H-6''), 2.38-1.80 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 170.9, 170.5, 170.0, 169.8 (C=O), 145.2 (C-2''), 123.7 (C-3''), 97.0 (C-1), 81.5 (C-3), 80.2 (C-2), 76.5 (C-4), 71.1 (C-3'), 70.7 (C-5,6), 68.2 (C-4'), 67.0 (C-2'), 66.3-65.0 (C-1''), 61.3 (C-5'), 60.1 (C-6'), 49.8 (C-1',6''), 47.9 (C-4''), 27.3 (C-5''), 21.2, 21.0, 20.9 (CH_3).

Compound 29:

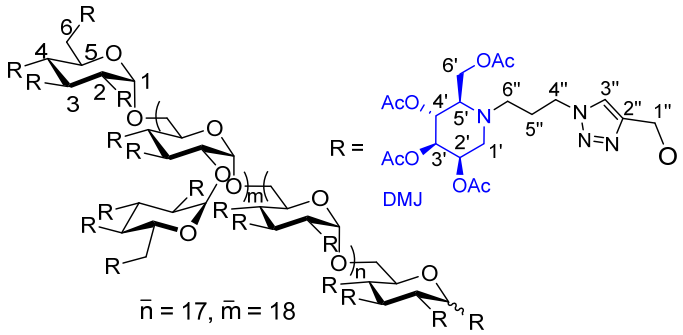


Obtained following method B. This reaction afforded the compound **29** (27 mg, 49%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.25-7.70 (m broad, 3H, H-3''), 5.30-5.10 (m broad, 7H, H-1,2',4'), 5.03-4.94 (m broad, 3H, H-3'), 4.92-4.59 (m broad, 6H, H-1''), 4.51-4.28 (m broad, 6H, H-4''), 4.26-4.13 (m broad, 6H, H-6'), 3.82-3.37 (m broad, 6H, H-2, 3, 4, 5, 6), 3.12-2.99 (m broad, 3H, H-1'$_A$), 2.92-2.74 (m broad, 6H, H-1'$_B$,5'), 2.76-2.56 (m broad, 6H, H-6''), 2.23-1.85 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 171.9, 170.5, 170.0,

169.9 (C=O), 145.0 (C-2''), 124.2 (C-3''), 97.1 (C-1), 81.4 (C-3), 79.7 (C-2), 76.6 (C-4), 71.2 (C-3'), 70.5 (C-5,6), 68.1 (C-4'), 67.0 (C-2'), 66.2-63.8 (C-1''), 61.2 (C-5'), 60.0 (C-6'), 50.0 (C-1',6''), 47.9 (C-4''), 27.0 (C-5''), 21.2, 21.0, 20.9 (CH_3).

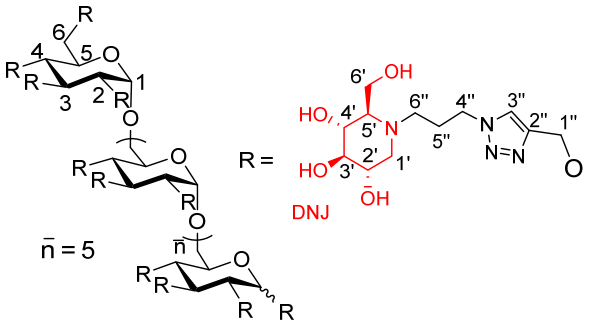
Compound 30:



Obtained following method B. This reaction afforded the compound **30** (25 mg, 44%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ(ppm): 8.18-7.71 (m broad, 3H, H-3''), 5.31-5.08 (m broad, 7H, H-1,2',4'), 5.05-4.95 (m broad, 3H, H-3'), 4.89-4.54 (m broad, 6H, H-1''), 4.53-4.31 (m broad, 6H, H-4''), 4.30-4.10 (m broad, 6H, H-6'), 3.93-3.37 (m broad, 6H, H-2, 3, 4, 5, 6), 3.16-3.00 (m broad, 3H, H-1'$_A$), 2.95-2.76 (m broad, 6H, H-1'$_B$,5'), 2.76-2.55 (m broad, 6H, H-6''), 2.23-1.73 (m broad, 42H, H-5'',O=CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ(ppm): 170.7, 170.3, 169.8, 169.6 (C=O), 145.7 (C-2''), 124.0 (C-3''), 96.7 (C-1), 81.2 (C-3), 80.1 (C-2), 76.5 (C-4), 71.1 (C-3'), 70.9 (C-5,6), 68.1 (C-4'), 66.9 (C-2'), 66.4 (C-1''), 61.1 (C-5'), 60.0 (C-6'), 49.8 (C-1',6''), 47.9 (C-4''), 27.0 (C-5''), 21.1, 20.9, 20.8, 20.7 (CH_3).

Compound 31:

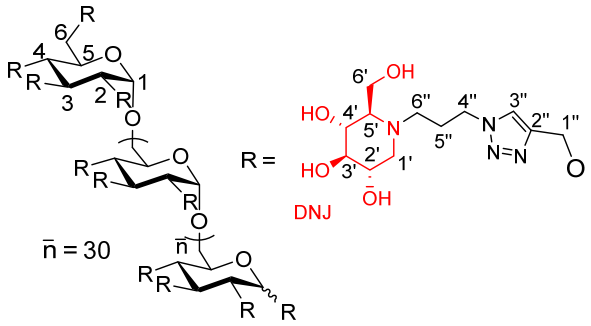


Obtained following method C. This reaction afforded the compound **31** (22 mg, 55%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.20-7.95 (m broad, 3H, H-3''), 5.28-5.10 (m broad, 1H, H-1), 5.03-4.70 (m broad, 6H, H-1''), 4.56-4.30 (m broad, 6H, H-4''), 4.03-3.67 (m broad, 11H,

H-3, 4, 5, 6, 6'), 3.62-3.49 (m broad, 4H, H-2, 2'), 3.49-3.36 (m broad, 3H, H-4'), 3.35-3.23 (m broad, 3H, H-3'), 3.07-2.91 (m broad, 3H, H-1'$_A$), 2.91-2.74 (m broad, 3H, H-6''$_A$), 2.74-2.57 (m broad, 3H, H-6''$_B$), 2.38-2.18 (m broad, 6H, H-1'$_B$, 5'), 2.18-1.94 (m broad, 6H, H-5''); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.4 (C-2''), 125.0 (C-3''), 96.5 (C-1), 80.7 (C-3), 79.3 (C-2), 78.5 (C-3'), 77.0 (C-4), 70.1 (C-5, 4'), 69.0 (C-2'), 65.7 (C-6), 65.0 (C-5'), 63.3 (C-1''), 57.7 (C-6'), 55.6 (C-1'), 48.9 (C-6''), 48.7 (C-4''), 24.2 (C-5'').

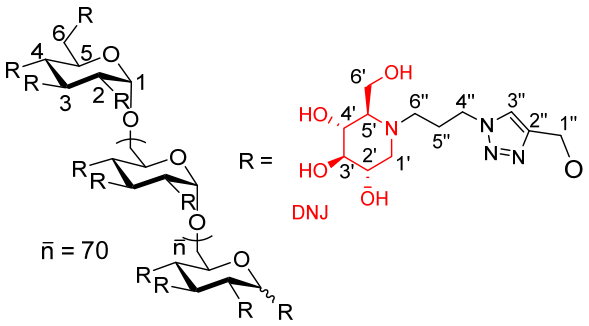
Compound 32:



Obtained following method C. This reaction afforded the compound **32** (25 mg, 53%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.20-7.98 (m broad, 3H, H-3''), 5.25-5.07 (m broad, 1H, H-1), 5.03-4.81 (m broad, 6H, H-1''), 4.49-4.36 (m broad, 6H, H-4''), 4.05-3.69 (m broad, 11H, H-3, 4, 5, 6, 6'), 3.64-3.52 (m broad, 4H, H-2, 2'), 3.49-3.36 (m broad, 3H, H-4'), 3.35-3.26 (m broad, 3H, H-3'), 3.07-2.95 (m broad, 3H, H-1'$_A$), 2.91-2.76 (m broad, 3H, H-6''$_A$), 2.76-2.60 (m broad, 3H, H-6''$_B$), 2.41-2.21 (m broad, 6H, H-1'$_B$, 5'), 2.21-2.00 (m broad, 6H, H-5''); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.2 (C-2''), 125.1 (C-3''), 96.0 (C-1), 80.8 (C-3), 79.1 (C-2), 78.4 (C-3'), 77.4 (C-4), 72.4 (C-5), 70.1 (C-4'), 69.0 (C-2'), 65.8 (C-6), 65.0 (C-5'), 63.3 (C-1''), 57.7 (C-6'), 55.5 (C-1'), 48.9 (C-6''), 48.7 (C-4''), 24.1 (C-5'').

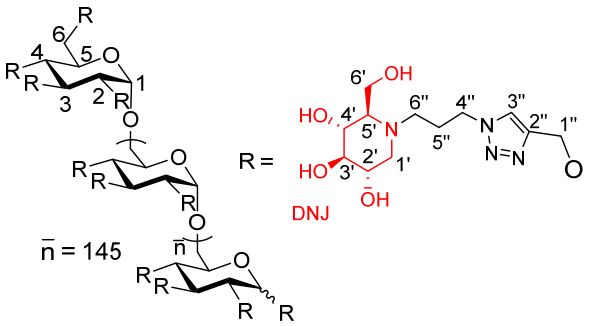
Compound 33:



Obtained following method C. This reaction afforded the compound **33** (24 mg, 60%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.23-7.98 (m broad, 3H, H-3''), 5.30-5.07 (m broad, 1H, H-1), 5.04-4.75 (m broad, 6H, H-1''), 4.57-4.28 (m broad, 6H, H-4''), 4.05-3.67 (m broad, 11H, H-3, 4, 5, 6, 6'), 3.67-3.47 (m broad, 4H, H-2, 2'), 3.46-3.3.34 (m broad, 3H, H-4'), 3.34-3.23 (m broad, 3H, H-3'), 3.05-2.93 (m broad, 3H, H-1'$_A$), 2.90-2.74 (m broad, 3H, H-6''$_A$), 2.74-2.55 (m broad, 3H, H-6''$_B$), 2.37-2.20 (m broad, 6H, H-1'$_B$, 5'), 2.20-2.00 (m broad, 6H, H-5''); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.2 (C-2''), 125.0 (C-3''), 96.2 (C-1), 80.7 (C-3), 79.1 (C-2), 78.4 (C-3'), 77.1 (C-4), 72.4 (C-5), 70.1 (C-4'), 69.0 (C-2'), 65.6 (C-6), 65.0 (C-5'), 63.1 (C-1''), 57.7 (C-6'), 55.5 (C-1'), 48.9 (C-6''), 48.7 (C-4''), 24.2 (C-5'').

Compound 34:

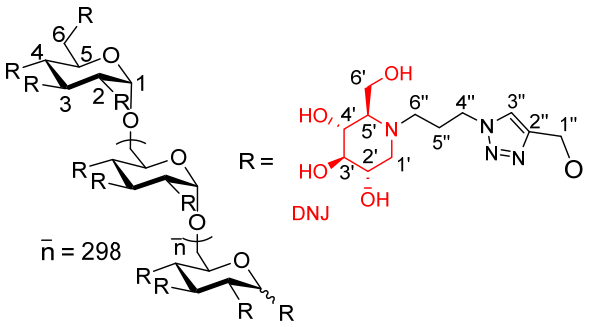


Obtained following method C. This reaction afforded the compound **34** (20 mg, 50%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.20-7.99 (m broad, 3H, H-3''), 5.27-5.16 (m broad, 1H, H-1), 5.04-4.81 (m broad, 6H, H-1''), 4.60-4.28 (m broad, 6H, H-4''), 4.05-3.66 (m broad, 11H, H-3, 4, 5, 6, 6'), 3.66-3.50 (m broad, 4H, H-2, 2'), 3.46-3.34 (m broad, 3H, H-4'), 3.34-3.23 (m broad, 3H, H-3'), 3.10-2.90 (m broad, 3H, H-1'$_A$), 2.90-2.72 (m broad, 3H, H-6''$_A$), 2.72-2.54 (m broad, 3H, H-6''$_B$), 2.38-2.18 (m broad, 6H, H-1'$_B$, 5'), 2.20-1.95 (m broad, 6H, H-5''); ^{13}C

NMR (125 MHz, D_2O) δ(ppm): 144.4 (C-2''), 125.0 (C-3''), 96.6 (C-1), 80.6 (C-3), 79.1 (C-2), 78.4 (C-3'), 77.0 (C-4), 71.9 (C-5), 70.1 (C-4'), 68.8 (C-2'), 65.6 (C-6), 65.0 (C-5'), 63.2 (C-1''), 57.7 (C-6'), 55.6 (C-1'), 48.9 (C-6''), 48.7 (C-4''), 24.1 (C-5'').

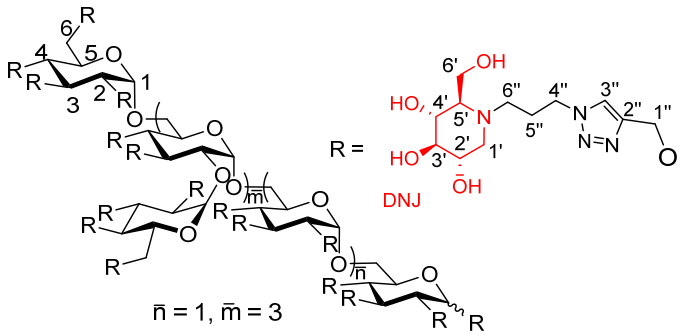
Compound 35:



Obtained following method C. This reaction afforded the compound **35** (25 mg, 62%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.22-7.99 (m broad, 3H, H-3''), 5.27-5.09 (m broad, 1H, H-1), 5.05-4.89 (m broad, 6H, H-1''), 4.58-4.37 (m broad, 6H, H-4''), 4.03-3.70 (m broad, 11H, H-3, 4, 5, 6, 6'), 3.67-3.48 (m broad, 4H, H-2, 2'), 3.47-.3.35 (m broad, 3H, H-4'), 3.35-3.24 (m broad, 3H, H-3'), 3.09-2.95 (m broad, 3H, H-1'$_A$), 2.90-2.76 (m broad, 3H, H-6''$_A$), 2.76-2.61 (m broad, 3H, H-6''$_B$), 2.40-2.22 (m broad, 6H, H-1'$_B$, 5'), 2.20-2.00 (m broad, 6H, H-5''); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.3 (C-2''), 125.2 (C-3''), 96.3 (C-1), 80.6 (C-3), 79.1 (C-2), 78.4 (C-3'), 76.5 (C-4), 72.5 (C-5), 70.2 (C-4'), 69.0 (C-2'), 65.8 (C-6), 65.0 (C-5'), 63.1 (C-1''), 57.7 (C-6'), 55.6 (C-1'), 48.9 (C-6''), 48.7 (C-4''), 24.1 (C-5'').

Compound 36:

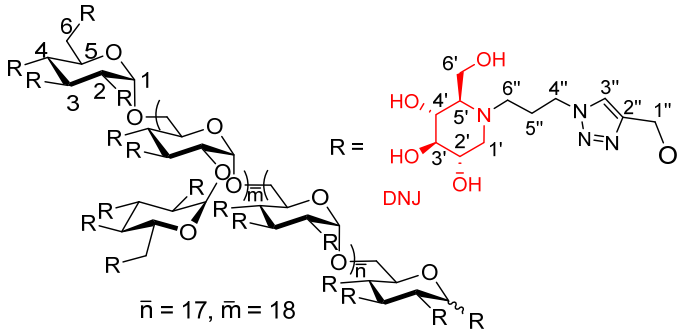


Obtained following method C. This reaction afforded the compound **36** (16 mg, 45%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.19-7.92 (m broad, 3H, H-3''), 5.42-5.22 (m broad, 1H,

H-1), 5.20-4.72 (m broad, 6H, H-1’’), 4.55-4.29 (m broad, 6H, H-4’’), 4.00-3.66 (m broad, 11H, H-3, 4, 5, 6, 6’), 3.66-3.48 (m broad, 4H, H-2, 2’), 3.48-3. 35 (m broad, 3H, H-4’), 3.34-3.21 (m broad, 3H, H-3’), 3.08-2.93 (m broad, 3H, H-1’$_A$), 2.89-2.75 (m broad, 3H, H-6’’$_A$), 2.73-2.52 (m broad, 3H, H-6’’$_B$), 2.38-2.18 (m broad, 6H, H-1’$_B$, 5’), 2.18-1.97 (m broad, 6H, H-5’’); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.3 (C-2’’), 125.0 (C-3’’), 96.4 (C-1), 80.6 (C-3), 79.3 (C-2), 78.3 (C-3’), 76.7 (C-4), 70.0 (C-5, 4’), 68.9 (C-2’), 65.5 (C-6), 64.9 (C-5’), 63.4 (C-1’’), 57.6 (C-6’), 55.5 (C-1’), 48.9 (C-6’’), 48.7 (C-4’’), 24.1 (C-5’’).

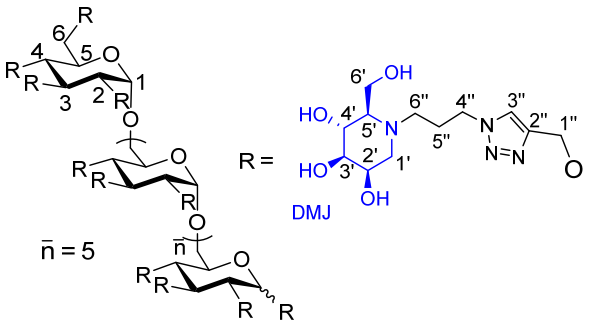
Compound 37:



Obtained following method C. This reaction afforded the compound **37** (22 mg, 46%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.19-7.92 (m broad, 3H, H-3’’), 5.41-5.15 (m broad, 1H, H-1), 5.06-4.65 (m broad, 6H, H-1’’), 4.60-4.27 (m broad, 6H, H-4’’), 3.95-3.66 (m broad, 11H, H-3, 4, 5, 6, 6’), 3.65-3.48 (m broad, 4H, H-2, 2’), 3.48-3. 34 (m broad, 3H, H-4’), 3.34-3.21 (m broad, 3H, H-3’), 3.06-2.90 (m broad, 3H, H-1’$_A$), 2.90-2.72 (m broad, 3H, H-6’’$_A$), 2.72-2.53 (m broad, 3H, H-6’’$_B$), 2.38-2.18 (m broad, 6H, H-1’$_B$, 5’), 2.18-1.92 (m broad, 6H, H-5’’); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.3 (C-2’’), 124.9 (C-3’’), 96.6 (C-1), 80.7 (C-3), 79.3 (C-2), 78.4 (C-3’), 76.9 (C-4), 70.1 (C-5, 4’), 69.0 (C-2’), 65.7 (C-6), 65.0 (C-5’), 63.0 (C-1’’), 57.7 (C-6’), 55.6 (C-1’), 48.9 (C-6’’), 48.7 (C-4’’), 24.3 (C-5’’).

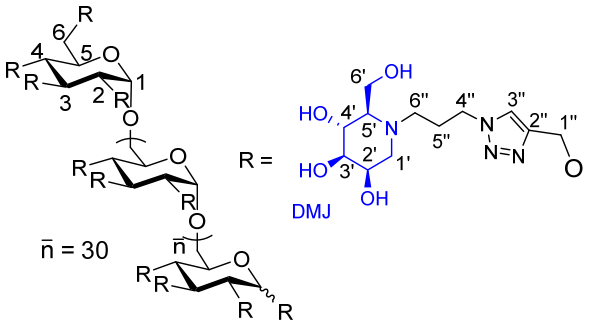
Compound 38:



Obtained following method C. This reaction afforded the compound **38** (12 mg, 75%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.19-8.00 (m broad, 3H, H-3''), 5.26-5.10 (m broad, 1H, H-1), 5.04-4.82 (m broad, 6H, H-1''), 4.56-4.32 (m broad, 6H, H-4''), 4.07-3.61 (m broad, 18H, H-2, 3, 4, 5, 6, 2', 4', 6'), 3.54-3.44 (m broad, 3H, H-3'), 3.04-2.87 (m broad, 3H, H-1'$_A$), 2.87-2.72 (m broad, 3H, H-6''$_A$), 2.69-2.55 (m broad, 3H, H-6''$_B$), 2.55-2.39 (m broad, 3H, H-1'$_B$), 2.33-1.99 (m broad, 9 H, H-5', 5''); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.4 (C-2''), 125.0 (C-3''), 96.4 (C-1), 80.8 (C-3), 79.3 (C-2), 77.0 (C-4), 74.5 (C-3'), 70.1 (C-5), 67.9 (C-2',4'), 65.4 (C-6, 5'), 63.5 (C-1''), 58.0 (C-6'), 54.9 (C-1'), 49.0 (C-6''), 48.7 (C-4''), 24.6 (C-5'').

Compound 39:

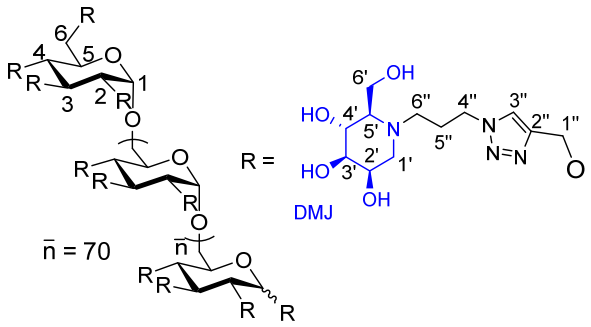


Obtained following method C. This reaction afforded the compound **39** (16 mg, 75%) as a pale yellow solid.

^{1}H NMR (700 MHz, D_2O) δ(ppm): 8.18-7.97 (m broad, 3H, H-3''), 5.24-5.10 (m broad, 1H, H-1), 5.04-4.81 (m broad, 6H, H-1''), 4.55-4.33 (m broad, 6H, H-4''), 4.02-3.57 (m broad, 18H, H-2, 3, 4, 5, 6, 2', 4', 6'), 3.53-3.45 (m broad, 3H, H-3'), 3.04-2.88 (m broad, 3H, H-1'$_A$), 2.88-2.72 (m broad, 3H, H-6''$_A$), 2.71-2.57 (m broad, 3H, H-6''$_B$), 2.57-2.44 (m broad, 3H, H-1'$_B$), 2.27-2.00 (m broad, 9 H, H-5', 5''); ^{13}C NMR (175 MHz, D_2O) δ(ppm): 144.1 (C-2''), 124.9 (C-3''), 96.2 (C-1), 80.8 (C-3), 79.0 (C-2), 77.4 (C-4), 74.4 (C-3'), 69.7 (C-5), 67.8 (C-2',4'),

65.3 (C-6, 5'), 63.2 (C-1''), 57.8 (C-6'), 54.8 (C-1'), 48.9 (C-6''), 48.6 (C-4''), 24.4 (C-5'').

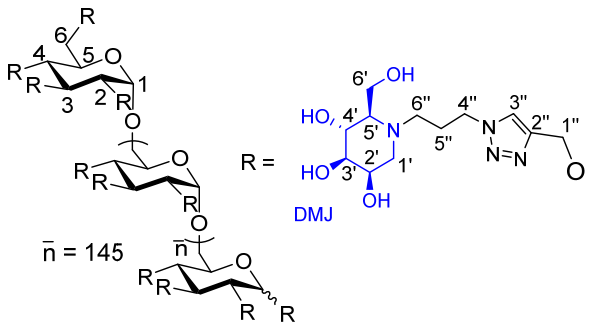
Compound 40:



Obtained following method C. This reaction afforded the compound **40** (9.5 mg, 71%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.17-7.91 (m broad, 3H, H-3''), 5.20-5.05 (m broad, 1H, H-1), 5.00-4.78 (m broad, 6H, H-1''), 4.54-4.22 (m broad, 6H, H-4''), 4.03-3.48 (m broad, 18H, H-2, 3, 4, 5, 6, 2', 4', 6'), 3.53-3.36 (m broad, 3H, H-3'), 2.99-2.11 (m broad, 3H, H-1'$_A$), 2.81-2.64 (m broad, 3H, H-6''$_A$), 2.64-2.50 (m broad, 3H, H-6''$_B$), 2.50-2.33 (m broad, 3H, H-1'$_B$), 2.23-1.85 (m broad, 9 H, H-5', 5''); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.2 (C-2''), 125.0 (C-3''), 96.4 (C-1), 81.0 (C-3), 78.8 (C-2), 77.3 (C-4), 74.5 (C-3'), 70.0 (C-5), 67.8 (C-2',4'), 65.3 (C-6, 5'), 63.1 (C-1''), 57.9 (C-6'), 54.9 (C-1'), 49.0 (C-6''), 48.7 (C-4''), 24.5 (C-5'').

Compound 41:

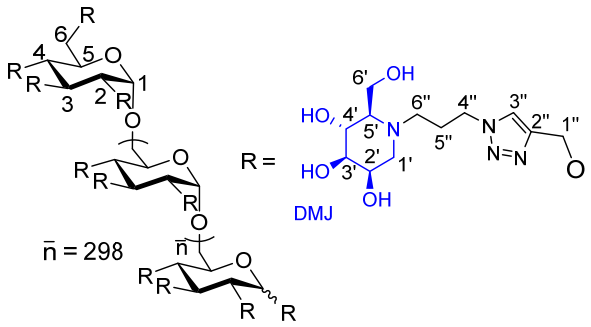


Obtained following method C. This reaction afforded the compound **41** (13 mg, 69%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.19-7.97 (m broad, 3H, H-3''), 5.27-5.14 (m broad, 1H, H-1), 5.06-4.84 (m broad, 6H, H-1''), 4.56-4.30 (m broad, 6H, H-4''), 4.15-3.55 (m broad, 18H, H-2, 3, 4, 5, 6, 2', 4', 6'), 3.55-3.38 (m broad, 3H, H-3'), 3.05-2.86 (m broad, 3H, H-1'$_A$), 2.86-2.69 (m broad, 3H, H-6''$_A$), 2.69-2.37 (m broad, 6H, H-1'$_B$, 6''$_B$), 2.30-1.91 (m broad, 9 H, H-

5', 5''); ^{13}C NMR (125 MHz, D_2O) δ(ppm): 144.3 (C-2''), 124.9 (C-3''), 96.6 (C-1), 80.9 (C-3), 79.2 (C-2), 76.9 (C-4), 74.6 (C-3'), 70.1 (C-5), 67.9 (C-2',4'), 65.4 (C-6, 5'), 63.3 (C-1''), 58.0 (C-6'), 54.9 (C-1'), 49.0 (C-6''), 48.7 (C-4''), 24.7 (C-5'').

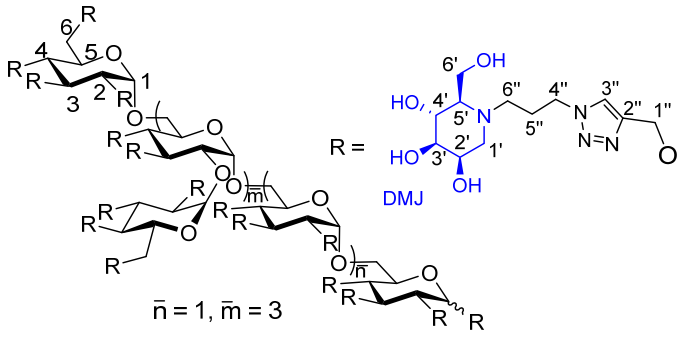
Compound 42:



Obtained following method C. This reaction afforded the compound **42** (7 mg, 53%) as a pale yellow solid.

^{1}H NMR (700 MHz, D_2O) δ(ppm): 8.22-8.00 (m broad, 3H, H-3''), 5.20-5.06 (m broad, 1H, H-1), 5.06-4.82 (m broad, 6H, H-1''), 4.61-4.33 (m broad, 6H, H-4''), 4.10-3.55 (m broad, 18H, H-2, 3, 4, 5, 6, 2', 4', 6'), 3.55-3.41 (m broad, 3H, H-3'), 3.47-2.89 (m broad, 3H, H-1'$_A$), 2.89-2.73 (m broad, 3H, H-6''$_A$), 2.73-2.43 (m broad, 6H, H-1'$_B$, 6''$_B$), 2.33-1.98 (m broad, 9 H, H-5', 5''); ^{13}C NMR (175 MHz, D_2O) δ(ppm): 144.2 (C-2''), 124.7 (C-3''), 96.3 (C-1), 80.5 (C-3), 78.8 (C-2), 77.8 (C-4), 74.2 (C-3'),69.6 (C-5), 67.7 (C-2',4'), 65.4 (C-6, 5'), 63.2 (C-1''), 57.8 (C-6'), 54.8 (C-1'), 49.1 (C-6''), 48.6 (C-4''), 24.6 (C-5'').

Compound 43:

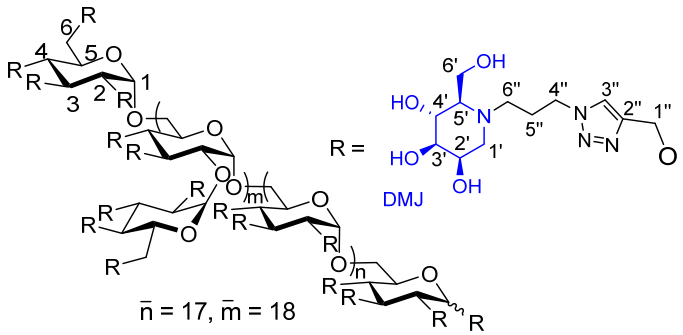


Obtained following method C. This reaction afforded the compound **43** (12 mg, 71%) as a pale yellow solid.

^{1}H NMR (700 MHz, D_2O) δ(ppm): 8.25-7.83 (m broad, 3H, H-3''), 5.41-5.14 (m broad, 1H, H-1), 5.06-4.83 (m broad, 6H, H-1''), 4.55-4.20 (m broad, 6H, H-4''), 4.0-3.54 (m broad, 18H,

H-2, 3, 4, 5, 6, 2', 4', 6'), 3.54-3.38 (m broad, 3H, H-3'), 3.02-2.84 (m broad, 3H, H-1'$_A$), 2.84-2.67 (m broad, 3H, H-6''$_A$), 2.73-2.35 (m broad, 6H, H-1'$_B$, 6''$_B$), 2.29-1.87 (m broad, 9 H, H-5', 5''); ^{13}C NMR (175 MHz, D_2O) δ(ppm): 144.2 (C-2''), 124.8 (C-3''), 74.3 (C-3'), 67.6 (C-2',4'), 65.3 (C-5'), 57.7 (C-6'), 54.7 (C-1'), 48.9 (C-6''), 48.5 (C-4''), 24.4 (C-5''). Only caracteristics peaks of iminosugars were observed on ^{13}C spectra.

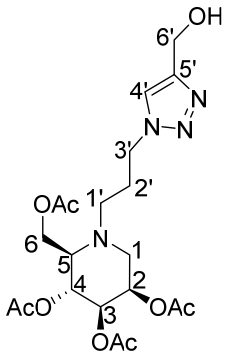
Compound 44:



Obtained following method C. This reaction afforded the compound **44** (11 mg, 70%) as a pale yellow solid.

^{1}H NMR (500 MHz, D_2O) δ(ppm): 8.30-7.88 (m broad, 3H, H-3''), 5.36-5.22 (m broad, 1H, H-1), 5.09-4.82 (m broad, 6H, H-1''), 4.54-4.22 (m broad, 6H, H-4''), 4.06-3.56 (m broad, 18H, H-2, 3, 4, 5, 6, 2', 4', 6'), 3.54-3.37 (m broad, 3H, H-3'), 3.05-2.87 (m broad, 3H, H-1'$_A$), 2.87-2.69 (m broad, 3H, H-6''$_A$), 2.69-2.33 (m broad, 6H, H-1'$_B$, 6''$_B$), 2.33-1.88 (m broad, 9 H, H-5', 5''); ^{13}C NMR (175 MHz, D_2O) δ(ppm): 144.4 (C-2''), 124.9 (C-3''), 96.6 (C-1), 80.5 (C-3), 78.7 (C-2), 77.8 (C-4), 74.5 (C-3'), 70.2 (C-5), 67.9 (C-2',4'), 65.5 (C-6, 5'), 63.2 (C-1''), 58.0 (C-6'), 54.9 (C-1'), 49.0 (C-6''), 48.7 (C-4''), 24.7 (C-5'').

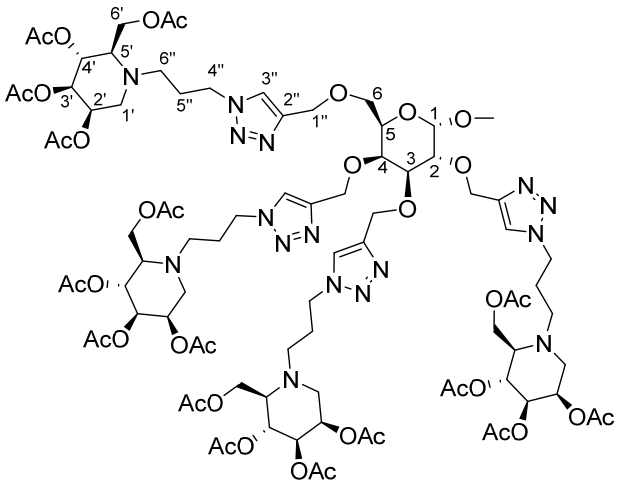
Compound 6'



To a solution of the propargyl alcohol (3.8 μL, 66.4 μmol) and **16** (25.0 mg, 60.3 μmol) in dioxane-H_2O (3 mL, 2–1) copper sulfate (4.5 mg, 18.3 μmol) and sodium ascorbate (7.2 mg, 36.2 μmol) were added and the mixture was irradiated at 80 °C for 45 minutes in a sealed vessel. The crude product was quenched with EDTA under agitation. The mixture was poured into an NH_4Cl satd. solution (10 mL) and extracted with ethyl acetate. The organic layer was dried ($MgSO_4$), filtered and the solvent removed under reduced pressure. The crude product was purified by flash chromatography on silica gel with 98/2 (DCM/MeOH) like eluent affording **6'** (22 mg, 79% of yield) as a uncolorless oil.

$[\alpha]_D^{20}$ = -37.5 (c = 1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 7.55 (s, 1H, H-4'), 5.26-5.20 (m, 1H, H-2), 5.17 (t, 1H, *J* = 6.7 Hz, H-4), 5.05 (dd, *J* = 3.2 Hz, *J* = 7.2 Hz, 1H, H-3), 4.79 (s, 2H, H-6'), 4.51-4.41 (m, 1H, H-3'$_A$), 4.41-4.31 (m, 1H, H-3'$_B$), 4.26-4.13 (m, 2H, H-6), 3.04 (dd, *J* = 6.6 Hz, *J* = 12.9 Hz, 1H, H-1$_A$), 2.87-2.79 (m, 1H, H-5), 2.79-2.70 (m, 1H, H-1'$_A$), 2.66-2.55 (m, 2H, H-1$_B$, 1'$_B$), 2.09, 2.07, 2.06, 2.04 (4s, 14H, H-2', C=OC*H3*); ^{13}C NMR (100 MHz, $CDCl_3$) δ(ppm): 171.0, 170.5, 169.9, 169.7 (C=O), 148.1 (C-5'), 122.0 (C-4'), 70.6 (C-3), 68.4 (C-4), 66.8 (C-2), 61.2 (C-5), 60.3 (C-6), 56.9 (C-6'), 49.5 (C-1'), 49.1 (C-1), 47.6 (C-3'), 27.4 (C-2'), 21.1, 21.0, 20.8 (CH_3); MALDI, TOF-HRMS $[M+H]^+$ m/z 471.2086 requires for $C_{20}H_{31}N_4O_9$; found 471.2085; FT-IR: (cm^{-1}) 3396 (ν_{OH}), 3029 (ν_{CH}), 2951, 2869 (ν_{CH2}), 2868 (ν_{C-H}), 1496 ($\nu_{C=C}$), 1452 ($\nu_{C=C}$), 1363, 1266 (ν_{C-O}), 1208 (ν_{C-C-N}), 1174, 1140 (ν_{C-O-C}), 1101 (ν_{C-O-C}), 1063, 1028, 1014, 737, 699 (ν_{C-H}).

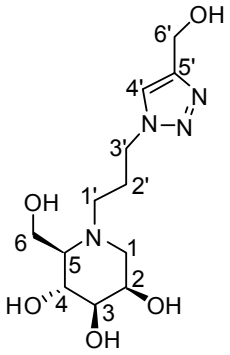
Compound 7':



Copper sulfate (43 mg, 173 μmol) and sodium ascorbate (69 mg, 350 μmol) were added to a solution of methyl 2,3,4,6-tetra-*O*-propargyl-α-D-galactopyranoside (20 mg, 57.8 μmol) and **16** (153 mg, 253.3 μmol) in dioxane-H_2O (5 mL, 4–1) and the mixture was irradiated at 80 °C for 45 min in a sealed vessel. Residual copper was removed with EDTA (101 mg, 346 μmol), and the mixture was poured into a NH_4Cl satd. solution (15 mL) and extracted with ethyl acetate (15mL). The organic layer was dried ($MgSO_4$), filtered and the solvent removed under reduced pressure. The crude product was purified by flash chromatography on silica gel with 97/3 (DCM/MeOH) as eluent to afford the compound **7'** (71 mg, 61% yield) as a colorless oil.

$[\alpha]_D^{20}$ = -19.6 (*c* = 1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ(ppm): 7.87, 7.74, 7.73, 7.64 (4 H, s, H-3''), 5.27-5.22 (m, 4H, H-2'), 5.22-5.15 (m, 4H, H-4'), 5.07-4.99 (m, 4H, H-3'), 4.93 (d, *J* = 12.3 Hz, 1H, H-1''), 4.86-4.71 (m, 6 H, H-1, 1''), 4.60 (q, *J* = 12.3 Hz, *J* = 28.2 Hz, 2H, H-1''), 4.51-4.39 (m, 4H, H-4''$_A$), 4.39-4.30 (m, 4H, H-4''$_B$), 4.30-4.17 (m, 8H, H-6'), 4.02-3.99 (m, 1H, H-4), 3.97 (dd, *J* = 3.5 Hz, *J* = 10.0 Hz, 1H, H-2), 3.90-3.84 (m, 2H, H-3, 5), 3.54 (d, *J* = 6.5 Hz, 2H, H-6), 3.35 (s, 3H, CH_3), 3.13-3.02 (m, 4H, H-1'$_A$), 2.91-2.77 (m, 8H, H-5', 6''$_A$), 2.74-2.58 (m, 8H, H-1'$_B$, 6''$_B$), 2.12-1.99 (m, 56H, H-5'', C=OC*H3*); ^{13}C NMR (100 MHz, $CDCl_3$) δ(ppm): 171.8, 170.4, 169.9, 169.7 (C=O), 145.4, 145.3, 145.2 144.8 (C-2''), 123.5, 123.3, 123.0 (C-3''), 98.5 (C-1), 78.8 (C-3), 78.5 (C-2), 74.8 (C-4),71.0, 70.9, 70.8 (C-3'), 69.1 (C-5), 69.0 (C-6), 68.3 (C-4'), 66.8 (C-2'), 66.1, 65.1, 64.8, 64.7 (C-1''), 61.3 (C-5'), 56.9 (C-6'), 60.1 (C-1'), 55.4 (CH_3), 49.7, 49.6, 49.4, 49.3 (C-1', 6''), 47.8, 47.7 (C-4''), 27.5, 27.3 (C-5''), 21.1, 21.0, 20.9, 20.8 (CH_3); MALDI, TOF-HRMS $[M+Na]^+$ m/z 2025.8311 requires for $C_{87}H_{126}N_{16}O_{38}Na$; found 2025.8234.

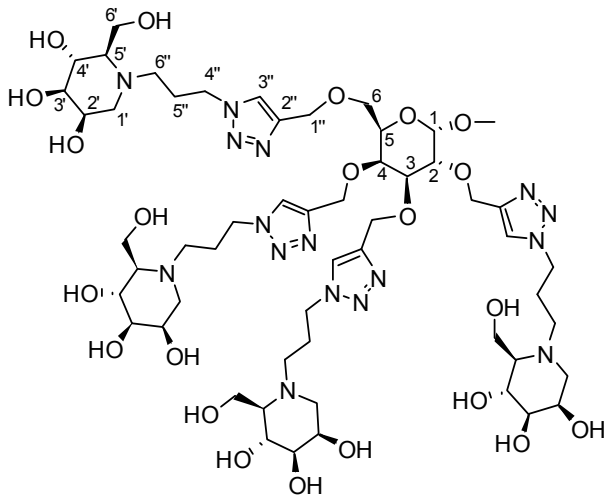
Compound 46



The protected compound **6'** (20 mg, 42.5 μmol) was dissolved in MeOH/H_2O (1:1, 5 mL). Amberlite resin IRA 400 1.25 meq/mL (570 mg) was added, and the mixture was stirred overnight at room temperature. The resin was then filtered and washed with methanol and water. The solvents were evaporated affording final compound **46** (11 mg, 86%) as pale yellow solid.

$[\alpha]_D^{20}$ = - 42.4 (*c* = 0.5, H_2O); ^{1}H NMR (400 MHz, D_2O) δ(ppm): 7.97 (s, 1H, H-4'), 4.69 (s, 1H, H-6'), 4.47-4.38 (m, 2H, H-3'), 3.96-3.90 (m, 1H, H-2), 3.75 (d,2H, H-6), 3.61 (t, 1H, *J* = 9.5 Hz, H-4), 3.43 (dd, *J* = 3.5 Hz, *J* = 9.5 Hz, 1H, H-3), 2.90 (dd, *J* = 3.5 Hz, *J* = 12.8 Hz, 1H, H-1_A), 2.78-2.65 (m, 1H, H-$1'_A$), 2.63-2.54 (m, 1H, H-$1'_B$), 2.51 (dd, *J* = 1.3 Hz, *J* = 12.8 Hz, 1H, H-1_B), 2.21-2.13 (m, 1H, H-5), 2.13-2.01 (m, 1H, H-2'); ^{13}C NMR (100 MHz, D_2O) δ(ppm): 146.8 (C-5'), 124.1 (C-4'), 74.3 (C-3), 67.7 (C-4, 2), 65.1 (C-5), 57.8 (C-6), 54.6 (C-1), 54.5 (C-6'), 48.8 (C-1'), 48.5 (C-3'), 24.0 (C-2'); ES-HRMS $[M+H]^+$ m/z 303.1663 requires for $C_{12}H_{23}N_4O_5$; found 303.1665; FT-IR: (ATR, cm^{-1}) 3402 (ν_{OH}), 1631 (ν_{NH}), 1459 (ν_{C-N}), 1081 (ν_{CH-OH}), 1049(ν_{CH-OH}), 960, 895.

Compound 48:



The protected compound **7'** (50 mg, 24.9 μmol) was dissolved in MeOH/H2O (1:1, 5 mL). Amberlite resin IRA 400 OH- 1.25 meq/mL (335 mg) was added, and the mixture was stirred overnight at room temperature. The resin was then filtered and washed with methanol and water. The solvent were evaporated giving final compound **48** (30 mg, 91%) as a white solid.

$[\alpha]_D^{20}$ = -15.3 (*c* = 1, H_2O); ^{1}H NMR (400 MHz, D_2O) δ(ppm): 8.12, 8.11, 8.10, 8.09 (4 H, s, H-3''), 4.96-4.63 (m, 9H, H-1, 1''), 4.60-4.44 (m, 8H, H-4''), 4.18-4.14 (m, 1H, H-3), 4.07-3.98 (m, 5H, H-5, 2'), 3.98-3.89 (m, 2H, H-2, 4), 3.89-3.79 (m, 8H, H-6'),3.75-3.66 (m, 4H, H-4'), 3.65-3.58 (m, 1H, H-6), 3.56-3.47 (m, 5H, H-6, 3'), 3.39 (s, 3H, CH_3), 3.05-2.94 (m, 4H, H-1'$_A$), 2.90-2.76 (m, 4H, H-6''$_A$), 2.73-2.61 (m, 4H, H-6''$_B$), 2.61-2.50 (m, 4H, H-1'$_B$), 2.30-2.20 (m, 4H, H-5'), 2.20-2.07 (m, 8H, H-5''); ^{13}C NMR (100 MHz, D_2O) δ(ppm):144.2, 143.8 (C-2''), 125.2, 125.1, 125.0 (C-3''), 97.7 (C-1), 77.5 (C-3), 75.6 (C-2), 74.7 (C-4),74.4 (C-3'), 68.8 (C-6), 68.5 (C-5), 67.8 (C-4'), 67.7 (C-2'), 65.3 (C-5'), 64.9, 63.5, 63.4, 63.0 (C-1''), 57.9 (C-6'), 55.1 (CH_3), 54.7 (C-1'), 48.9(C-6''), 48.6 (C- 4''), 24.3 (C-5''); MALDI, TOF-HRMS $[M+H]^+$ m/z 1331.6807 requires for $C_{55}H_{95}N_{16}O_{22}$; found 1331.6851.

Production of recombinant TmFuc, TmGal and AgaB

Recombinant strains expressing AgaB, TmFuc or TmGal genes were grown in 1 L of LB medium at 37°C overnight, centrifuged and resuspended in 35 mL of lysis buffer (0.1 M phosphate, pH 8, 0.5 M NaCl, 10 mM imidazole, 10 $\mu g.mL^{-1}$ DNAse I). After sonication and centrifugation, 6xHis-tagged proteins were purified by immobilized ion metal-affinity chromatography (IMAC): 250 µL of Ni-NTA Superflow (Qiagen) was added to the supernatant and stirred for 1 h at 4°C, then loaded onto a 10 mL column. The column was washed with 25 mL of buffer (0.1 M phosphate, pH 8, 0.5 M NaCl, 25 mM imidazole), then 5 mL of elution buffer was added (0.1 M phosphate, pH 8, 0.5 M NaCl, 250 mM imidazole). Purity of the final product was checked by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Enzyme concentrations were determined by UV absorbance at 280 nm using NanoDrop 1000 (Thermo Scientific).

Production of recombinant ManIIb

Pichia pastoris host strain GS115 was transformed with pPICZαHisFlag-GMIIb plasmid with cloned Golgi mannosidase ManIIb from *Drosophila melanogaster*. The ManIIb mannosidase was expressed and purified as described in the *Pichia* expression manual (Invitrogen) and Nemčovičová *et al.* 2013 (Nemčovičová I, Šesták S, Rendić D, Plšková M, Mucha J, Wilson IBH. (2013) Characterisation of class I and II α-mannosidases from *Drosophila melanogaster*. Glycoconj J. 30(9): 899-909). The isolated mannosidase was stabilized by adding glycerol to 30% final concentration and kept at -20 °C.

Production of recombinant Uhgb_MP

First, Uhgb_MP encoding gene was subcloned from pDEST17 vector (Ladeveze et al., 2013) into pET28a between NdeI and BamHI restriction sites, yielding a 6-histidine N-terminal thrombin cleavable tagged protein. Primers forward 5' TACGCTAGCCATATGAGTATGAGTAGCAAAGTT 3' and reverse 5' TACGCTGGATCCTCAGATGATGCTTGTA 3' were used to amplify the gene, using the Phusion® High-Fidelity DNA Polymerase (New England Biolabs), and to introduce NdeI and BamHI restriction sites on 5' and 3' ends of the PCR product. Both the PCR product and the pET28a plasmid were subsequently digested with NdeI and BamHI, purified and ligated. After *E. coli* TOP10 transformation, the construct was extracted and verified by sequencing.

To allow heterologous UhgbMP production, *E.coli* BL21-AI cells (Invitrogen) transformed with the UhgbMP::pET28a plasmid were cultured at 20°C for 24 hours in 200mL ZYM-5052 autoinduction medium (Studier et al., 2005) supplemented with 50 µg/mL kanamycin, inoculated at OD_{600nm}=0.1. Cells were harvested and resuspended in 20mM Tris HCl, pH 7.0, 300 mM NaCl, and lysed by sonication. Soluble lysate was applied to a TALON resin loaded with cobalt (GE Healthcare) equilibrated in 20mM Tris HCl, pH 7.0, 300 mM NaCl. After column washing with 8 volumes of the same buffer supplemented with 10mM imidazole, the protein was eluted in 20mM Tris HCl, pH 7.0, 300 mM NaCl, 150 mM imidazole. Finally, the protein sample was desalted on a PD-10 column (GE Healthcare) and eluted in 20 mM Tris-HCl pH7.0, Tween 80 0.1% (vol/vol).

General procedure for the inhibition assays on glycoside hydrolases

Inhibition constant (K_i) values were determined spectrophotometrically, by measuring the residual hydrolytic activities of the glycosidases against *p*-nitrophenyl α-D-glycopyranoside for JbMan, TmGal and TmFuc. Each assay was performed in phosphate-citrate buffer at the optimal pH for the enzymes. The reactions were initiated by addition of enzyme to a solution of the substrate in the absence or presence of various concentrations of inhibitors. The mixture was incubated for 10-30 min at 37 ºC and the reaction was quenched by addition of 1 M Na_2CO_3. Reaction times were appropriate to obtain less than 10% conversion of the substrate in order to achieve linear rates. The absorbance of the resulting mixture was determined at 405 nm. Note that for AgaB, the absorbance was directly monitored throughout the reaction each 30 secondes during 45 min at 30°C. Then initial reaction rates were calculated from the slope of the first-order plot of product concentration (*p*NP) against reaction time. Approximate values of K_i were determined using a fixed concentration of substrate (around the K_M value for the different glycosidases) and various concentrations of inhibitors. All measurements were made in triplicate. Full K_i determinations and enzyme inhibition mode were determined from the slope of Lineweaver-Burk plots and double reciprocal analysis. AgaB, TmFuc and TmGal were fitted to the competitive or uncompetitive equations using the nls package for R 2.15.0 (*www.**r**-project.org)*.

General procedure for the inhibition assays on Uhgb_MP

Effect of the iminosugars on Uhgb_MP phosphorolytic activity was assessed on *p*NP-β-D-mannopyranose. All reactions were carried out with 0.1 $mg.ml^{-1}$ of purified enzyme, at 37 °C

(UhgbMP optimal temperature) in Tris HCl 20mM, pH 7.0 (UhgbMP optimal pH). Activity was measured in presence of 10 mM inorganic phosphate, 1mM *p*NP-β-D-mannopyranose, and 500 µM iminosugar. The *p*NP release rate was monitored at OD_{405nm} on a Carry-100 UV-visible spectrophotometer (Agilent technologies).

The percentage of activity inhibition and activation factor (for compound **44** only) were calculated as follows, Ai_{Im} and Ai corresponding to the initial rates of *p*NP release in presence and in absence of 500 mM iminosugars, respectively:

$$\%\, Inhibition = 100 - \frac{Ai_{Im} \times 100}{Ai}$$

$$Activation\ factor = \frac{Ai_{Im}}{Ai}$$

Representative Lineweaver-Burk Plot

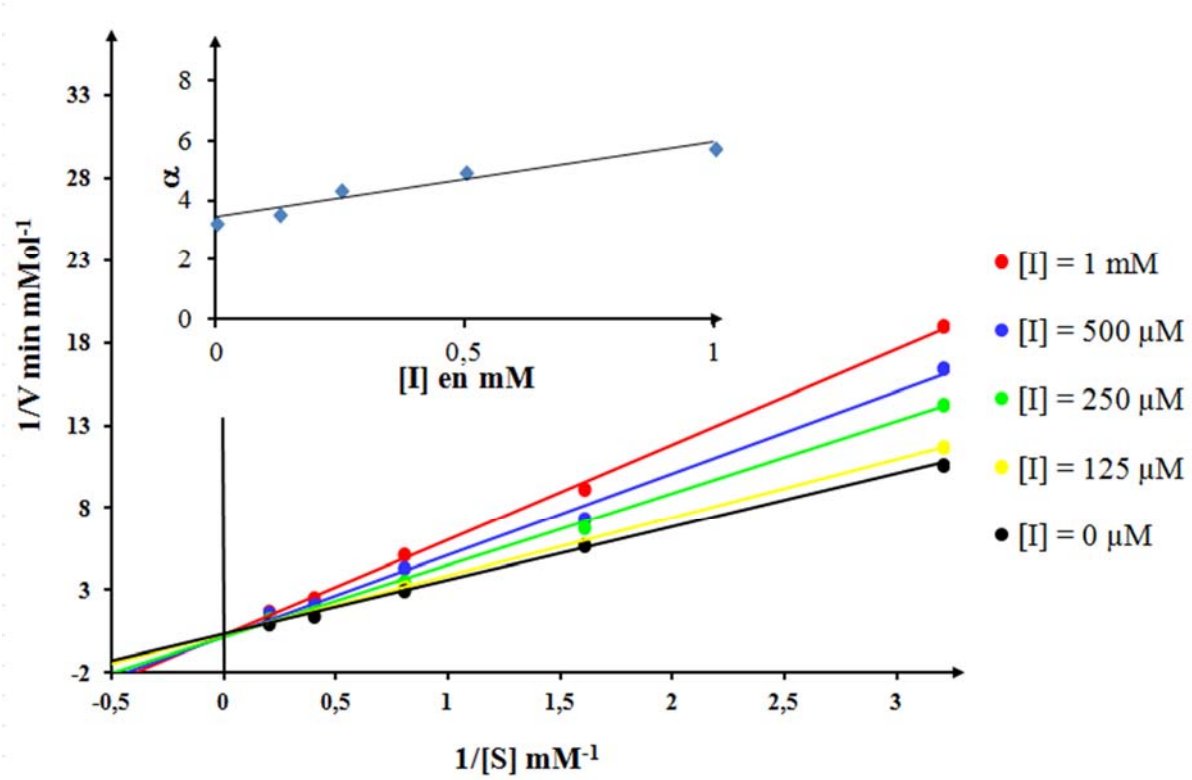


Figure S2: Lineweaver-Burk Plot for K_i determination (394 µM) of **45** against JbMan (pH 5.5).

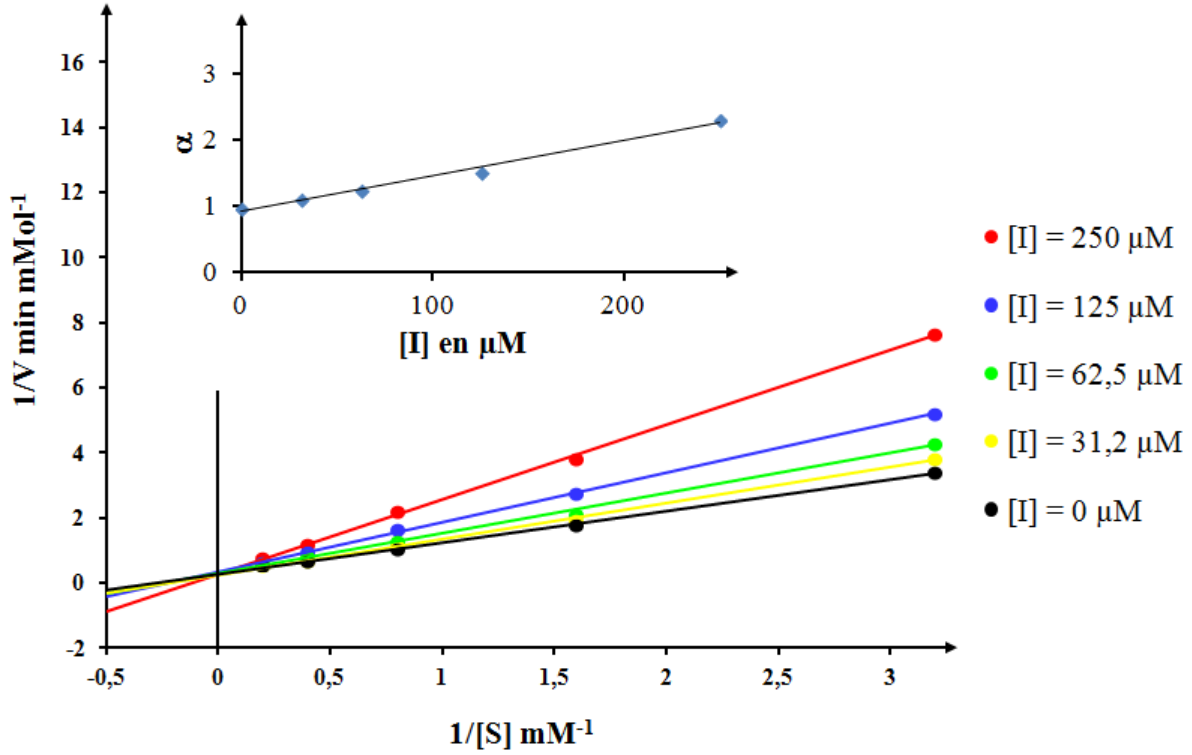


Figure S3: Lineweaver-Burk Plot for K_i determination (167 µM) of **46** against JbMan (pH 5.5).

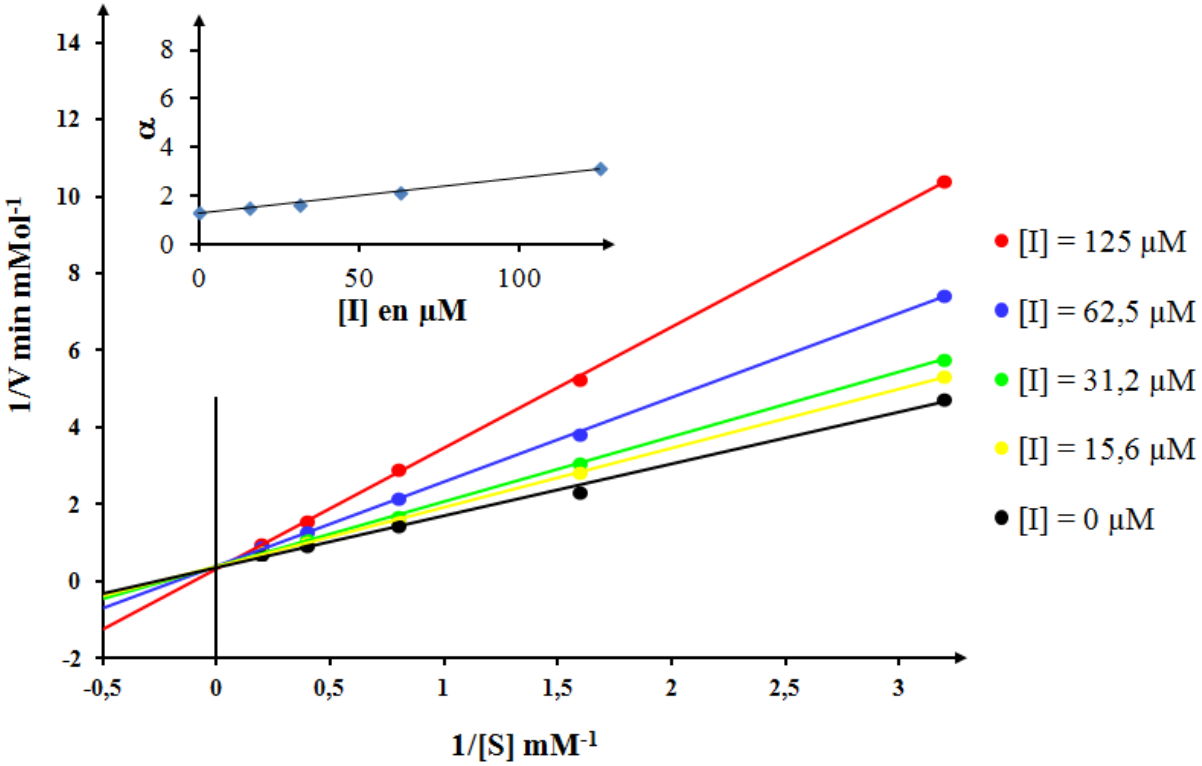


Figure S4: Lineweaver-Burk Plot for K_i determination (196 µM) of **47** against JbMan (pH 5.5).

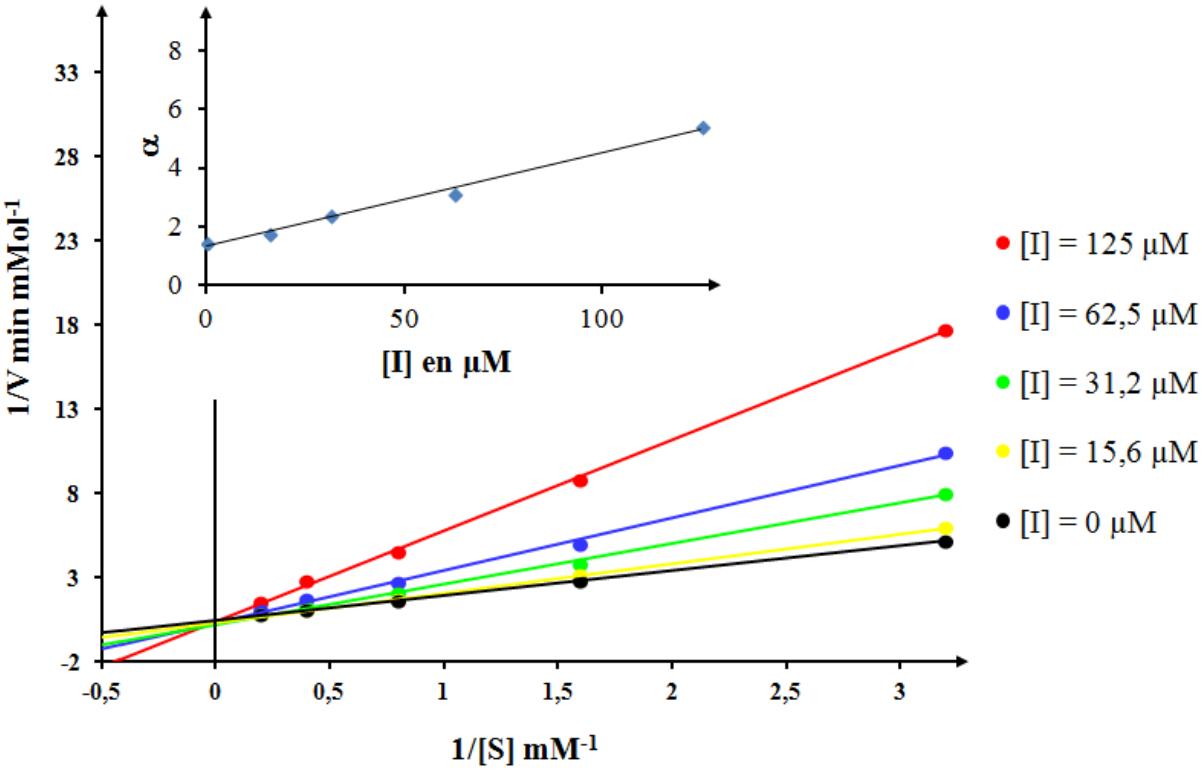


Figure S5: Lineweaver-Burk Plot for K_i determination (132 µM) of **48** against JbMan (pH 5.5).

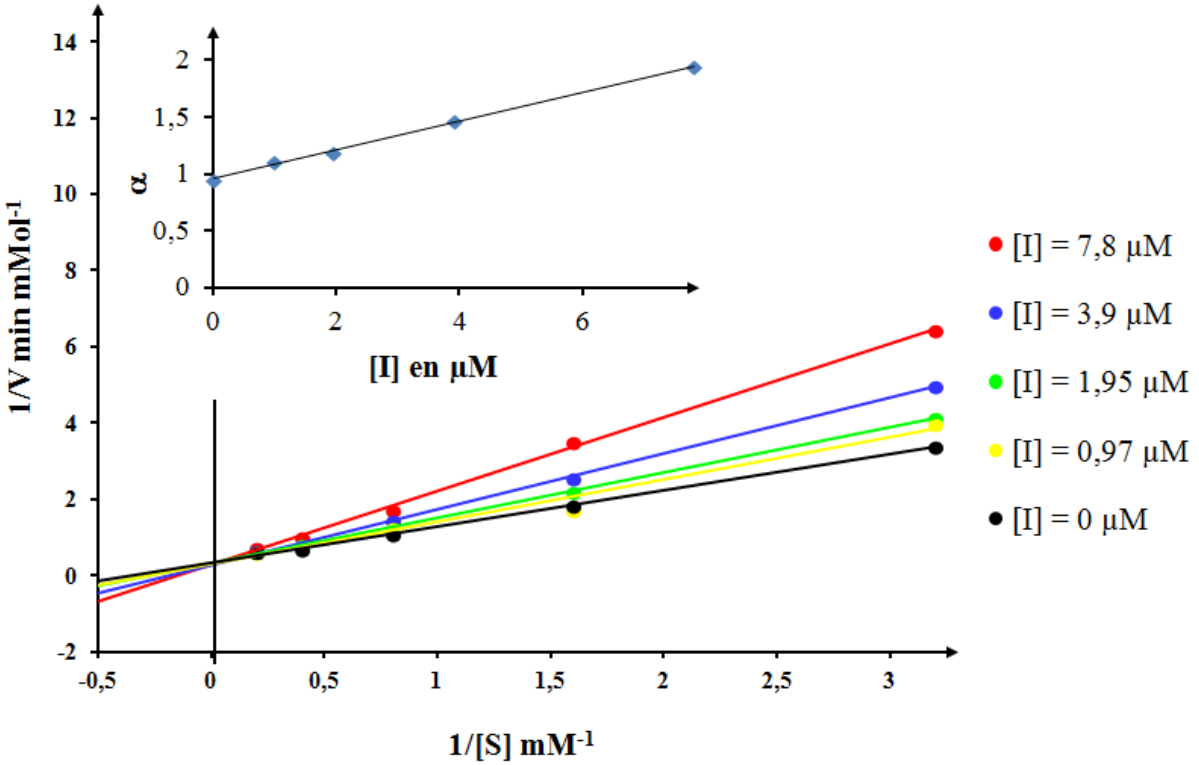


Figure S6: Lineweaver-Burk Plot for K_i determination (6 µM) of **31** against JbMan (pH 5.5).

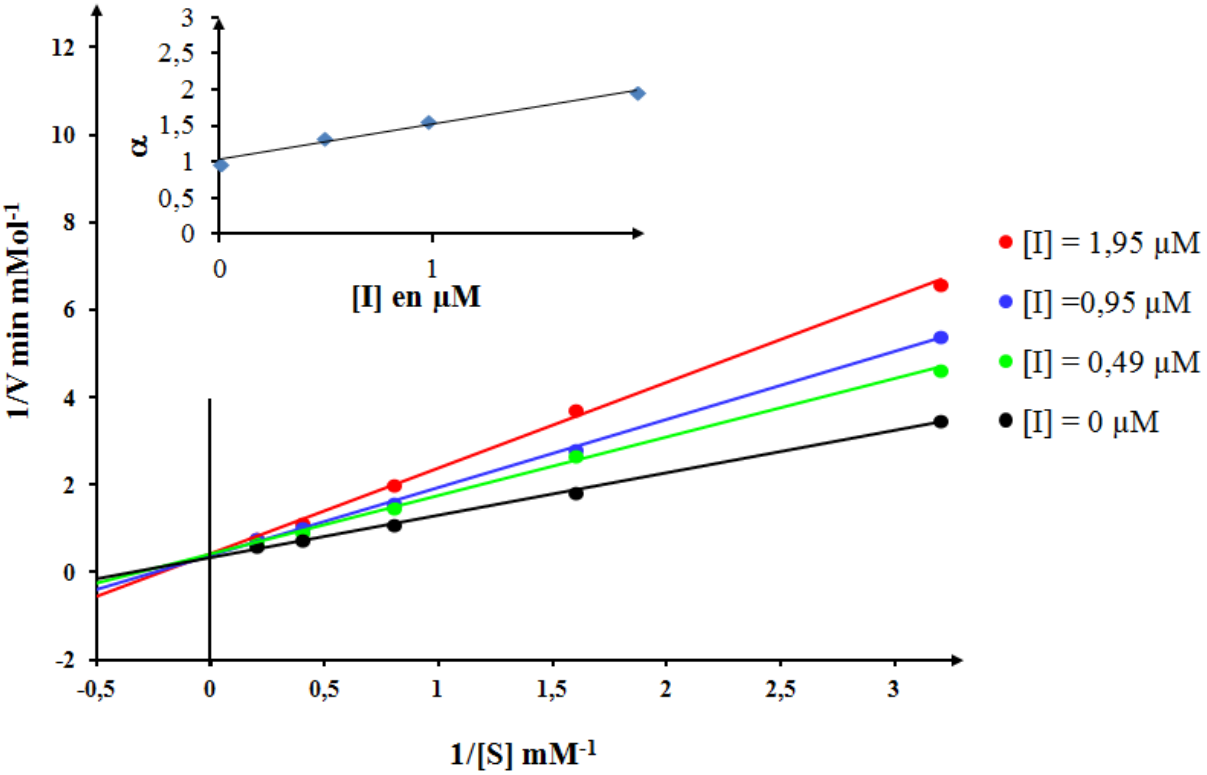


Figure S7: Lineweaver-Burk Plot for K_i determination (2 µM) of **38** against JbMan (pH 5.5).

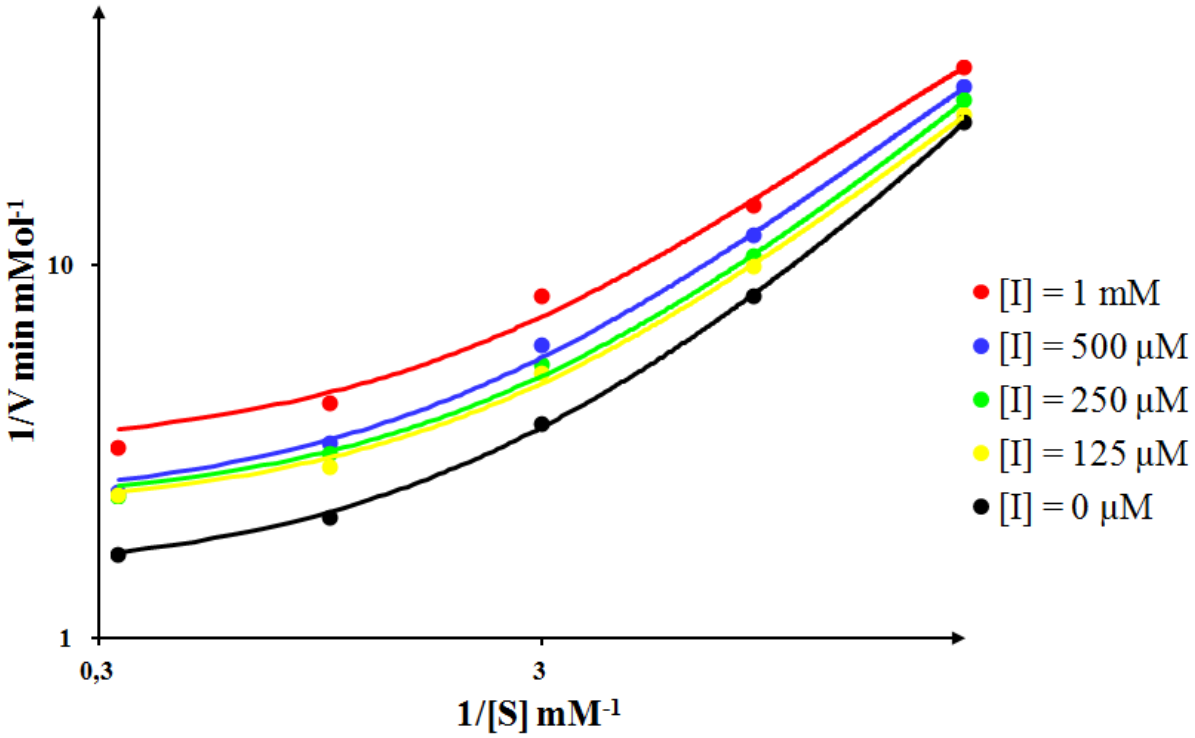


Figure S8: Logarithmic Lineweaver-Burk Plot for K_i determination (960 µM) of **45** against TmGal (pH 6.0).

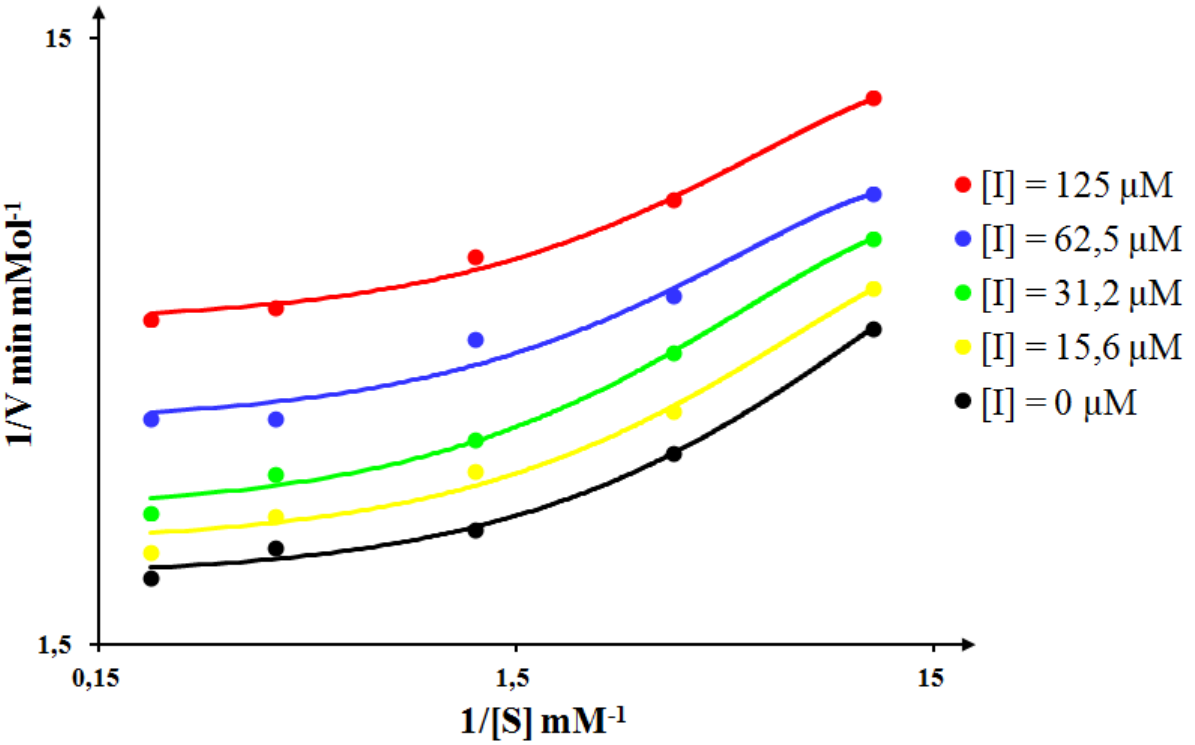


Figure S9: Logarithmic Lineweaver-Burk Plot for K_i determination (276 µM) of **47** against TmGal (pH 6.0).

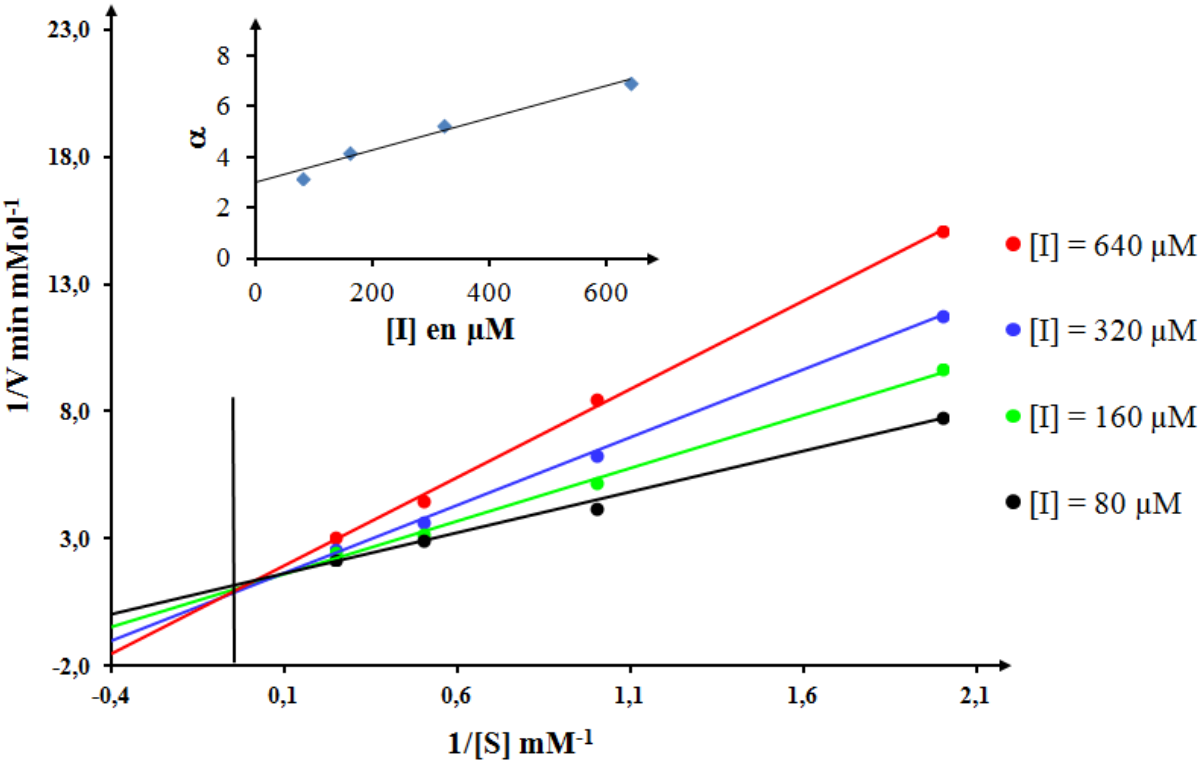


Figure S10: Lineweaver-Burk Plot for K_i determination (435 µM) of **46** against ManIIb (pH 6.0).

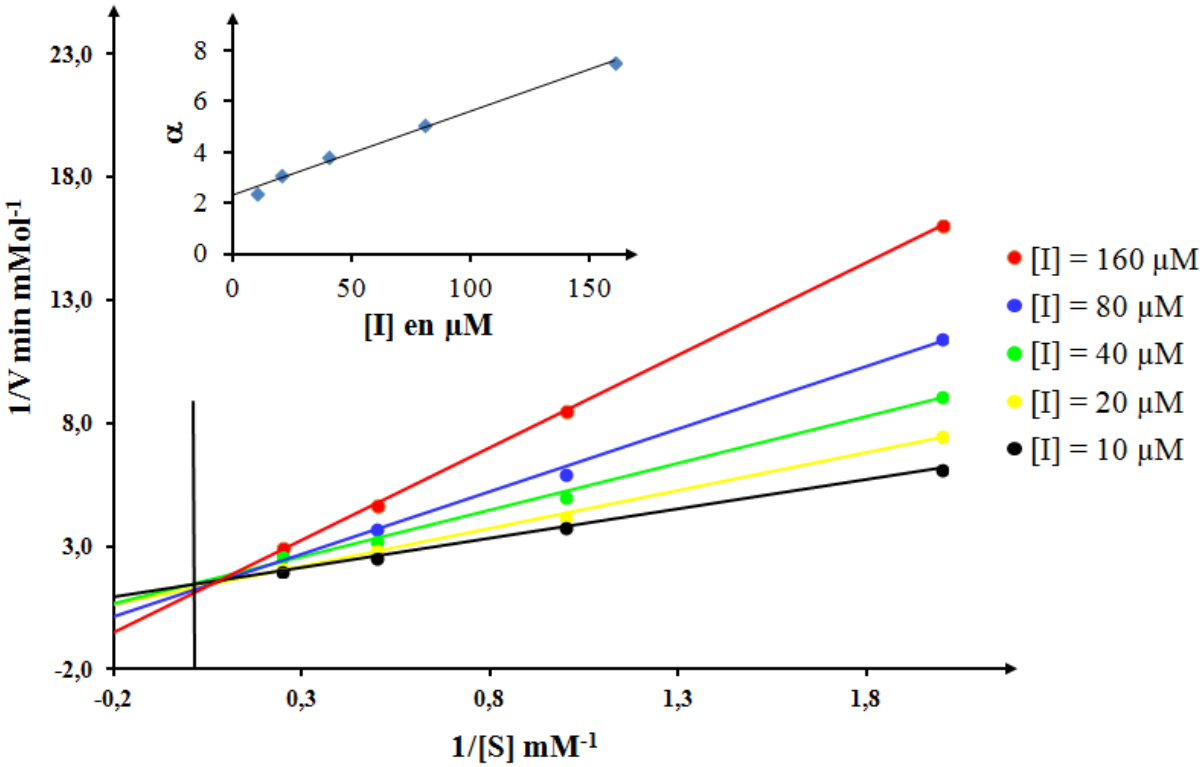


Figure S11: Lineweaver-Burk Plot for K_i determination (62 µM) of **48** against ManIIb (pH 6.0).

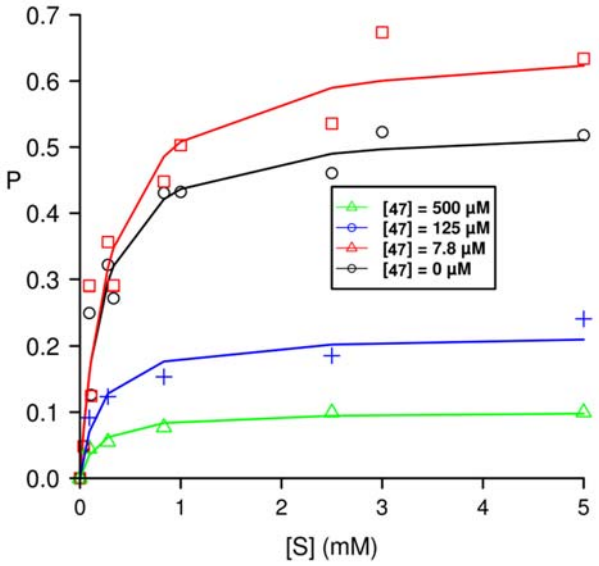


Figure S12: Activation and inhibition of TmGal at specific concentrations of tetravalent **47**

DLS measurements

DLS study was carried out using purified Uhgb_MP Tris in 20 mM Tris-HCl pH7.0, Tween 80 0.1% (vol/vol), supplemented with various concentrations of **44**, to achieve a molar ratio **44** to protein of 1:1; 10:1 and 200:1 (i.e 2.5, 25 and 500µM of **44**). Scattered intensities were recorded at 293 K using a DLS Nanostar (Wyatt Technology) and processed using DYNAMICS V7.1.5.6. Each measurement consisted in 10 data acquisition of 5s with the laser power set to automatic mode.

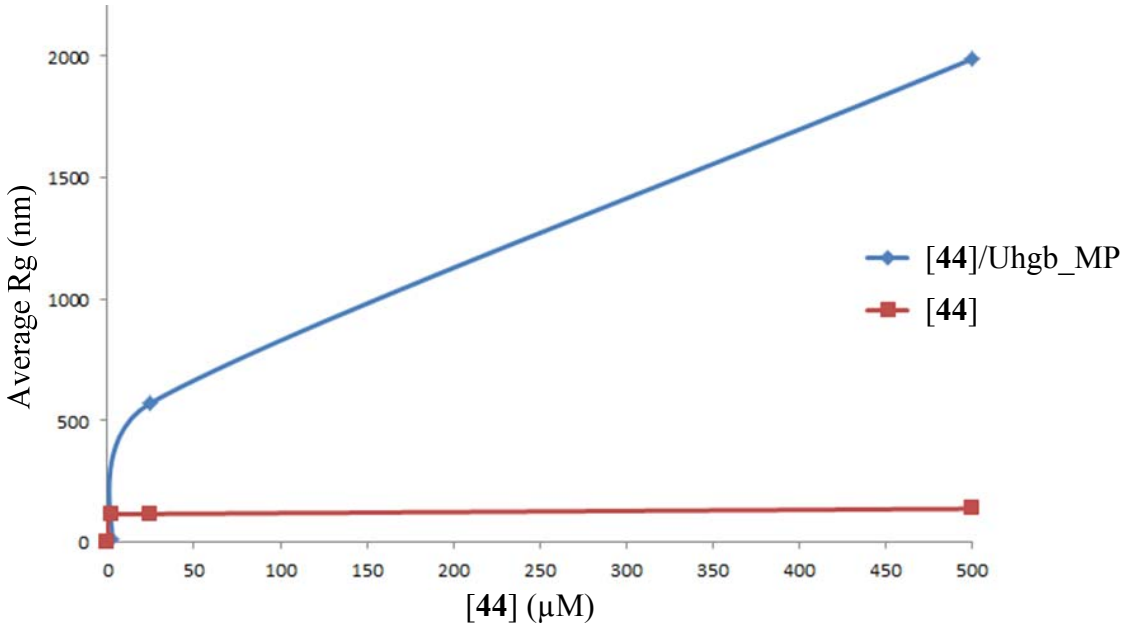


Figure S13: DLS analysis of the Uhgb_MP-**44** aggregates

Fourth article

Analysis of the GH130 content of the human gut microbiome highlighted highly prevalent and abundant sequences, in the GH130_1, GH130_2 and GH130_NC clusters. However, so far, too few of these enzymes encoded by highly prevalent genes, have been characterized. This is the case in particular for enzymes belonging to the GH130_NC cluster, in which four members belonging to gut bacteria have been crystallized, but not functionally characterized. This lack of functional data is thus preventing us to understand what is their role in the ecosystem. Moreover, we proposed in the first part of the 'Results' section, basing on GH130 primary structure analysis, that members of the NC cluster would be hydrolases. To address this specific question, a functional analysis of the enzyme BACOVA_03624, belonging to the GH130_NC cluster, for which a 3D structure is available, was undertaken. BACOVA_03624 is one of the 7 GH130 enzymes of the human gut symbiont *Bacteroides ovatus* ATCC 8483. In order to get relevant data on their role in carbohydrate foraging by *B. ovatus*, a GH130 targeted transcriptomic study was initiated, thanks to Pascale Mosoni and Jordane Despres (INRA Clermont-THeix). The first results are presented in the following chapter.

The human gut symbiont *Bacteroides ovatus* possesses a highly polyvalent GH130 enzyme acting on galactosides and mannosides

Simon Ladevèze[1,2,3], Angeline Rizzo[1,2,3], Elisabeth Laville[1,2,3], Jordane Despres[4], Pascale Mosoni[4], Sandrine Morel[1,2,3] and Gabrielle Potocki-Véronèse[1,2,3*]

[1]Université de Toulouse; INSA, UPS, INP; LISBP. 135 Avenue de Rangueil, F-31077 Toulouse, France.
[2]CNRS, UMR5504, F-31400 Toulouse, France.
[3]INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France.
[4]INRA, UR454 Microbiologie, F-63122 Saint-Genès Champanelle, France
*corresponding author: Gabrielle Potocki-Véronèse, LISBP, INSA, CNRS UMR5504, INRA UMR792, 135 avenue de Rangueil, 31077 Toulouse, France; Tel: +33 5 61 55 94 87; Fax : +33 5 61 55 94 00 ; E-mail : veronese@insa-toulouse.fr

Abstract

BACOVA_03624 is one of the seven enzymes of *Bacteroides ovatus* ATCC 8483 that are classified in the family 130 of glycoside hydrolases. Up to now, the only characterized GH130 enzymes were glycoside phosphorylases, specifically acting on β-mannosides. Here, the biochemical characterization of BACOVA_03624 revealed that it is rather a hydrolase than a phosphorylase. This enzyme is thus the first known hydrolase of the GH130 family. Moreover, in presence of inorganic phosphate, it is able to breakdown either galactosides and mannosides by using the two mechanisms. It is also capable of catalysing both transglycosylation and reverse-phosphorolysis reactions, to glycosylate a large range of acceptors. The large flexibility of its -1 and +1 subsites is a unique feature among glycoside phosphorylases, and makes it a highly polyvalent enzyme for carbohydrate breakdown and synthesis. Analysis of the genomic context of *B. ovatus* GH130 sequences, coupled to a preliminary transcriptomic study, suggests that in the human gut, BACOVA_03624 would be involved in the catabolism of human or yeast high mannose N-glycans, by hydrolyzing the β-D-Gal*p*-1,4-D-Glc*p*NAc and β-D-Man*p*-1,4-D-Glc*p*NAc motives.

Introduction

Glycoside phosphorylases (GPs) are carbohydrate active enzymes (CAZymes), which catalyze the breakdown of a glycosidic linkage from oligosaccharides or polysaccharides with concomitant phosphate glycosylation, to yield a glycosyl-phosphate product and a glycan of

reduced polymerization degree. Contrary to glycoside hydrolases (GHs), GPs are also able to perform carbohydrate synthesis (in the so-called 'reverse phosphorolysis' reaction) to form a glycosidic bond between the glycosyl unit originating from the glycosyl-phosphate, which acts as donor, and a carbohydrate acceptor. This property makes them highly powerful biocatalytic tools for synthesis of high added value carbohydrates (O'Neill and Field 2015), in contrast to classical Leloir-type glycosyltransferases (GTs) that use expensive and non-readily available nucleotide sugars.

In vivo, GPs would be more advantageous than hydrolysis, since the chemical energy of glycosidic bonds is kept in the form of glycosyl-phosphates, which are directly metabolized. This would be especially the case under particular physiologic conditions, such as in anoxic environments, where the ATP required for carbohydrate kinases involved in the conventional catabolic pathways cannot be efficiently produced by the respiratory chain-linked phosphorylation process (Zhang and Lynd 2005).

Despite this interesting feature, the diversity of known GPs is much lower than that of GHs and GTs. Nevertheless, their prevalence in all living organisms and their role in carbohydrate metabolism may be underestimated, as they may often be misannotated as GHs or GTs in genomic and metagenomics datasets. Indeed, GPs share strong structural and mechanistic similarities with true GTs and GHs, and are thus classified in 6 GH and 2 GT families in the CAZy classification (Lombard et al. 2014). In all cases, they present the same specificity towards glycosides than those of the true GHs and GTs which closely look like them, belonging to the same CAZy family or subfamily. To date, only 132 GPs have been functionally characterized (1 GH3, 16 GH13, 15 GH65, 24 GH94, 12 GH112, 8 GH130, 2 GT4 and 54 GT35). Most of them are able to phosphorolyse α- or β-linked glucans and gluco-oligosaccharides, or glycans containg α-D-linked glucosyl residues at their non-reducing end (retaining GT4, GT35, GH13 GPs, and inverting GH65 and GH94 GPs). Others, represented by only 23 characterized enzymes, can act onto various oligosaccharides with β-linked *N*-acetylglucosaminyl (1 retaining GH3 and 2 inverting GH94), galactosyl (12 inverting GH112) or mannosyl residues (8 inverting GH130) at their non-reducing end.

The GH130 family is the most recent GPs containing family. In January 2015, it counts 647 sequences, mostly bacterial ones. Only 8 GH130 enzymes have been biochemically characterized, their physiologic function being assigned to bacterial metabolization of plant, mammal or fungal mannosides. Indeed, they all catalyze the reversible phosphorolysis of β-mannosides (β-D-Man*p*-1,4-D-Glc, β-D-1,4 linked manno-oligosaccharides and mannan, β-D-1,2-linked manno-oligosaccahrides, and the core N-glycan motives β-D-Man*p*-1,4-D-GlcNAc and β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc) to form α-D-mannose-1-phosphate with an inversion of its anomeric configuration.Thus all of them are highly specific for mannosyl donors.

Like other GH-related inverting GPs, GH130 enzymes use a single displacement mechanism, involving phosphate itself and a single catalytic residue for the nucleophilic attack to the glycosidic bond. However, as suggested by Nakae *et al*. (Nakae et al. 2013) and confirmed by Uhgb_MP structural analysis (Ladevèze et al. 2015), for GH130 GPs, the assistance of C_3-OH would be required to relay proton transfer because the proton donor is too far away from the interosidic oxygen to be able to directly interact with it. This mechanism is still very close to that of typical inverting GHs, the main difference being that for GHs, the nucleophilic attack of the glycoside C_1 is performed by a water molecule activated by a second catalytic residue, which plays the role of base.

Based on sequence similarities within its members, two GH130 subfamilies have been created in 2013 (Ladevèze et al. 2013) (Figure 1). The GH130_1 gathers sequences resembling those of *Bf*MGP, *Ra*MP1 and *Rm*MGP, three enzymes exhibiting a narrow specificity for β-D-Man*p*-1,4-D-Glc. *Bf*MGP is the only enzyme of the GH130_1 subfamily of which 3D structure was solved (PDB code 4kmi). GH130_2, on the contrary, contains sequences coding for enzymes of which specificity towards mannosides is much broader, like Uhgb_MP, *Bt*1033 and *Ra*MP2. Two structures are available for GH130_2 enzymes, that of the human N-glycan phosphorylase Uhgb_MP (PDB code 4udi), and that of *Tm*1225, a protein from *Thermotoga maritima* MSB8 of unknown function (PDB code 1vkd).

The GH130_NC group contains 435 sequences which are too distantly related to those of the GH130_1 and GH130_2 subfamilies to belong to them. Four structures are available for the GH130_NC cluster (PDB codes 3taw, 3qc2, 4onz and 3r67), all belonging to a specific branch of the GH130 phylogenetic tree, which gathers 60 sequences (Figure 1). However, none of these putative enzymes have been biochemically characterized to date, the only two enzymes of known function (Teth514_1788 and Teth514_1789, (Chiku et al. 2014)) belonging to the most 'diffuse' GH130_NC part of this tree (Figure 1). On the 435 sequences of the GH130_NC cluster, 396 (not including the Teth514_1788 and Teth514_1789 sequences) present two common traits. First, the tyrosine residue at position 103 in Uhgb_MP, which lies in the +1 subsite, is strictly conserved in subfamilies GH130_1 and _2 but is replaced by a glutamic acid in most of the GH130_NC sequences (Figure 2). Secondly, residues Arg150, Arg168, and Asn151 (Uhgb_MP numbering) that are perfectly conserved among the GH130_1 and GH130_ 2 sequences, and which strongly stabilize phosphate in enzyme active site Ladevèze et al, 2015), are not perfectly conserved in the GH130_NC cluster. Ladevèze *et al.* (Ladevèze et al. 2013) thus suspected that many enzymes belonging to the GH130_NC cluster would be capable to act as hydrolases rather than phosphorylases, the glutamic acid corresponding to Uhgb_MP Tyr103 playing the role of base.

In order to test this hypothesis and to explore the spectrum of specificities towards glycosyl donors in the GH130 family, we characterized the BACOVA 03624 protein from *Bacteroides ovatus* ATCC 8483. This protein is a key representative of the GH130_NC group. First, its structure (3qc2) has been released in the Protein Data Bank in 2011, but to date, no function has been attributed to this protein. Secondly, it presents less than 30 % protein sequence identity (23% with *Ra*MP1, 22% with Uhgb_MP, 20% with *Ra*MP2, 21% with *Bf*MGP, 24% with Teth514_1789 and 24% with Teth514_1788) with the characterized GH130 GPs, which is a good argument to look for new functionalities. Thirdly, contrarily to the other characterized GH130, it is predicted to possess a signal peptide, which indicates that it may participate in carbohydrate degradation in a different way than the other characterized GH130 do. Finally, BACOVA_03624 is encoded by a highly prevalent and abundant gene in the human gut microbiome (Ladevèze et al. 2013), indicating that it probably plays a key role in carbohydrate foraging. This functional study, together with a GH130 targeted transcriptional analysis of *Bacteroides ovatus* ATCC 8483, allowed to get insights into the physiologic role of this protein in the human gut.

Results and discussion

BACOVA 03624 activity detection

BACOVA_03624 activity was first tested on *p*NP-β-D-Mannopyranoside (*p*NP-β-Man*p*), used as β-mannoside model substrate, and inorganic phosphate (P_i). Surprisingly, no activity was detected. Even if, as observed for Uhgb_MP (Ladevèze et al. 2013), *p*NP-β-Man*p* could be a bad glycosyl donor compared to β-D-manno-oligosaccharides, we suspected that BACOVA_03624 could exhibit a different substrate specificity, compared to all characterized GH130 enzymes. Its activity was thus tested on various *p*NP-β-glycosides (*p*NP-β-D-galactopyranose (*p*NP-β-Gal*p*), *p*NP-β-D-lactopyranose, *p*NP-β-D-galactofuranose, *p*NP-cellobioside, *p*NP-β-D-glucopyranoside, *p*NP-β-D-xylopyranoside, *p*NP-β-D-Glc*p*NAc), in presence of P_i. Activity was only detected on *p*NP-β-Gal*p*, indicating that BACOVA_03624 is able to act on galactosides. This is the first evidence of this substrate specificity for enzymes belonging to the GH130 family.

Influence of pH and temperature on BACOVA_03624 activity was then studied, in order to determine the optimal conditions for its characterization (Figure 3). Optimal pH and temperature are 7 and 65°C, respectively. This high optimal temperature is surprising for such an enzyme, produced by a mesophilic human gut bacterium. However, 28 % of its maximal activity is retained at 37°C, which is probably sufficient to insure its role for carbohydrate foraging in the gut. Moreover, a total loss of activity was observed after only 25 min and 90 min of incubation at 65°C and 50°C, respectively, while only 50 % of initial

activity was retained after 27h at 37°C. Thus, despite its high optimal temperature, this enzyme cannot be considered as thermostable.

BACOVA_03624: an hydrolase, phosphorylase or transglycosidase?

As explained in the introduction, we hypothesized that BACOVA_03624 could be a hydrolase, the Glu144 residue corresponding to Uhgb_MP Tyr103 playing the role of base. To investigate this question, kinetic analyses were performed at optimal pH and temperature, by varying *p*NP-β-Gal*p* and P_i concentrations ranging from 0 to 10mM. First, the enzyme displays a typical Michaelis-Menten behavior both in presence and in absence of P_i (Figure 4), contrary to inverting GPs, which follow ordered or random Bi Bi kinetics involving the formation of a ternary complex of two substrates (the glycosyl donor and acceptor) with the enzyme. Moreover, as revealed by V_{max} and K_M values, BACOVA_03624 is nearly as efficient to breakdown *p*NP-β-Gal*p* in absence of P_i as in presence of 10mM P_i. This result reveals that, as anticipated by primary structure analysis, BACOVA_03624 is a hydrolase, the first characterized one in the GH130 family.
Nevertheless, the kinetic profiles are slightly, but positively affected in large P_i excess (P_i concentrations ≥ 5 mM), indicating that BACOVA_03624 may also act as a phosphorylase in these conditions.

Thus we wanted to know which kinds of reactions are involved in glycoside breakdown, in absence and in presence of 10mM P_i. Reactions were performed at 37°C, the optimal temperature for BACOVA_03624 use, being the best compromise between enzyme activity and stability. As explained further in this paper, the genomic context of the BACOVA_03624 encoding gene indicates that this enzyme may be involved in N-glycan breakdown. We thus tested its activity on β-D-Gal*p*-1,4-D-Glc*p*NAc (*N*-acetyllactosamine), a disaccharide present in human hybrid and complex N-glycans. As shown on Figure 5, the commercialized form of this disaccharide is not pure, limiting accurate quantification of reaction products. However, in absence of P_i, their HPAEC-PAD profiles before and after BACOVA_03624 action are clearly different, revealing the release of GlcNAc and galactose, in equivalent amounts (0.6mM), corresponding to 6 % of the initial β-D-Gal*p*-1,4-D-Glc*p*NAc concentration (10mM). BACOVA_03624 is thus capable of hydrolyzing this N-glycan core oligosaccharide, even slowly. Indeed, after 4h of reaction, substrate conversion yield only reached 2%.
In presence of P_i, 1.96 times more GlcNAc (1.18mM, corresponding to 11.8% of the initial donor amount) was produced at the same reaction time. However, only 1.8 (1.08mM) times more galactose were produced, the deficit in galactosyl units being due to the apparition of slight amounts of α-D-galactose-1-phosphate (Gal1P), corresponding to 1% of the initial donor amount. In presence of P_i, BACOVA_03624 thus catalyses both β-D-Gal*p*-1,4-D-Glc*p*NAc hydrolysis, which would count for 91.5% of the reaction, and phosphorolysis, which

would count for at least 8.5% of the reaction. However, it is difficult to exactly calculate the ratio between hydrolysis and phosphorolysis, as reverse-phosphorylation products, coming from galactosyl transfer from Gal1P onto Gal, GlcNAc, β-D-Gal*p*-1,4-D-Glc*p*NAc or one of the contaminant detected in the commercial substrate sample, may be hidden in the forest of peaks detected in the HPAEC-PAD profiles.

To check the specificity of BACOVA_03624 towards mannosyl donors with real glycosides, which are often better recognized by GH130 enzymes that *p*NP-substrates, the BACOVA_03624 activity towards manno-oligosaccharides of polymerization degree from 2 to 5 and was then tested. Contrary to what was observed onto *p*NP-β-Man*p*, this enzyme was able to breakdown these substrates, with various efficiencies. The best substrate conversion yield (28%) was obtained with mannotriose in presence of P_i, while it only reached 1, 16 and 4% with mannobiose, mannotetraose and mannopentaose, respectively. In these conditions, mannotriose was converted into mannobiose, mannotetraose and α-D-mannose-1-phosphate (Man1P). As no trace of contaminant of DP >4 was detected in the initial reaction medium, mannotetraose was produced by mannosyl transfer onto mannotriose. However, as no kinetic analysis of this reaction was performed at this stage of the study, it is difficult to know which one between mannotriose or Man1P played the role of mannosyl donor. Based on the amount of released Man1P, in this case, phosphorolysis would thus count for at least 3.6% of the mannotriose breakdowing reaction.

Without P_i, only 11% of mannotriose was consumed, and converted into mannobiose and manotetraose. Here, presence of mannotetraose testifies of the BACOVA_03624 ability to catalyse transglycosylation reaction. However, no transglycosylation or reverse-phosphorolysis product of DP >4 was observed, regardless of the manno-oligosaccharide substrate. For this reason, and because mannopentaose is not a good substrate for this enzyme, we think that BACOVA_03624 active site topology would prevent accommodation of oligosaccharides of DP >4. This is in accordance with what was hypothesized from structural analysis of GH130 enzymes (Ladevèze et al, 2015). Indeed, in enzymes belonging to the GH130_NC cluster, a particularly long loop (corresponding to the BACOVA_03624 E94-R104 loop), would clog the catalytic tunnel exit. Solving the BACOVA_03624 crystallographic structure in complex with manno- or galacto-oligosaccahrides will allow to define the -1 to +3 subsites, and, potentially, confirm this hypothesis.

Reverse phosphorolysis reactions catalysed by BACOVA_03624

The study of *p*NP-β-Gal*p*, β-D-Gal*p*-1,4-D-Glc*p*NAc and manno-oligosaccharides degradation by BACOVA_03624 revealed that this original enzyme mainly acts as a hydrolase, but is also able to catalyze glycoside phosphorolysis. It also gave insights into specificity of its -1 subsite towards galactosyl, but also, to a much less extent, towards mannosyl residues. However, information regarding the specificity of the +1 subsite were biased by the disaccharide (or the *p*NP-glycosyl compound used as donor) conformation itself, and by affinity of the -1

subsite towards the residue at its non-reducing end. To characterize the BACOVA_03624 +1 subsite specificity, we used its ability to catalyze phosphorolysis, and thus, reverse-phosphorolysis. First, contrary to what is observed for others GH130, like Uhgb_MP, no significant hydrolysis of Gal1P or Man1P was observed when no carbohydrate acceptor was added to the reaction medium.
Either using Gal1P or Man1P, several monosaccharides acted as good acceptors for BACOVA_03624, given that in their presence, 10 to 38% of glycosyl phosphate was consumed (Table 1) to generate one or several glycosylated products, visible on HPAEC-PAD profiles at higher retention times than that of the initial acceptors. With Gal1P as donnor, lyxose, glucose, mannose and GalNAc were the best recognized acceptors, yielding to one unique glycosylated product for each of them. In presence of GlcNAc, only 6% of Gal1P was consumed, to produce traces of β-D-Gal*p*-1,4-D-Glc*p*NAc. This weak specificity of the +1 subsite towards GlcNAc indicates that, *in vivo*, β-D-Gal*p*-1,4-D-Glc*p*NAc may not be the preferred substrate of BACOVA_03624. With Man1P as donor, the +1 subsite specificity is still broader, as fucose, GlcNAc, GalNAc, glucose, fructose, galactose and lyxose behave as the best acceptors, mannose arriving in the eighth position. For all of them, excepted lyxose, series of glycosylated acceptors of DP from 2 to 4 were observed. Accommodation of a mannosyl residue in the -1 subsite thus seems to positively affect accessibility and or acceptor accommodation in the +1 subsite, compared to what happens when a galactosyl residue is bound in the -1 subsite.

Because of the flexibility of BACOVA 03624 -1 and +1 subsites, it is difficult, only based on *in vitro* biochemical characterization of this polyvalent enzyme, to understand what would be its role for carbohydrate foraging in the human gut. In this context, analysis of the genomic context of the BACOVA_03624 encoding gene, and transcriptional analysis of *Bacteroides ovatus* ATCC 8483 on various galactosides and mannosides, appears as the best strategy to investigate the function of the different GH130 enzymes displayed by this bacterial strain.

Genomic context of the BACOVA_03624 encoding gene

B.ovatus ATCC 8483 genome harbors 7 GH130 encoding genes, each belonging to a physically-linked set of genes referred to as polysaccharide utilization loci (PUL) (Terrapon and Henrissat 2014) (Figure 6). No data are still available regarding the function of the predicted PUL94, while the 6 others appear as experimentally validated in the Polysaccharide Utilization Loci DataBase (PULDB) (Terrapon and Henrissat 2014), even if 3 of them are not listed as up- or down-regulated by N-glycans or plant mannans (Figure 7) in Martens *et al.*, (Martens et al. 2011), the paper cited in the PULDB regarding their experimental validation (Table 2).

This is the case of PUL82, which contains the gene encoding BACOVA_03624, the polyspecific enzyme targeted in the present study. In addition to the BACOVA_03624 encoding gene, PUL82 contains 2 GH92, 1 GH125 and 2 GH76 encoding genes. Despite the fact that these *B. ovatus* GH are still not biochemically characterized, the known activities for their families (α-1,2,3,4, 6-mannosidases, α-1,6-mannosidase and endo-α-1,6-mannanase, respectively) seem to be suited for degradation of human high-mannose, hybrid or complex N-glycans. Based on -1 and +1 subsite specificities of BACOVA_03624, deduced from analysis of its hydrolytic, phosphorolytic and reverse-phosphorolytic activities, in this case this enzyme may act by breakdowning the core oligosaccharide β-D-Man*p*-1,4-D-Glc*p*NAc, even if we still have to test this substrate as glycosyl donor to validate its activity towards the β-1,4 linked linkage. It may also act by breakdowning β-D-Gal*p*-1,4-D-Glc*p*NAc in hybrid and complex N-glycans. However in mammals, this disaccharide is capped by an α-1,6 linked sialyl residue, and PUL82 does not contain any putative sialidase encoding gene.

In addition, mammals N-glycans contain a maximum of 3 α-1,6-linked mannosyl residues, while GH76 endo-1,6 mannases would require a much longer α-1,6-linked mannoside chain as substrate (Maruyama and Nakajima 2000). These long α-1,6-mannan chains are found in yeast N-glycans, which also contain α-1,3 and α-1,2 linked-mannosyl residues, Man*p*-α-1-phosphate-6-*O*-Man*p* motives, and sometimes, like in the pathogenic yeast *C. albicans*, β-1,2 linked mannosides. Recently, the involvement of *B. thetaiotamicron* VPI-5482 MAN-PUL2 (BT3773–92), which resembles PUL82, in yeast N-glycan metabolization, was demonstrated. MAN-PUL2 (BT3773–92) also contains GH125, GH92, GH76 and GH130 sequences. However, the role of the BT GH130 enzyme (BT3780, which presents 87.47% identity with BACOVA 03624) in high mannose N-glycan (HMNG) catabolism has still not been studied.
In the present study, we showed that BACOVA_03624 is able to synthesize a β-linked Gal-GlcNAc disaccharide, but also β-linked Manno-oligosaccharides. BACOVA_03624 may thus be able to act both onto β-D-Man*p*-1,4-D-Glc*p*NAc and β-1,2-linked-manno-oligolosaccharides, which can be found in *C. albicans* HMNG. NMR analysis of the reverse-phosphorolysis products obtained from the GlcNAc and Man acceptors, and testing hydrolysis of these particular substrates should soon allow one to validate this hypothesis. Given PUL82 GH content and BACOVA_03624 specificity, we thus suspect that the principal PUL82 target may be yeast N-glycans, rather than human N-glycans or plant mannans.

GH130 targeted transcriptomic analysis of Bacteroides ovatus ATCC 8483

To investigate the physiological function of BACOVA_03624 in the human gut, we initiated a quantitative transcriptomic study, targeting the 7 GH130 of *B. ovatus* ATCC 8483. Only preliminary results are presented here, at this stage of the study.

First, *B. ovatus* ATCC 8483 was able to grow both on β-mannan and on a *S. cerevisiae* highly N-glycosyated protein, used as only carbon sources (Figure 8). This is in accordance with the results of Martens *et al*., (Martens et al. 2011) who observed growth on galactomannan, glucomannan, and many other plant and host-derived glycans, and with those of Cuskin *et al*. (Cuskin et al. 2015), who observed growth of 25 *B. ovatus* strains on *S. cerevisiae* mannan. Expression of most of *B. ovatus* GH130 encoding genes was induced by β-mannan (Figure 9). Overexpression of the two genes encoding the uncharacterized enzymes classified in the GH130_1 subfamily (BACOVA_02634 and 02090), belonging to PUL46 and PUL42, confirms our previous hypothesis regarding involvement of GH130_1 enzymes in plant mannan, galactomannan and glucomannan catabolism, acting in concert with enzymes of the GH26 and GH36 families. These results are in accordance with the induction of PUL42 and PUL86 during *in vivo* or *in vitro* growth on plant polysaccharides (Martens et al. 2011). However, the weak overexpression of BACOVA_02634 may explain why no significant induction of PUL42 was previously observed during *in vitro* growth on plant glycans, while it was induced in plant fed mice (Martens et al. 2011).

Surprisingly, expression of the gene encoding BACOVA_04885, an uncharacterized GH130_NC enzyme, is highly induced by β-mannan. It belongs to the predicted PUL94, surrounded by 3 GH92, 1 GH31, 1 GH78, 1 GH2 and 1 GH38 encoding genes. The activities related to these families are all related to degradation of N-glycans. However, the GH2 and BACOVA_04885 enzymes may either be involved in the breakdown of the β-D-man*p*-1,4-D-Glc*p*NAc core oligosaccharide of N-glycans, or of plant β-1,4-mannosides. Similarly, expression of BACOVA_04110 (GH130_2) encoding gene was induced by β-mannan, but moderately, while it is surrounded in PUL85 by GH18 and GH92 encoding genes, related to N-glycan breakdown. PUL85 was up-regulated in plant fed mice, while it is homologous to a *B. thetaiotaomicron* PUL associated with host glycan utilization *in vitro* and *in vivo*. Here again, BACOVA_04110 and maybe also the uncharacterized associated GH18 enzyme, could be involved both in β-mannan and in N-glycan breakdown. Based on these transcriptomic data, which complete previously published ones (Martens et al. 2011), we thus suspect that PUL94 and PUL85 may be involved in the breakdown of both plant cell wall or eukaryotic N-glycans, depending on physiological conditions.

Expression of the BACOVA_02161 (GH130_NC) and BACOVA_02887 (GH130_NC) encoding genes belonging to PUL44 and 58, was also induced by β-mannans. In addition to BACOVA_02161, PUL44 contains 9 GH encoding genes, all related to plant cell wall degrading functions, while PUL58 contains one unique predicted CAZyme corresponding to BACOVA_02887. Even appearing as experimentally validated in the PUL Database, these PULs were not mentioned in Martens *et al*., as being induced by plant glycans.

Finally, by using a ≥10-fold cutoff threshold, the sole GH130 encoding gene which was not upregulated during growth on plant mannan was that encoding BACOVA_03624 (PUL82), the target described in the present paper. This result shows that despite its ability to breakdown β-1,4-manno-oligosaccharides *in vitro*, BACOVA_03624 is not involved in the metabolisation of plant mannosides.

Conclusion

In this paper, the biochemical characterization of BACOVA_03624, one of the seven GH130 enzymes of *B. ovatus* ATCC 8483, revealed the polyvalence of this original enzyme. First, BACOVA_03624 is today the first characterized hydrolase of the GH130 family. To our knowledge, it is also the sole known enzyme that is able to catalyze the breakdown of β-D-galactosides and β-D-mannosides both by hydrolysis and phosphorylation, or their synthesis by transglycosylation and reverse-phosphorolysis. The resolution of its crystallographic structure in complex with its numerous ligands will allow identifying the molecular bases of its fascinating properties. It will also offer new perspectives to predict GPs specificity towards glycoside donors, and to modulate it by rational engineering.
In addition, integration of the biochemical and genomic data presented here suggests that BACOVA_03624, and its associated PUL, would be involved in metabolization of yeast glycans, or host glycan. The comparative study of GH130 targeted transcriptomic data acquired during growth on plant and yeast mannans will allow us to refine the understanding of the role of BACOVA_03624, and more generally, of the gut symbiont *B. ovatus*, in carbohydrate foraging.

Materials and methods

BACOVA_03624 cloning, production and purification
The BACOVA_03624 protein sequence deposited on the NCBI was used as template for gene synthesis with codon optimization for *E. coli* heterologous production, and was subsequently cloned into the pET28a vector to allow for production of an N-terminal hexahistidine tagged protein (Biomatik, Cambridge, Ontario, Canada). To allow heterologous BACOVA_03624 production, *E.coli* BL21-AI cells (Invitrogen) transformed with the BACOVA_03624::pET28a plasmid were cultured at 20°C for 24 hours in 200mL ZYM-5052 auto-induction medium (Studier 2005) supplemented with 50µg/mL Kanamycin, inoculated at OD_{600nm} of 0.1. Cells were harvested and resuspended in 20mM Tris-HCl, pH7.0, 300mM NaCl, and lysed by sonication. Soluble lysate was applied to a TALON resin loaded with cobalt (GE Healthcare) equilibrated in 20mM Tris-HCl, pH7.0, 300mM NaCl. After column washing with 15 volumes of the same buffer supplemented with 10mM imidazole, the protein was eluted in 20mM Tris-HCl, pH7.0, 300mM NaCl, 150mM imidazole. The purity of the purified BACOVA_03624 protein was evaluated higher than 95 % by SDS page electrophoresis using Mini-PROTEAN®

TGX stain free™ Precast Gel (BIO-RAD). After migration, proteins were detected on a Gel Doc™ EZ imaging system (BIO-RAD). Protein concentrations were determined by spectrophotometry using a NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The purification yield was 9.2mg of protein per liter of culture. The calculated extinction coefficient of the purified BACOVA_03624 fused to an N-terminal His(6) tag is 75477 $M^{-1}.cm^{-1}$.

Enzyme assays

Temperature profile was obtained by determining BACOVA_03624 specific activity on 1mM p*NP*-β-D-galactopyranose (Sigma Aldrich, ref N1252) and 10mM inorganic phosphate, using 0.1mg/ml purified protein, in 20mM Tris-HCl, pH7.0. Enzyme stability was determined by monitoring residual specific activity of purified BACOVA_03624 samples stored at increasing temperatures (4, 37, 50 and 65°C), in 20mM Tris-HCl, pH7.0, at 65°C, by using 1mM p*NP*-β-D-galactopyranose and 10mM inorganic phosphate as substrates. The p*NP* release was monitored at OD_{405nm} on a Eon™ Microplate Spectrophotometer (BioTek™) after basification of the samples by addition of 150µL Na_2CO_3 1M onto a 50µL reaction sample, in order to ensure proper measurement of the released *p*NP concentration.

HPLC analysis

BACOVA_03624 phosphorolysis, reverse-phosphorolysis, hydrolysis, and transglycosylation properties were assessed by measuring product release from 10mM Mannobiose, Mannotriose, Mannotetraose, Mannopentaose, *N*-acetyl-lactosamine, or 0,5g/L galactan in presence or absence, respectively of 10mM P_i, using 0.1mg/ml of purified protein in 20mM Tris-HCl pH7.0, at 37°C.

All reverse-phosphorolysis reactions were carried out using 10mM of Gal1P or Man1P as donors, and 10mM acceptors .(D-mannose, D-glucose, D-fructose, D-galactose, L-xylose, *N*-acetylglucosamine, L-arabinose, L-fucose, *N*-acetylgalactosamine, (Sigma)).

Timepoints were taken at 0, 1h, 3h and 72h, and boiled 5min at 95°C to stop reaction. Reaction product analysis and quantification was performed by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Carbohydrates were separated on a 2x250mm Dionex Carbopac PA100 column. A gradient of sodium acetate (from 0 to 150mM in 15 min) followed by a descending one from 300 to 0mM in 5min, in 150mM NaOH was applied at a 0.25mL.min^{-1} flowrate. Detection was performed using a Dionex ED40 module with a gold working electrode anda Ag/AgCl pH reference.

B. ovatus culture conditions

B. ovatus ATCC 8483 was cultivated anaerobically at 37°C, in a chemically defined medium (Rasmussen et al. 1988) that was modified (no supplementation with phenylpropionic and phenylacetic acids) and which contained 0.1% of glucose, 0.1% of 1,4-β-D-mannan

(Megazyme, Bray, Ireland), or 0.1% of the glycoprotein invertase from *S. cerevisiae* (Reference I4504, Sigma Aldrich, Saint-Quentin Fallavier, France). The liquid media were prepared and dispensed in Balch tubes using strictly anaerobic techniques (Hungate 1969). A 2.5% (v/v) inoculum of culture pre-adapted on each substrate was used for inoculation. Bacterial growth on glucose and on the glycoprotein invertase from *S. cerevisiae* was recorded directly in Balch tubes by measuring optical density at 600nm using a Jenway 6320D spectrophotometer. As suspension of 1,4-β-D-mannan interfered with OD_{600} measurements, growth on this substrate was followed by determining bacterial protein concentration using the Bradford Protein Assay (Bradford 1976). Two to three independent cultures were performed for each substrate condition. Cells were harvested at mid-log phase for subsequent gene expression analysis.

RNA isolation, reverse transcription and relative quantitative PCR

Total RNAs were isolated from cultures harvested at mid-log phase using a modified guanidinium–phenol–chloroform procedure (Béra-Maillet et al. 2009). Briefly, for each substrate, bacterial cultures (2 x 8ml) were centrifuged for 15min at 3,000 × g at 4°C. The pellets were resuspended in 9ml of a RNA-E solution containing solution D, water saturated phenol, sodium acetate 0.2M pH4.0 and 2-mercaptoethanol (1:0.1:1:0.007). Cells were then disrupted by bead-beating for 1min with 0.1g zirconia beads (0.1mm) followed by a 2-min incubation at 60°C. These two steps were then repeated once. After addition of 3.75ml of chloroform, the samples were briefly mixed, incubated for 15min on ice and centrifuged (12,000 × g, 20min, 10°C). The RNAs contained in the aqueous supernatants (approximately 6ml) were precipitated with 0.25 volume isopropanol and washed with 1 volume 75% cold ethanol in DEPC-treated water. Total RNAs were solubilized in 100µl of DEPC-treated water. Genomic DNA was removed by Turbo DNA-Free (Ambion, France) for 30min at 37°C. RNAs were quantified using a ND-2000 spectrophotometer (Nanodrop Technologies).

Total RNAs (1µg) were reverse-transcribed into cDNAs using random hexamer primers (Invitrogen, France) and 200U SuperscriptII Rnase H^- reverse transcriptase (Invitrogen, France) according to the manufacturer's recommendation. For each RNA sample, a negative RT (no addition of reverse transcriptase) was performed and used as negative control in subsequent qPCRs.

The relative expression of target genes (7 GH130 genes) in the mannan condition versus the glucose condition was determined by quantitative PCR using a Rotor-Gene Q real-time PCR detector (Qiagen, France) and Quantifast SYBR Green PCR mastermix (Qiagen, France) using supplier's instructions. The designed specific primers are listed in Table 3 and were purchased from Eurogentec (Seraing, Belgium). Reaction mixtures were set up to a final volume of 20µL using 1.2µL of each primer (5µM), 10µL of QuantiFast SYBR Green PCR mastermix (Qiagen) and 2µL of cDNA template (0.5ng/µL). Thermal cycling consisted of an

initial denaturation step (95°C for 5min) followed by 45 cycles of amplification (95°C for 15s, 60°C for 45s), with a single fluorescence measurement at the end of each cycle, then finished by the melting curve program (65–95°C). Negative controls with RNA not reverse-transcribed were run with every assay to rule out DNA contaminations. At this stage of the study, the fold change in gene expression (β-mannan versus glucose) was calculated from one biological replicate using two technical qPCR replicates according to Livak and Schmittgen, 2001 (Livak and Schmittgen 2001) using the RpoD reference gene for normalization. The primer set used for RpoD is given in Table 3.

Figure 1:

Phylogenetic tree of the GH130 family, revisited in January 2015. In this radial phylogram, branches corresponding to subfamilies GH130_1 and GH130_2 are colored in blue and red, respectively. Branches in black correspond to sequences that are not classified into any subfamily (GH130_NC). The protein names of all the structurally and functionally characterized GH130 members appear explicitly. For enzymes whose structure is known, a picture of the 3D structure is shown, with an insert indicating the corresponding pdb codes and ligands. Red star: enzymes for which the structure is known, in the *apo* form, but for which no functional data is available.

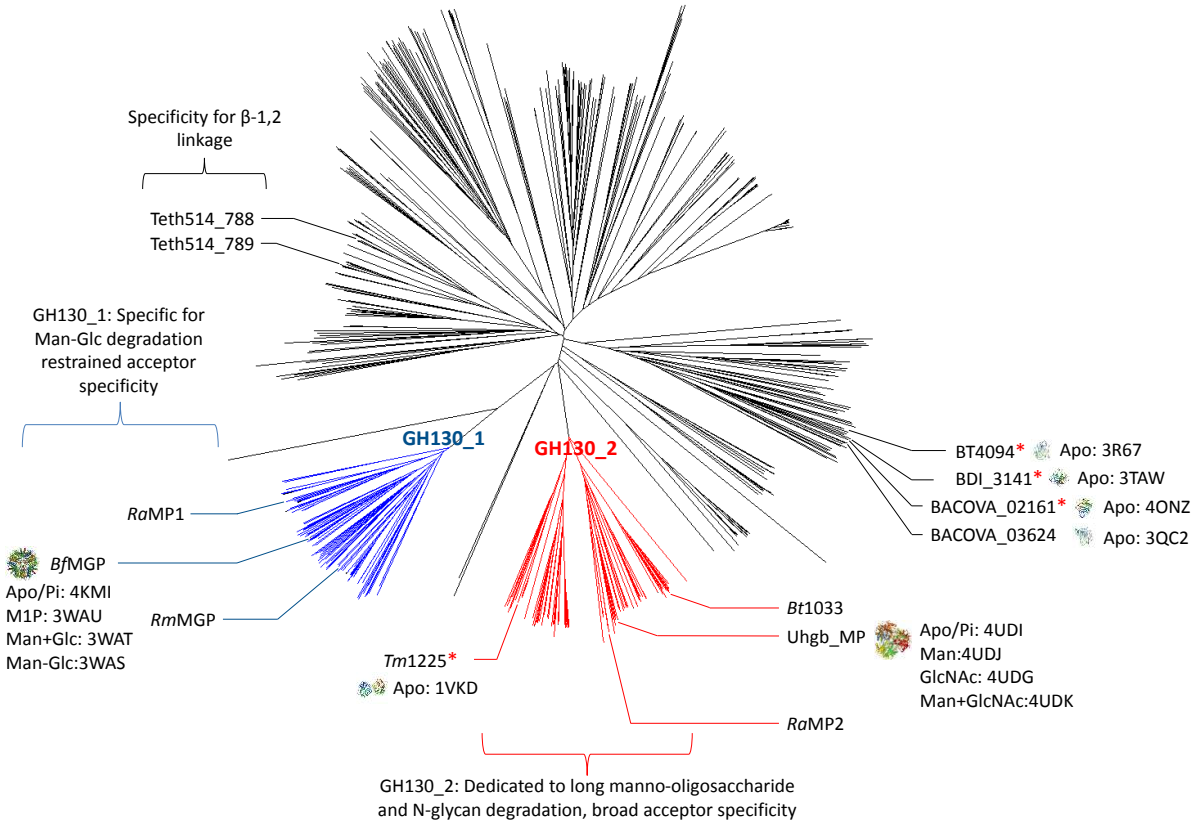


Figure 2:

Multiple alignment of GH130 characterized protein sequences and others GH130_NC from human gut inhabitants, showing the conservation in the GH130_1 and GH130_2 GPs subfamilies, of residue Y103 (which plays the role of nucleophile for glycoside phosphorolysis), as well as R150, R168, and N151 (Uhgb_MP numbering), which maintain P_i in the GPs active site. On the contrary, in most of GH130_NC sequences, Y103 is replaced by an acidic residue (E144, BACOVA 03624 numbering), which would play the role of second catalytic residue in the case of glycoside hydrolases like BACOVA_03624. In all cases, residue D104 (D145 BACOVA_03624 numbering) is the catalytic proton donor required for hydrolysis and phosphorolysis.

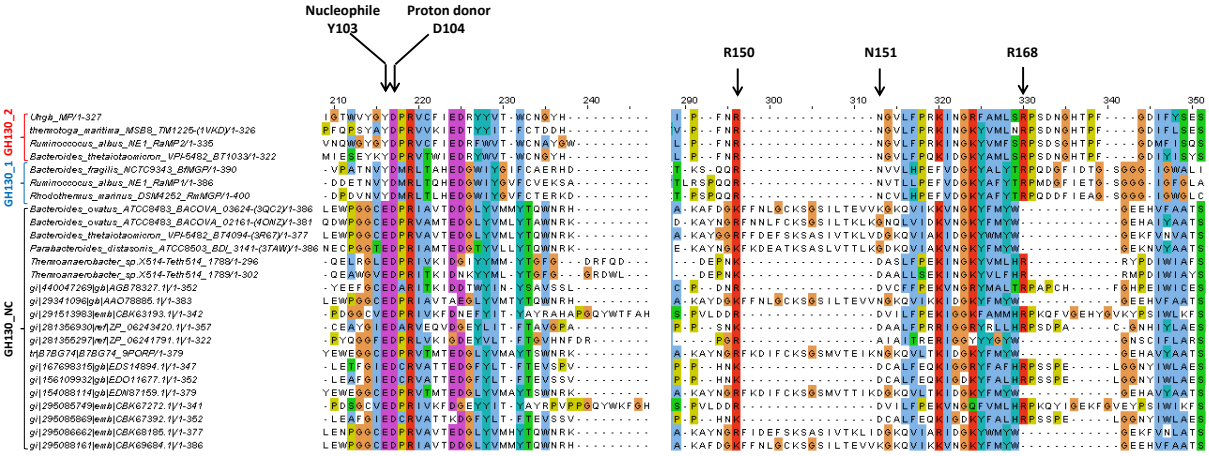


Figure 3:

Effect of temperature on BACOVA_03624 activity, measured in presence of 1mM *p*NP-β-D-Gal*p* and 1mM P_i. A) Temperature dependency of BACOVA_03624 activity. B) BACOVA_03624 stability at 4, 37, 50 and 65°C.

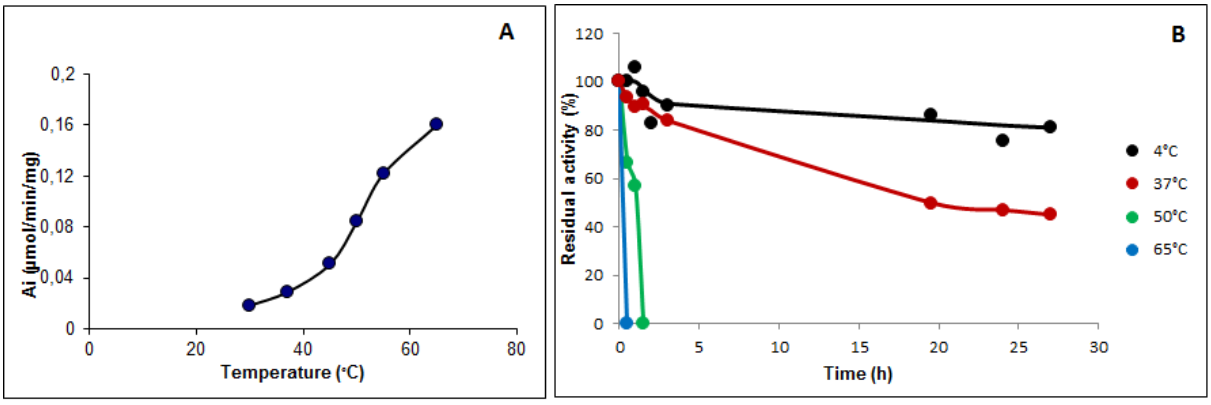


Figure 4:

Double-reciprocal plots for *p*NP-β-D-Gal*p* degradation by BACOVA_03624. The initial velocities of *p*NP release were measured at various concentrations of *p*NP-β-D-Gal*p* in

absence or in presence of 10mM P_i (A), or from 10mM *p*NP-β-D-Gal*p* at various concentrations of P_i (B). Data are the mean values ± S.D. for three independent experiments.

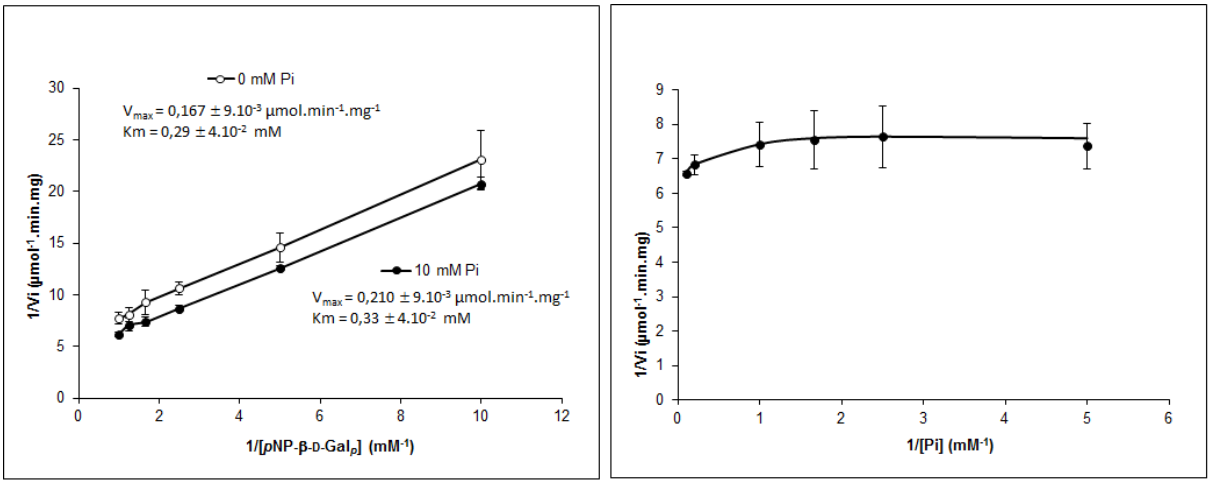


Figure 5:

HPAEC-PAD analysis of the reaction products obtained from 10mM β-D-Gal*p*-1,4-D-Glc*p*NAc (A) and mannotriose (B), in absence and in presence of 10mM P_i. Final reaction time corresponds to 72h.

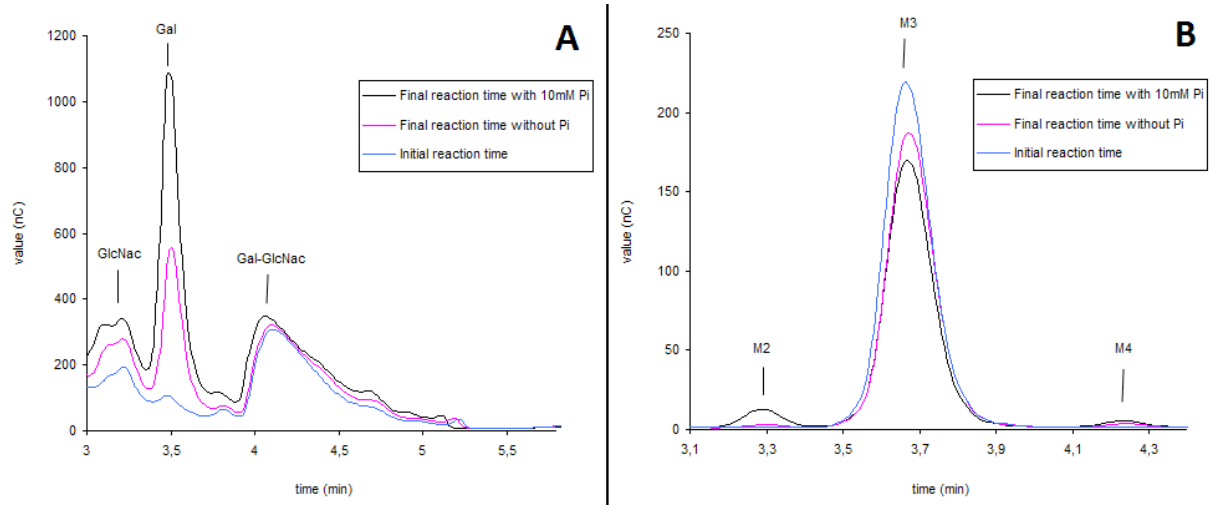


Figure 6:

Polysaccharide utilization loci (PULs) of *Bacteroides ovatus* ATCC 8483, containing GH130 encoding genes. GH130 subfamily classification is represented by the bar color: blue, GH130_1; red, GH130_2; black, GH130_NC. The neighbouring CAZyme encoding genes part of each PUL are indicated as gray bars. PUL numbers were retrieved from the Polysaccharide Utilization Loci DataBase (PULDB), in which all appear as experimentally validated, excepted PUL94.

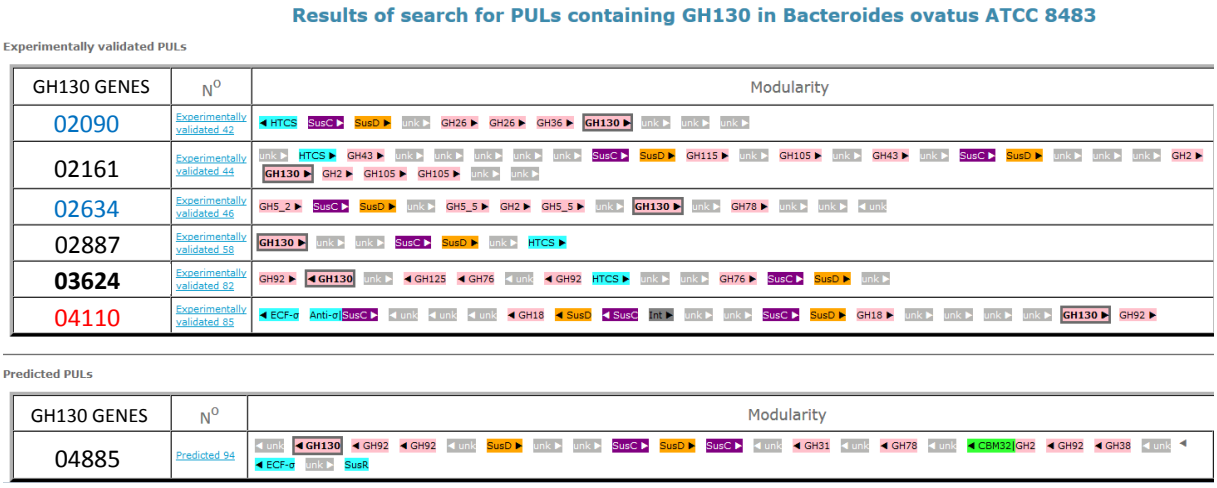


Figure 7:

Schematic representation of the main mannosides present in the human gut.

A. Plant β-mannan, an hemicellulosic component. B. Human N-glycans, showing the different types of structures found on mature glycoproteins formed from a common $Glc_3Man_9GlcNAc_2$ precursor. C. Yeast very high mannose containing N-glycan (yeast mannan); the structure depicted here is that of *C. albicans*, holding additional β-1,2-linked mannosyl ramifications on the long lateral α-1,6 mannosyl branches.

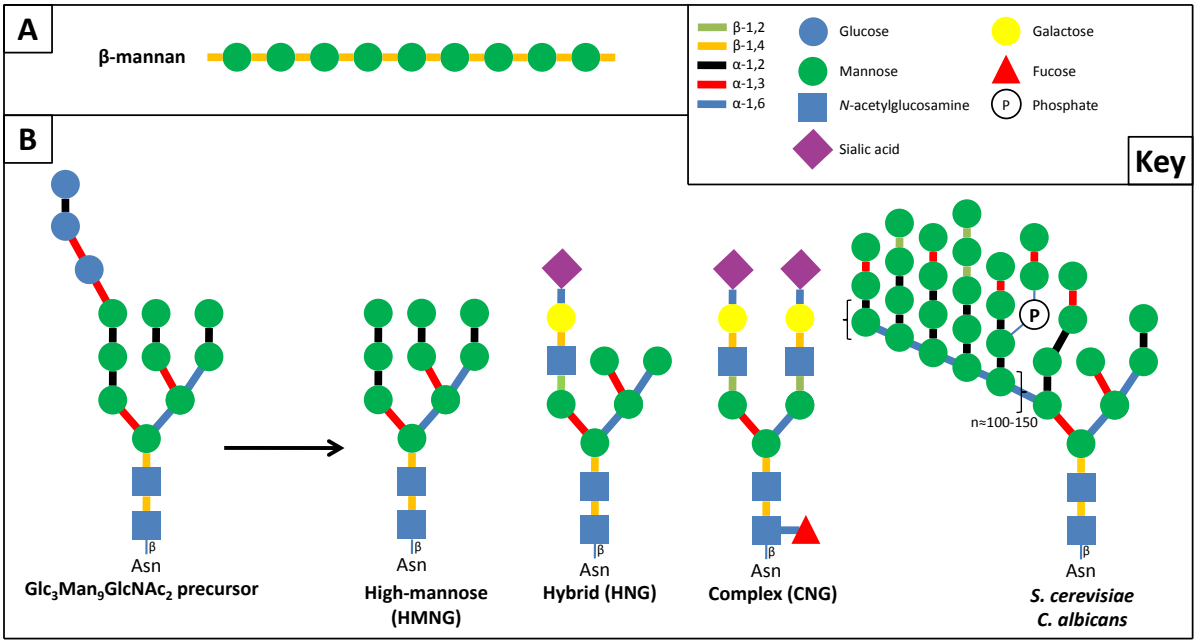


Figure 8:

Growth of *B. ovatus* ATCC 8483 on β-mannan and on the *N*-glycosylated *S. cerevisiae* invertase protein used as sole carbon sources. A. Growth curves of *B. ovatus* ATCC 8483 on invertase and glucose. B. Growth of *B. ovatus* ATCC 8483 on β-mannan monitored by protein expression over time.

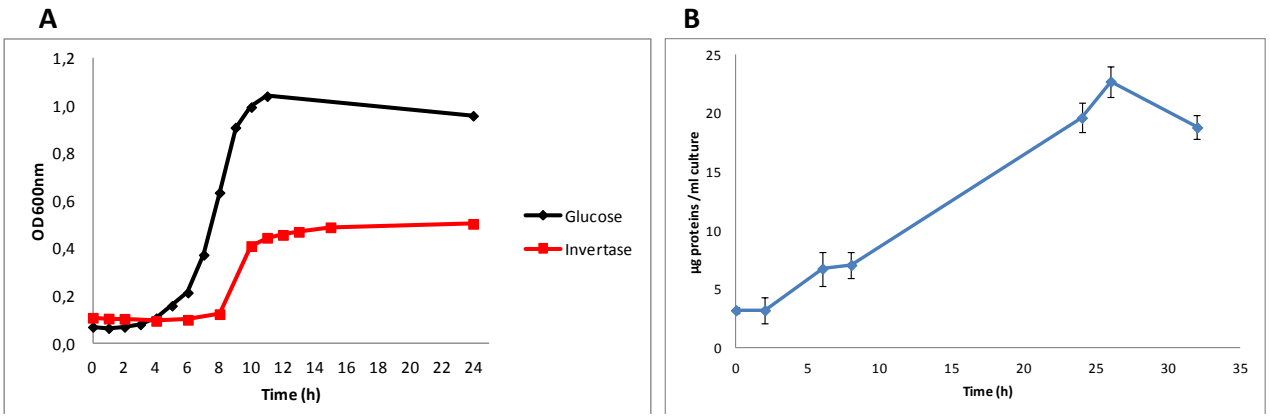


Figure 9:

Expression of the 7 GH130 encoding genes of *B. ovatus* ATCC 8483, during growth on β-mannan as only carbon source.

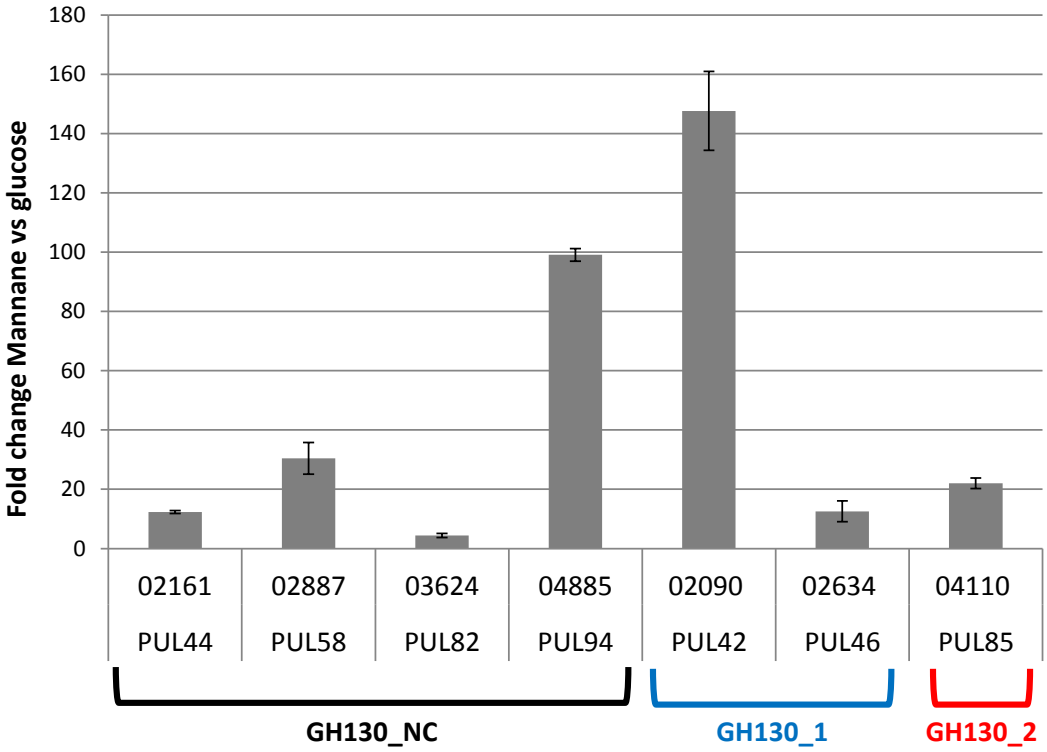


Table 1:

Reverse-phosphorolysis yields after 72 h reaction in presence of 10mM α-D-mannopyranose-1-phosphate (α-D-Man*p*1P) or α-D-Galactopyranose-1-phosphate (α-D-Gal*p*1P) as glycosyl donors, and 10mM of various acceptors.

Glycosyl acceptor	α-D-Gal*p*1P conversion yield (%)	α-D-Man*p*1P conversion yield (%)
/	0	0
N-acetylgalactopyranosamine	11,7	20,5
N-acetylglucopyranosamine	6,7	35,4

L-Arabinose	6,5	13,2
D-Fructose	1,0	17,5
L-Fucose	1,1	38,9
D-Galactose	0	16,3
D-Glucose	27,8	18,2
D-Lyxose	36,9	13,5
D-Mannose	20,3	5,2

Table 2:

Genomic environment and probable function of the *Bacteroides ovatus* ATCC 8483 GH130 genes and their associated PULs. HMNG : High mannose N-glycans.

Bacova-GH130 genes	Sub-family	PUL number	Genomic environment	Known specificities of the GH families	Expression induced byβ-mannan*	PUL expression induced by plant-glycans according to Martens et al. (2011)	Functional specificity
03624	GH130_NC	Exp82	GH92 (x2) GH125 (x1) GH76 (x2)	α-1,2-3-4-6-mannosidase α-1,6-mannosidase endo-α-1,6-mannanase	no	not listed**	Yeast and/or human N-glycans
02090	GH130_1	Exp42	GH26 (x2) GH36 (x1)	β-man,-xyl, -glucanase α-gal, -N-acetylgalactosaminidase	highly	galactomannan glucomannan	Plant mannans*
04885	GH130_NC	Pred94	GH92 (x3) GH31 (x1) GH78 (x1) GH2 (x1) GH38 (x1)	α-1,2-3-4-6-mannosidase α-gluc, -xyl, -mannosidase α-L-rhamnosidase β-gal, -man,-glucuronidase, α-Arabinofuranosidase α-mannosidase	highly	not listed**	Host N-glycans Plant mannans*
02161	GH130_NC	Exp44	GH43 (x2) GH115 (x1) GH78 (x1) GH2 (x2) GH105 (x3)	β-gal, -xylosidase;arabinanaseα-arabinofuranosidase; α-glucuronidase α-L-rhamnosidase β-gal, -man,-glucuronidase,exo-β-glucosaminidase; α-arabinofuranosidase rhamnogalacturonyl-,glucuronyl-hydrolase	moderately	not listed**	Plant mannans*
02887	GH130_NC	Exp58	no	no	moderately	not listed**	Plant mannans*
02634	GH130_1	Exp46	GH5 (x3) GH2 (x1) GH78 (x1)	cellulase, gluca-, xyl-, mannanase; β-gluco, -mannosidase β-gal, -man,-glucuronidase, α-Arabinofuranosidase α-L-rhamnosidase	moderately	induced *in vivo* by plant fed mice	Plant mannans*
04110	GH130_2	Exp85^{BT}	GH18 (x1) GH92 (x1)	Chitinase; β-N-acetylglucosaminidase α-1,2-3-4-6-mannosidase	moderately	induced *in vivo* by plant fed mice	Host N-glycans Plant mannans*

^{BT}; PUL 85 is homologous to a *B. thetaiotaomicron* PUL associated with host glycan utilization *in vitro* and *in vivo* (Martens et al., 2011).

*; Experimentally validated in this study

Not listed**, the GH130 encoding gene and the corresponding PUL were not listed in Martens et al. (2011) as being up- or down-regulated during growth on plant substrates, nor *in vivo* in plant fed mice.

Table 3:

List of primers used for qRT-PCR analysis.

Targeted GH130 enzyme	Forward primers (5'-3')	Reverse primers (5'-3')	PCR Product size (bp)
BACOVA_02090	CCAGCAAGAATGTGCCTTTGG	CGCCAGAAGACAGGAGTATGG	94
BACOVA_02634	GCAGAACATGCACCACTCATC	ATTGATCTTGATGGCTCCCGT	114
BACOVA_04110	CTGTCGGATGTACGGACGTTA	GCCTGCAAAACCATCCTCAA	125
BACOVA_02161	ATGTGGATTTACCAATGCGCC	TCCTTCAGGACGCAAGAATGG	91
BACOVA_02887	AAAGCGGGAATCAACGGAATG	CTGTATAGGTCTCTCTGCGACG	108
BACOVA_03624	GATGACCTGTACTGCCTACGG	CAAAACCACCGAAAGCCCAAT	77
BACOVA_04885	GCTGCCGCATTAATCACTTCC	TTAATTCCTTCCGGACGCTGG	95
BACOVA_00615(RpoD)	CGGCCGTGAAGACCTGATTA	ACGAAACGAAGATTGGCACG	111

Bibliographic references:

Béra-Maillet C, Mosoni P, Kwasiborski A, Suau F, Ribot Y, Forano E (2009) Development of a RT-qPCR method for the quantification of Fibrobacter succinogenes S85 glycoside hydrolase transcripts in the rumen content of gnotobiotic and conventional sheep. J Microbiol Methods 77:8–16. doi: 10.1016/j.mimet.2008.11.009

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254.

Chiku K, Nihira T, Suzuki E, Nishimoto M, Kitaoka M, Ohtsubo K, Nakai H (2014) Discovery of Two β-1,2-Mannoside Phosphorylases Showing Different Chain-Length Specificities from Thermoanaerobacter sp. X-514. PloS One 9:e114882. doi: 10.1371/journal.pone.0114882

Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Pudlo NA, Porter NT, Urs K, Thompson AJ, Cartmell A, Rogowski A, Hamilton BS, Chen R, Tolbert TJ, Piens K, Bracke D, Vervecken W, Hakki Z, Speciale G, Munōz-Munōz JL, Day A, Peña MJ, McLean R, Suits MD, Boraston AB, Atherly T, Ziemer CJ, Williams SJ, Davies GJ, Abbott DW, Martens EC, Gilbert HJ (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature 517:165–169. doi: 10.1038/nature13995

Hungate RE (1969) Chapter {IV} A Roll Tube Method for Cultivation of Strict Anaerobes. In: Norris JR, Ribbons DW (eds). Academic Press, pp 117 – 132

Ladevèze S, Cioci G, Roblin P, Mourey L, Tranier S, Potocki-Véronèse G (2015) Structural bases for N-glycan processing by mannoside phosphorylase. Acta Crystallogr D Biol Crystallogr 71:1335–1346. doi: 10.1107/S1399004715006604

Ladevèze S, Tarquis L, Cecchini DA, Bercovici J, André I, Topham CM, Morel S, Laville E, Monsan P, Lombard V, Henrissat B, Potocki-Véronèse G (2013) Role of glycoside phosphorylases in mannose foraging by human gut bacteria. J Biol Chem 288:32370–32383. doi: 10.1074/jbc.M113.483628

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 25:402–408. doi: 10.1006/meth.2001.1262

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–495. doi: 10.1093/nar/gkt1178

Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, Gordon JI (2011) Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. PLoS Biol 9:e1001221. doi: 10.1371/journal.pbio.1001221

Maruyama Y, Nakajima T (2000) The aman6 gene encoding a yeast mannan backbone degrading 1,6-alpha-D-mannanase in Bacillus circulans: cloning, sequence analysis, and expression. Biosci Biotechnol Biochem 64:2018–2020. doi: 10.1271/bbb.64.2018

Nakae S, Ito S, Higa M, Senoura T, Wasaki J, Hijikata A, Shionyu M, Ito S, Shirai T (2013) Structure of Novel Enzyme in Mannan Biodegradation Process 4-O-β-d-Mannosyl-d-Glucose Phosphorylase MGP. J Mol Biol 425:4468–4478. doi: 10.1016/j.jmb.2013.08.002

O'Neill EC, Field RA (2015) Enzymatic synthesis using glycoside phosphorylases. Carbohydr Res 403:23–37. doi: 10.1016/j.carres.2014.06.010

Rasmussen MA, Hespell RB, White BA, Bothast RJ (1988) Inhibitory Effects of Methylcellulose on Cellulose Degradation by Ruminococcus flavefaciens. Appl Environ Microbiol 54:890–897.

Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234. doi: 10.1016/j.pep.2005.01.016

Terrapon N, Henrissat B (2014) How do gut microbes break down dietary fiber? Trends Biochem Sci 39:156–158. doi: 10.1016/j.tibs.2014.02.005

Zhang Y-HP, Lynd LR (2005) Cellulose utilization by Clostridium thermocellum: bioenergetics and hydrolysis product assimilation. Proc Natl Acad Sci U S A 102:7321–7325. doi: 10.1073/pnas.0408734102

Conclusion and Prospects

As we saw, human gut bacteria produce a wide variety of CAZymes to ensure their survival in the swarming jungle of this ecosystem. Among them, the enzymes belonging to the recently created GH130 family constitute particularly interesting targets. Indeed, at the beginning of this thesis project, only one GH130 enzyme was biochemically characterized, revealing a new function of mannoside breakdown by phosphorolysis. Moreover, as previously shown in Tasse *et al*. (2010), and further revisited in the frame of this project, some highly prevalent genes of the human gut microbiome encode for putative enzymes classified in 2011 in the GH130 family. However, no function was assigned to them until July 2013, when our first article describing the function of Uhgb_MP was published.

Uhgb_MP, a GH130 enzyme produced by an uncultivated bacterium, was the main target of this project. The first part of this PhD work was thus to fully characterize its substrate specificity towards mannosyl donors and acceptors, and its kinetic behavior. These biochemical data revealed that this enzyme is capable of efficiently phosphorolyzing the N-glycan core oligosaccharide structure D-Man*p*-β-1,4-D-Glc*p*NAc-β-1,4-D-Glc*p*NAc. By using site-directed mutagenesis, we identified key residues for catalysis, in particular D104, which is the most probable proton donor.

Moreover, the mapping of 369 GH130 sequences on the human fecal metagenome of 411 individuals highlighted the high prevalence of Uhgb_MP homologs in the microbiome of patients suffering from inflammatory bowel diseases. Finally, thanks to the analysis of the genomic context of 65 known GH130 sequences belonging to human gut bacteria, we proposed a novel pathway of N-glycan catabolism, involving Uhgb_MP homologs working in synergy with glycoside hydrolases of the GH92 and GH18 families. These results, and those published concomitantly regarding the GH130 enzyme *Bt*1033 from *Bacteroides tethaiotamicron* VPI-5482, allowed one to identify a new mechanism of human N-glycan degradation by gut bacteria, for which only glycoside hydrolases had been considered to be involved in, until 2013 (Thompson et al., 2012; Zhu et al., 2010). They also allowed one to expand the known enzyme functional range in the GH130 family, which contained in 2013 only 3 characterized enzymes, all being involved in the catabolism of plant mannans in the bovine rumen and the human gut ecosystems (Kawahara et al., 2012).

This work also permitted the creation of two GH130 subfamilies, related to plant β-mannan and N-glycan degradation by gut bacteria. Uhgb_MP itself is able, *in vitro*, to breakdown plant mannan, which makes it the first mannoside phosphorylase acting on a polysaccharide. Finally, we revealed in this study that Uhgb_MP is an enzyme with high potential for industrial biotechnology, as it is the most efficient enzyme known, to date, able to synthesize by reverse-phosphorolysis the precious D-Man*p*-β-1,4-D-Glc*p*NAc-β-1,4-D-Glc*p*NAc and D-Man*p*-β-1,4-D-Glc*p*NAc oligosaccharides, of which the commercial price is nowadays over $ 10,000 per mg. These high added value compounds can even be

synthesized from mannan, this low cost hemicellulosic compound, in a two-step process involving only Uhgb_MP as catalyst. These properties have been patented in 2013.

Optimization of the reaction yields of this two-step process, in particular by shifting the reaction equilibrium towards the phosphorolysis and synthesis directions, respectively for steps one and two, will therefore be necessary to envisage the commercialization of these products.

The second part of this thesis project was dedicated to the structural characterization of this fascinating enzyme. Its crystallographic structure has been solved at 1.85 Å, in its *apo* form and in complex with mannose, *N*-acetylglucosamine and both. This is the first 3D structure of a N-glycan phosphorolytic enzyme. This work allowed us to confirm both the involvement of D104 as proton donor, and its catalytic mechanism, involving an unpredictable structural rearrangement involving N44, S45 and D104. Analysis of this first structure solved for a member of the GH130_2 subfamily also revealed H174 and M67 as determinants for specificity towards N-glycan motives and mannan. SAXS and crystal structure analysis also revealed a hexameric structure, a specific feature of GH130 enzymes among other glycoside-phosphorylases.

Such GH130 targeted genomic and transcriptomic studies have been initiated to study their role in mannoside metabolization by *Bacteroides ovatus* ATCC 8483, a prominent gut symbiont. The results show that expression 6 of the 7 *B. ovatus* GH130 encoding genes is induced by plant β-mannan. The single one of which expression is not affected by this polymer, encodes the protein BACOVA_03624, classified in the GH130_NC subgroup. Its biochemical characterization revealed that it is a very original enzyme, capable of performing a large set of reactions. Acting as an hydrolase, it is also able to perform transglycosylation, phosphorolysis and reverse-phosphorolysis reactions. This enzyme, contrary to Uhgb_MP and all other characterized GH130 members, catalyses hydrolytic reactions thanks to the presence of a second acid catalytic residue (BACOVA_03624 E144), which probably plays the role of base. BACOVA_03624 is thus the first characterized hydrolase of the GH130 family. In addition, it is the only characterized member that is able to accommodate both mannosyl and galactosyl residues in its -1 subsite. This high flexibility is also demonstrated by the wide substrate tolerance of the +1 subsite. Based on these biochemical data, and on results of genomic context analysis, we suspect that BACOVA_03624 physiological role would be hydrolysis of the core β-D-Man*p*-1,4-D-Glc*p*NAc motif of human and yeast N-glycans, and/or β-D-Gal*p*-1,4-D-Glc*p*NAc (*N*-acetyllactosamine), a disaccharide existing in hybrid and complex human N-glycans. We also suspect at this stage that BACOVA_03624 could also act on β-1,2-linked-manno-oligosaccharides or phosphated mannosides, which can be found in the high mannose N-glycans of the pathogen *C. albicans*. NMR analysis of the reverse-phosphorolysis products obtained from the GlcNAc and Man

acceptors, and testing hydrolysis and *B. ovatus* growth on these particular substrates should soon allow one to be more conclusive on BACOVA_03624 targets in the human gut. In addition, the construction of the BACOVA_03624 E144Q and/or E144Y mutants will be necessary to demonstrate the role of E144 in hydrolysis. Finally, soaking or co-crystallization of BACOVA_03624 with its substrates will bring additional structural data that would help us to understand how both hydrolysis and phosphorolysis can occur, and which are the structural features promoting one or the other mechanism.

The results obtained in the frame of this thesis largely expanded our knowledge on the catalytic mechanism, structure-function relationships and physiologic roles of GH130 enzymes, and also highlighted their high potential for synthesis of high added value oligosaccharides (Table 2). However, even if this work and that of other research groups largely contributed to increase the amount of biochemical data available on this family, the number of characterized members is still limited (Fig. 20).

Protein	Subfamily	Structure	Function	Organism
BT1033	GH130_2	✗	✓	*Bacteroides thetaiotaomicron* VPI-5482
*Ra*MP2	GH130_2	✗	✓	*Ruminococcus albus 7*
Uhgb_MP	GH130_2	✓ *	✓ *	*Unknown*
*Tm*1225	GH130_2	✓	✓ *	*Thermotoga maritima* MSB8
ADD61810	GH130_2	✗	✗	*Unknown*
*Ra*MP1	GH130_1	✗	✓	*Ruminococcus albus 7*
*Rm*MGP	GH130_1	✗	✓	*Rodothermus marinus* ATCC43812
*Bf*MGP	GH130_1	✓	✓	*Bacteroides fragilis* NCTC9343
BT4094	GH130_NC	✓	✗	*Bacteroides thetaiotaomicron* VPI-5482
BDI_3141	GH130_NC	✓	✗	*Parabacteroides distasonis* ATCC8503
BACOVA_02161	GH130_NC	✓	✗	*Bacteroides ovatus* ATCC8483
BACOVA_03624	GH130_NC	✓	✓ *	*Bacteroides ovatus* ATCC8483
Teth514_1788	GH130_NC	✗	✓	*Thermoanaerobacter sp, X-514*
Teth514_1789	GH130_NC	✗	✓	*Thermoanaerobacter sp, X-514*

Figure 20: Overview of available data on GH130 enzymes in January 2015.
The family includes in total 13 different characterized proteins, either structurally or functionally. Three of them belong to subfamily 1, five to subfamily 2 and six are grouped in the pool of sequences that are not classified. It is worth noting that one single representative of each group of sequences has been both characterized functionally and structurally. Stars indicate the inputs of the present thesis to the knowledge of the family. Red arrows indicate data which are lacking about the listed enzymes. Validated data are indicated by a green tick. Note: functional characterization of *Tm*1225 is currently in progress.

Table 2: Summary of acceptor tolerance for synthetic reactions catalyzed by GH130 enzymes, based on available data in January 2015.

This table assumes the use of α-D-man*p*-1-phosphate as donor for all enzymes. +, ++, and +++ indicates the observed activity on the specified substrate. ND: not determined. [1] data generated during this thesis. [2] Both teth514_1788 and Teth514_1789 are specific for β–1,2 linkages and do not process β-1,4 bonds. Note: regarding Teth514_1788 and Teth514_1789, β-1,2-manno-oligosaccharides longer than DP2 (β–1,2-mannobiose) were not tested in the acceptor reactions. However, products as long as DP3 and DP5 were detected after a 24h incubation of Teth514_1789 and Teth514_1788, respectively. BACOVA_03624 is able to use both α-D-man*p*-1-phosphate and α-D-gal*p*-1-phosphate as donor, but the panel of acceptor tolerated differs dependind on the used donor. The activities measured using both substrates as donors are indicated as a pair A/B, A being the activity using α-D-man*p*-1-phosphate, and B being the activity using α-D-gal*p*-1-phosphate. All reactions were carried out using 10mM P_i. Preleminary characterization of *Tm*1225 showed production of manno-oligosaccharides DP2-20 from α-D-man*p*-1-phosphate alone.

Substrate	*Bf*MGP	*Ra*MP1	*Rm*MGP	*Ra*MP2	*Bt*1033	Uhgb_MP[1]	Teth514_1788	Teth514_1789	BACOVA_03624 A/B	*Tm*1225
D-Glucose	+++	+++	+++	+	+	++	0	0	++/+++	ND
2-Deoxy-D-glucose	0	0	ND	++	0	ND	0	0	ND/ND	ND
2-Deoxy-2-fluoro-D-glucose	0	0	ND	++	ND	ND	ND	ND	ND/ND	ND
D-Mannose	0	0	0	++	+	++	+++	+++	+/++	+++
N-Acetyl-D-glucosamine	0	0	0	+	+++	+++	0	0	+++/+	ND
3-Deoxy-D-glucose	0	0	ND	+	ND	ND	ND	ND	ND/ND	ND
3-Deoxy-3-fluoro-D-glucose	0	0	ND	+++	ND	ND	ND	ND	ND/ND	ND
D-Allose	0	0	0	+++	0	+	0	0	0/0	ND
3-O-Methyl-D-glucose	0	0	ND	+++	0	ND	0	0	ND/ND	ND
6-Deoxy-D-glucose	0	+	++	0	ND	ND	ND	ND	ND/ND	ND
6-Deoxy-6-fluoro-D-glucose	0	+	ND	0	ND	ND	ND	ND	ND/ND	ND
D-Xylose	0	0	0	0	0	0	ND	ND	0/0	ND
1,5-Anhydro-D-glucitol	0	0	++	++	0	ND	0	0	ND/ND	ND
Methylβ-D-glucoside	0	0	++	++	0	ND	0	0	ND/ND	ND
Methyl α-D-glucoside	0	0	0	++	0	ND	0	0	ND/ND	ND
Sophorose	0	0	ND	++	0	ND	0	0	ND/ND	ND

Laminaribiose	0	+	ND	++	0	ND	0	0	ND/ND	ND
Cellobiose	0	0	0	++	0	0	0	0	0/0	ND
Gentibiose	0	0	ND	++	0	ND	0	0	ND/ND	ND
Maltose	0	0	ND	++	0	ND	0	0	ND/ND	ND
N,*N*'-diacetylchitobiose	0	0	ND	++	++	+++	0	0	ND/ND	ND
β-1,4-Mannobiose	0	0	0	++	0	+++	++[2]	+++[2]	+/++	ND
β-1,4-Mannotriose	0	0	ND	++	ND	+++	+++[2]	0	+/+	ND
β-1,4-Mannotetraose	0	0	ND	+++	ND	+++	ND	ND	0/0	ND
β-1,4-Mannopentaose	0	0	ND	+++	ND	+++	ND	ND	0/0	ND
β-1,4-Mannohexaose	0	0	ND	+++	ND	+++	ND	ND	ND/ND	ND
β-1,4-mannan	ND	ND	ND	ND	ND	+++	ND	ND	ND/ND	ND
D-fructose	ND	ND	ND	ND	0	+++	+++	+++	++/+	ND
D-galactose	ND	ND	ND	ND	0	++	0	0	++/0	ND
D-altrose	ND	ND	ND	ND	0	traces	0	0	0/0	ND
D-Fucose	ND	ND	ND	ND	0	0	ND	ND	+++/+	ND
D-rhamnose	ND	ND	ND	ND	ND	0	ND	ND	ND/ND	ND
Xylitol	ND	ND	ND	ND	ND	0	ND	ND	ND/ND	ND
N-Acetyl-D-galactosamine	ND	ND	ND	ND	0	+	0	0	++/++	ND

That is why, in the frame of this thesis, we were interested in other GH130 enzymes. In particular, the protein ADD61810, which was also discovered by functional screening of the human gut metagenome, is an interesting target. Indeed, contrary to the other members of the GH130_2 subfamily but similarly to the GH130_1 ones, its gene is surrounded by GH5 and GH2 encoding genes, the PUL signature of plant mannan degrading enzymes. The ADD61810 encoding gene was cloned in the frame of this thesis, and its functional characterization will soon be started to investigate its substrate specificity.

We also expressed and started to characterize the function of *Tm*1225 from *Thermotoga maritima* MSB8, a GH130_2 protein which shares 61% sequence identity with Uhgb_MP. *Tm*1225 is the best β-mannoside polymerase known to date, and it is able to synthesize by reverse-phosphorolysis β-mannoside chains of DP 2 to 20, at least (fig.21). NMR analysis will soon be performed to determine the linkage specificity displayed by this enzyme. Its ability to perform mannosyl residue polymerization from a large panel of acceptors, including other glycans of various structures, will be tested to evaluate its interest for synthesis of high-added value oligosaccharides and of novel biosourced copolymers. In addition, this enzyme, whose *apo* structure was solved several years ago, appears to be the closest Uhgb_MP structural homolog. But contrary to Uhgb_MP, *Tm*1225 crystallized as a dimer. It will be interesting to confirm this state in solution, and to identify the molecular bases of multimerization in this particular CAZy family, which contains the sole known non-dimeric GPs. Solving *Tm*1225 structure in complex with its substrates will also allow one to identify the structural determinants of long mannoside polymerization. Together with the identification of the molecular bases of the Uhgb_MP and BACOVA_03624 specificity towards mannosides and galactosides, these works will open the way to rational engineering of GH130 enzymes, for optimization of their glycoside synthesis properties.

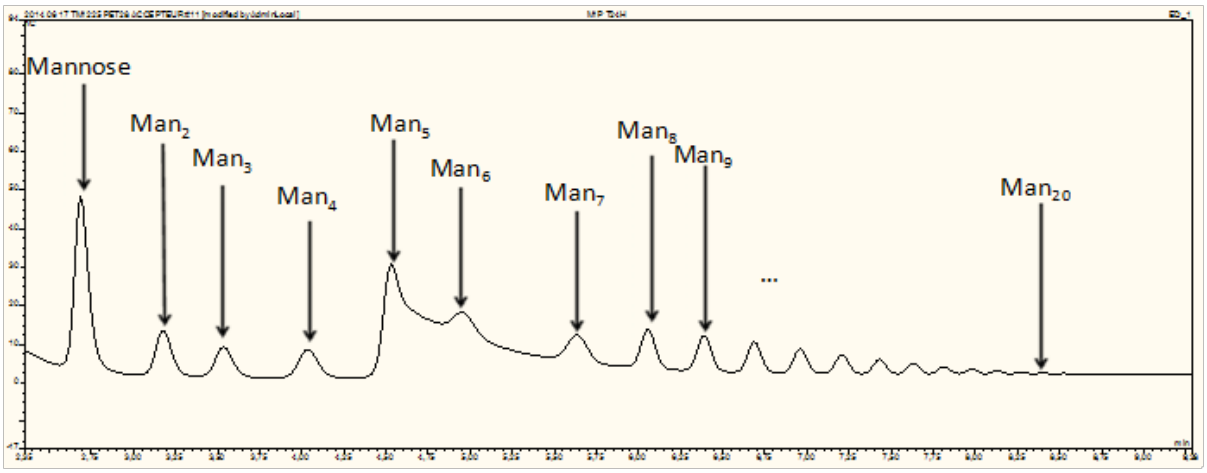


Figure 21: HPAEC-PAD profile of *Tm*1225 incubated with α-D-Man*p*-1-phosphate alone (reaction time 24h).

Finally, as highlighted by integration of biochemical, structural, genomic, metagenomics and transcriptomic data, GH130 enzymes can now be considered as new targets to study interactions between host and gut microbes, which is currently the focus of intense international research efforts. Inhibition of host glycan degrading enzymes could thus be considered as an innovating strategy to protect or restore the intestinal barrier in the context of inflammatory bowel diseases. In this context, we identified several monosaccharides, polyols and multivalent iminosugars acting as Uhgb_MP inhibitors. Surprisingly, we also observed that some imminosugars, that most of the time inhibit glycosidases, act as Uhgb_MP activators. A structural analysis of Uhgb_MP crystallized in complex with its inhibitors and activators would allow one to identify by which mechanism these molecules affect its activity. The structural characterization of GH130 enzymes produced by human gut bacteria, such as Uhgb_MP, opens the way to the design of specific inhibitors of N-glycan processing enzymes, by mimicking the interactions described in chapter 5 for substrate binding. However, further studies will be needed to confirm the physiological role of these enzymes, and the inhibitor effects, like metabolomics and transcriptomic analyses of gut bacteria producing GH130_2 enzymes in the presence of N-glycans as carbon source, with and without inhibitors. Finally, the over-representation of GH130_2 sequences in IBD suffering patient metagenomes, that we observed, will have to be confirmed on larger patient cohorts, in order to know on which extent GH130 sequences (and those encoding other mucus degrading enzymes) could be used as biomarkers for IBD detection.

Résumé de la thèse

Le microbiote intestinal humain est composé de 10^{14} bactéries par gramme de fèces, dont seulement 20 à 30% sont cultivées. Cet écosystème complexe joue un rôle crucial pour la santé humaine, notamment en métabolisant les fibres alimentaires, ces polysaccharides de structure très diverse, issus majoritairement de la paroi végétale (céréales, fruits et légumes). Ils participent ainsi pour une part non négligeable à l'assimilation des constituants issus de notre alimentation. Pour assurer leur croissance, certaines de ces bactéries sont également capables de dégrader les glycanes de l'hôte, telles les O- et N- glycosylations portées par les protéines tapissant la paroi intestinale et constituant les mucines, mais également ceux produits par les levures alimentaires.

Pour dégrader efficacement cette grande variété de polysaccharides, les bactéries intestinales ont développé un arsenal très diversifié d'enzymes capables de les dégrader, éléments fondamentaux de leur survie dans cet écosystème compétitif. Ces enzymes, appelées enzymes actives sur les sucres (Carbohydrate Active enZymes, CAZymes), peuvent aussi s'avérer constituer de puissants outils biotechnologiques, notamment pour la déconstruction de la biomasse d'origine végétale, ainsi que pour la synthèse d'oligosaccharides, polysaccharides et glycoconjugués pour de nombreuses applications, dans les domaines de l'agroalimentaire, de l'alimentation animale, de l'environnement, ou encore de la santé humaine.

La base de données CAZy recense et classe ces enzymes en familles selon leur similarité de séquence. Si l'évolution des méthodes de séquençage à haut débit a permis d'augmenter considérablement la vitesse de leur recensement ces dernières années, la caractérisation fonctionnelle et structurale de ces enzymes est loin de suivre le même rythme, limitant la compréhension des relations entre structure et fonction enzymatique. En effet, moins de 10% des enzymes recensées dans la base de données CAZy ont été caractérisées sur le plan fonctionnel, et moins de 1% ont une structure tridimensionnelle résolue.

En 2010, une étude de métagenomique fonctionnelle à haut-débit menée par le groupe « DiscOmics » de l'équipe de Catalyse et Ingénierie Moléculaire Enzymatiques du LISBP, ciblant le microbiome intestinal humain, a permis d'identifier plusieurs dizaines de CAZymes à fort potentiel de valorisation, en particulier pour la dégradation de la paroi végétale. Douze d'entre elles n'appartenaient à aucune famille CAZy connue. Mon projet de thèse a visé à caractériser d'un point de vue fonctionnel et structural l'une de ces nouvelles enzymes, afin i) de comprendre quel est son rôle (ainsi que celui joué par ses nombreux homologues) dans l'écosystème intestinal humain, ii) d'étendre les connaissances des relations entre structure et fonction des CAZymes, iii) et d'évaluer son potentiel applicatif pour la dégradation de la biomasse végétale et la synthèse d'oligosaccharides à haute valeur ajoutée.

Mon travail a donc porté en grande partie sur l'étude de l'enzyme que nous avons nommée « Uhgb_MP » (Unknown human gut bacterium Mannoside Phosphorylase), appartenant à la toute nouvelle famille des Glycosides Hydrolases 130 (GH130). La famille GH130, créée quelques mois avant le début de cette thèse, regroupe désormais plus de 648 séquences. Sa création fait suite à la caractérisation fonctionnelle de l'un de ses membres en mai 2011, qui est capable de phosphorolyser le β-D-Mannopyranosyl-1-4-D-Glucopyranose (Man-Glc), en libérant du mannose-α-1-phosphate et du glucose. En 2011, rien de plus n'était connu sur cette famille, excepté le fait que les structures 3D de quatre autres enzymes GH130, qui adoptent un repliement de type 5-bladed β-propeller, avaient été résolues dans le cadre d'initiatives de génomique structurale, mais sans fonction associée. En particulier, ni le résidu catalytique, ni les spécificités de ces enzymes n'étaient connues, alors même qu'il s'agissait de la seule famille connue comprenant une mannoside-phosphorylase hautement spécifique du disaccharide Man-Glc.

Nous avons donc entrepris la caractérisation fonctionnelle d'Uhgb_MP, produite sous forme étiquetée poly-histidine dans le vecteur pDEST17. La spécificité de l'enzyme vis-à-vis de ses substrats et produits, ainsi que ses paramètres cinétiques ont été déterminés par HPAEC-PAD, ^{13}C et ^{1}H RMN, et spectrométrie de masse. Contrairement à la première enzyme caractérisée de la famille GH130 (la Mannosyl-Glucose phosphorylase *Bf*MGP de l'organisme *Bacteroides fragilis*), Uhgb_MP est capable de phosphorolyser de nombreux substrats mannosylés, notamment les β-1,4-mannanes et manno-oligosaccharides issus de la fraction hemicellulosique de la paroi végétale, mais aussi le mannopyranosyl-β-1,4-D-*N*-acetylglucosamine (Man-GlcNAc) et le manno-chitobiose, les motifs hautement conservés des N-glycanes eucaryotes. Le manno-chitobiose est apparu comme le substrat pour lequel l'efficacité catalytique était la plus élevée.

De plus, la caractérisation cinétique des réactions de phosphorolyse inverse (réaction de synthèse), à partir de différents sucres accepteurs, en particulier la *N*-acetylglucosamine ou le chitobiose, et du mannose-α-1-phosphate (le seul sucre-phosphate donneur d' Uhgb_MP), a révélé une efficacité catalytique maximale pour la synthèse de manno-chitobiose, avec toutefois un phénomène d'inhibition par excès d'accepteur pour des concentrations supérieures à 1mM de chitobiose. Uhgb_MP catalyse les réactions de phosphorolyse et de phosphorolyse inverse de manno-oligosaccahrides par un mécanisme séquentiel de type Bi Bi, impliquant la formation d'un complexe ternaire entre les deux substrats et l'enzyme.

L'identification de ces nouvelles fonctions de dégradation de mannanes, un constituant des hémicelluloses, par phosphorolyse, et de synthèse de Man-GlcNAc (actuellement commercialisé à plus de 10 000€ le mg) et de manno-chitobiose par

phosphorolyse-inverse, fit l'objet d'un dépôt de brevet Européen par Toulouse Tech-Transfer, étendu en 2014, afin de couvrir leur usage au niveau mondial.

L'analyse de la structure primaire d'Uhgb_MP et de ses homologues de la famille GH130, et de son modèle structural construit par modélisation moléculaire, m'a permis d'identifier les résidus clefs de la catalyse, et plus particulièrement le résidu catalytique donneur de proton, l'acide aspartique 104. En 2013, l'obtention de la structure cristallographique de *Bf*MGP en complexe avec son substrat, permit de confirmer l'implication du D104 comme donneur de proton. En effet, l'acide aminé catalytique décrit par les auteurs est l'équivalent du D104 chez Uhgb_MP.

L'analyse approfondie de la structure primaire des enzymes de la famille GH130 fut conduite en collaboration avec l'équipe de Bernard Henrissat à l'AFMB (CNRS Marseille), fondateur du système de classification des CAZymes. L'alignement des 369 séquences protéiques recensées en janvier 2013, a permis de faire émerger très clairement deux sous-familles , GH130_1 et GH130_2, au sein de cet ensemble. Ces deux groupes homogènes d'un point de vue structural, contiennent respectivement 79 et 42 séquences. Le reste des séquences, les GH130 non classés (GH130_NC) est trop hétérogène pour pouvoir constituer une sous-famille. Dans cette nouvelle classification, Uhgb_MP apparut alors comme étant l'un des membres de la sous-famille 2 (fig. 22).

Par la suite, j'ai effectué l'analyse des environnements génomiques de 65 GH130 appartenant à 28 bactéries de l'intestin humain. Cette étude nous a permis de proposer deux voies distinctes de métabolisation des hydrates de carbone mannosylés par les enzymes de la famille GH130:

- Les gènes codant pour des enzymes de la sous-famille GH130_1 (comme *Bf*MGP) sont majoritairement entourées de glycoside hydrolases appartenant aux familles GH3, GH5 et GH28 (β-mannanases, β-mannosidases et β-glucosidases). Ces enzymes seraient donc impliquées dans la métabolisation des mannanes, galactomannanes et glucomannanes de la paroi végétale par les micro-organismes qui les possèdent.
- Les gènes codant pour des enzymes de la sous-famille GH130_2, comme Uhgb_MP, sont au contraire majoritairement entourées de glycoside hydrolases appartenant aux familles GH92, GH18 et GH97 (β-mannosidases, endo-β-*N*-acetylglucosaminidases et α-glucosidases) et n'ont jamais comme voisin des enzymes appartenant aux familles retrouvées autour des GH130_1, à l'exception de deux d'entre elles. Compte tenu du contexte génomique des gènes codant pour les GH130_2, celles-ci seraient donc impliquées dans la métabolisation des N-glycanes de l'hôte. En effet, ceux-ci sont formés par le cœur ubiquitaire des N-

glycanes, le trisaccharide manno-chitobiose avec des extensions α-1,3, α-1,6 et α-1,2-mannosyles sur le résidu mannosyl lié au chitobiose. Enfin, une des ramifications peut porter des résidus glucosyl terminaux liés en α-1,2. Dans ce modèle, la GH18 hydrolyserait la liaison entre les deux *N*-acetylglucosamines, libérant l'oligosaccharide de la protéine. Cet oligosaccharide serait ensuite importé dans l'espace périplasmique afin d'être déconstruit par les actions successives de la GH97, pressentie pour dégrader la liaison α-1,2 des résidus glucosyl terminaux, alors que les GH92 permettraient de décomplexifier la structure en éliminant les mannosyls liés en α. Enfin, le disaccharide β-D-man*p*-1,4-D-Glc*p*NAc serait dégradé par Uhgb_MP, comme démontré par les tests biochimiques. Cependant, Uhgb_MP ne possédant pas de peptide signal, ce disaccharide doit être nécéssairement transporté dans le milieu intracellulaire via une perméase codée par le même cluster multigénique. De nombreux Polysaccharide Utilization Loci (PULs, clusters multigéniques dédiés à la séquestration et à la dégradation d'un polysaccharide donné) homologues à celui codant pour Uhgb_MP ont été retrouvés dans les génomes de bactéries intestinales.

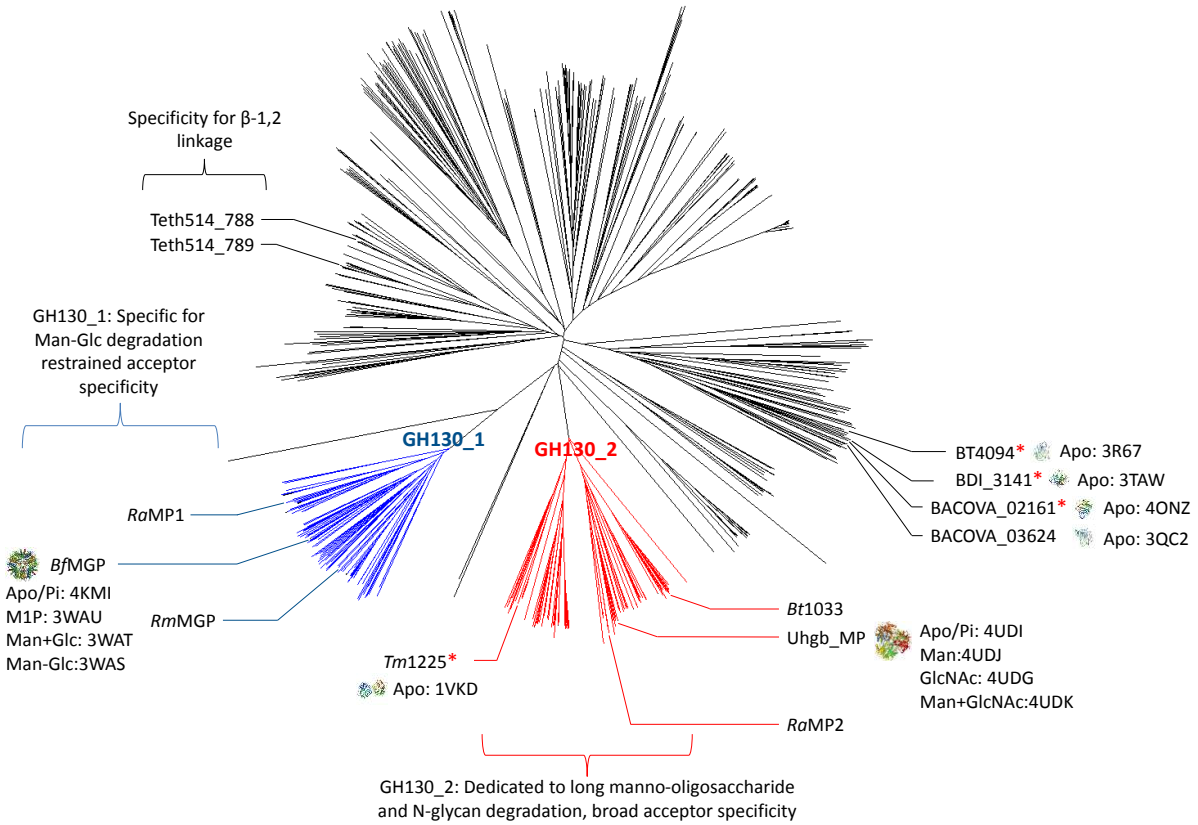


Figure 22: Représentation radiale de l'arbre phylogénétique de la famille GH130, en Janvier 2015.
Cet arbre, montre la présence de deux sous-familles clairement distinctes, GH130_1 et GH130_2, colorées respectivement en bleu et rouge. Le reste des séquences, trop hétérogène, reste dans un espace de séquences non classifié (GH130_NC). Sont également reportés sur cet arbre, les connaissances rapportées dans la littérature sur les

membres de cette famille (Janvier 2015). Sont en particulier indiqués, par leur nom, les enzymes caractérisées soit sur le plan fonctionnel, soit sur le plan structural, soit les deux. Lorsque la structure d'une enzyme est connue, une miniature de celle-ci est représentée ainsi que son/ses codes PDB. Les enzymes pour lesquelles la structure est connue mais pas leur fonction sont repérables par une astérisque rouge. Enfin, les fonctions démontrées des membres caractérisés de la famille sont indiqués pour chacune des sous-familles.

Par ailleurs, l'analyse de l'alignement des séquences de toutes les enzymes de la famille GH130 a permis de montrer que : i) la tyrosine 103 et l'asparagine 151, dont la chaine latérale est orientée en direction de la liaison inter-osidique clivée lors de la catalyse, sont parfaitement conservés au sein des sous-familles GH130_1 et GH130_2, qui rassemblent très probablement des glycoside phosphorylases ; ii) ces deux résidus sont quasi systématiquement remplacés par des acides aminés acides (E, D) dans l'ensemble des séquences non classées dans ces deux sous-familles. Ceci nous a conduit à faire l'hypothèse que les enzymes possédant cet acide aminé acide pourraient être des glycoside hydrolases vraies, lesquelles requièrent deux résidus catalytiques au lieu d'un pour les glycoside phosphorylases qui inversent la configuration anomérique de la liaison osidique.

Afin de valider cette hypothèse, la construction du mutant Y103E a été entreprise. L'activité de dégradation du *para*-nitrophénol-β-1,4-D-Mannopyranose en absence de phosphate (très faible pour la protéine sauvage) par ce mutant a été augmentée d'un facteur 2,7 par rapport à celle de la forme sauvage, ce qui suggère que le mutant serait capable d'hydrolyser ce substrat. Son instabilité n'a pas pu permettre d'approfondir davantage cette étude. Cependant, cette hypothèse sera confirmée par la suite dans le cadre de cette thèse, comme décrit ci-après.

Nous avons également utilisé l'ensemble des séquences GH130 et leur nouvelle classification pour estimer la prévalence et l'abondance des gènes codant ces protéines au sein des métagenomes intestinaux de 301 individus (issus des cohortes MetaHit et Human Microbiome Project). Ce travail a permis de montrer que 35 séquences GH130 provenant de micro-organismes intestinaux (sur 65 au total) sont retrouvées chez au moins 50 individus, dont celle d'Uhgb_MP qui est retrouvée dans les métagénomes de 93 individus sur 301. Les enzymes de la famille GH130 sont donc probablement des acteurs majeurs de la métabolisation des hémicelluloses et des N-glycanes eucaryotes par les bactéries intestinales. De plus, ces résultats ont permis de montrer que les séquences GH130, en particulier celles de la sous-famille 2, sont plus abondantes de 27% dans les métagénomes de patients atteints de maladies inflammatoires de l'intestin, telles que la maladie de Crohn, suggérant l'implication des micro-organismes qui les produisent dans ces pathologies.

L'ensemble de ces résultats a fait l'objet d'une publication parue dans le *Journal of Biological Chemistry* en 2013.

Je me suis focalisé en parallèle sur la caractérisation structurale d'Uhgb_MP. Afin d'obtenir sa structure cristallographique, j'ai dû optimiser la stabilité de la protéine, en sous-clonant son gène depuis le vecteur pDEST17 vers le vecteur pET28a. En effet, le modèle structural montrait que l'extrémité N-terminale était non structurée, ce qui pourrait être responsable de son instabilité. Dans la première partie de mes travaux, cette instabilité avait pu être contrôlée en présence de Tween, un surfactant incompatible avec la cristallisation protéique. En utilisant le vecteur pET28a, la taille du linker séparant l'étiquette poly-histidine et la protéine est réduite de cinq acides aminés. De plus, cette étiquette peut être clivée à l'aide d'une protéase, la thrombine, ce qui n'était pas le cas en utilisant le vecteur pDEST17. Après l'obtention de cette nouvelle construction, j'ai vérifié que la protéine était bien surproduite et active. Les résultats ont montré que la protéine était surproduite sous forme active, multimérique, à hauteur de 30 mg par litre de culture, même sans Tween 80. Afin d'éliminer le Tween 80 nécessaire à la stabilisation d'Uhgb_MP, ainsi que pour augmenter sa pureté, son homogénéité conformationnelle et les rendements de production, je me suis attaché à optimiser le protocole de purification de l'enzyme. L'analyse par Differential Scanning Fluorimetry de la stabilité de l'enzyme dans différents tampons de conditionnement a révélé deux phénomènes distincts : d'une part Uhgb_MP arbore deux températures de fusion, ce qui conforte l'hypothèse de l'oligomérisation, avec un premier point de fusion lors de la dissociation de l'oligomère, puis un second lors de la dénaturation des monomères isolés. D'autre part nous avons pu observer un effet stabilisateur très important (température de fusion augmentée de 5 à 8 degrés par rapport aux autres conditions) par les tampons phosphate de sodium ou de potassium, notamment à pH6 avec une concentration en NaCl de 150mM. L'effet stabilisateur du phosphate n'est pas surprenant car il s'agit d'un des substrats de l'enzyme, et donc stabilise probablement la protéine en se fixant dans le site actif. Cependant l'utilisation d'un tampon phosphate à pH6 n'est pas compatible avec la purification par chromatographie d'affinité en employant une étiquette poly-histidine, car cette valeur de pH correspond au pKa des chaines latérales des histidines. Nous avons donc opté pour un tampon phosphate de potassium 20mM pH7 avec 150mM de NaCl, car l'augmentation du pH de 6 à 7 ne diminue que de peu le gain de stabilité apporté (+6,4°C), et est compatible avec le protocole de purification. Dans ces nouvelles conditions, le rendement de purification est augmenté d'un facteur 3 (90mg de protéine pure par litre de culture) et l'activité spécifique de l'enzyme d'un facteur 10.

Uhgb_MP ainsi purifiée, j'ai pu conduire des essais de cristallisation au sein de l'équipe de biophysique de l'IPBS dirigée par Lionel Mourey, sous l'encadrement de Samuel Tranier. J'ai procédé à des essais préliminaires robotisés en utilisant des cribles commerciaux, pour des concentrations en protéine variables. Les résultats de ce criblage primaire ont été très encourageants, car pas moins de 94 conditions ont donné des cristaux, presque tous protéiques. Une condition a donné un cristal directement exploitable. Nous avons enregistré un jeu complet de données de diffraction des rayons X à l'ESRF de Grenoble

en mars 2014, ce qui a permis de résoudre la structure de la protéine sous sa forme *apo* (code PDB 4UDI) à une résolution de 1.85Å, grâce à une solution de remplacement moléculaire en utilisant comme modèle la structure de la protéine *Tm*1225, identique à 61%. Il s'agit de la première structure d'une enzyme capable de dégrader les N-glycanes humains par phosphorolyse.

La structure d'Uhgb_MP apporte de nouveaux éléments déterminants pour la compréhension de sa spécificité de substrat. Elle confirme tout d'abord que la protéine se trouve sous forme hexamérique, selon un arrangement en un trimère de dimères. L'analyse SAXS de la protéine, réalisée au Synchrotron Soleil en collaboration avec Pierre Roblin et Gianluca Cioci, a permis de confirmer cette organisation en solution, même si différences conformationelles ont été observées en fonction du tampon de conditionnement de la protéine. En effet, les enveloppes moléculaires calculées suggèrent une compaction de l'hexamère en présence de phosphate, faisant suite à un léger mouvement de rotation d'un des deux trimères par rapport à l'autre, entrainant par un effet de tassement une réduction du rayon de giration de la protéine.

La structure cristallographique révèle la présence d'un ion phosphate et une molécule de glycérol dans le site actif des 6 monomères, positionnés de façon symétrique. Le repliement en 5-bladed β-propeller et la topologie du site actif sont identiques à ceux observés dans les structures connues des autres membres de la famille. L'ion phosphate apparait très fortement stabilisé au cœur du site actif de l'enzyme, établissant de nombreuses interactions polaires et électrostatiques avec les chaines latérales des résidus adjacents. Ces résidus, R150, N151, R168, la K212, H231 et la Y242 sont parfaitement conservés au sein de la sous-famille GH130_2, ce qui souligne leur rôle déterminant, et conforte l'hypothèse d'une sous-famille GH130_2 exclusivement constituée de phosphorylases. Néanmoins, des différences majeures avec notre précédent modèle obtenu par modélisation moléculaire, nous conduisirent à le réviser. Ces différences se situent au niveau de l'orientation des chaines latérales formant le sillon catalytique, ainsi que les évidentes contraintes structurales liées à l'oligomérisation, qui ne pouvaient pas être prédites par modélisation moléculaire. En effet, de façon symétrique au sein de chaque monomère, une boucle d'un monomère voisin vient se loger au niveau des sites prédits de fixation des oligosaccharides. Par conséquent, un nouveau modèle de fixation des oligosaccharides a été élaboré, dans lequel le site actif n'apparait plus sous la forme d'un sillon mais sous celle d'un tunnel dont l'extrémité est orientée vers l'extérieur de l'hexamère (fig. 23).

Cette nouvelle hypothèse à d'abord été confortée par la publication de la structure de *Bf*MGP en complexe avec une molécule de Man-Glc. La superposition de la structure d'Uhgb_MP avec celle du complexe *Bf*MGP + Man-Glc montre une superposition quasi

parfaite du phosphate (translation d'1 Å) et surtout une orientation du disaccharide validant le nouveau modèle. Mais l'obtention de la structure d'Uhgb_MP en complexe avec ses substrats en septembre 2014 permit de valider définitivement cette hypothèse. En effet, des expériences de co-cristallisation d'Uhgb_MP avec du mannose, de la *N*-acetylglucosamine ou les deux à la fois permirent d'obtenir les co-cristaux correspondants. Les structures obtenues grâce à l'enregistrement de nouveaux jeux de données de diffraction sur le synchrotron ALBA de Barcelone permit de démontrer précisément le mode de fixation de ces ligands au sein du site actif. L'analyse de ces structures à également permis de confirmer le mécanisme catalytique atypique des enzymes de la famille GH130, qui avait été proposé en 2013 après analyse de la structure de *Bf*MGP. Dans celui-ci, le donneur de proton, le D104, qui se trouve à une trop grande distance de l'oxygène inter-osidique, fait appel à l'hydroxyle porté par le carbone 3 du résidu mannosyl situé dans le sous-site -1, afin de servir de relai pour le transfert de proton. De plus, l'analyse de ces structures à révélé un mouvement concerté des chaines latérales de certains résidus afin de permettre la catalyse (fig. 24). En effet, dans la forme *apo*, la chaine latérale du D104 n'est pas orientée correctement pour permettre la mise en place du relai de proton par le C_3-OH. Au contraire, lorsqu'un résidu mannosyl est fixé dans le site actif (code PDB 4UDJ, 1,94Å), celui-ci est fermement stabilisé par des liaisons hydrogènes formée avec les résidus adjacents, tel le D304 parfaitement conservé au sein de la famille. Par ailleurs, cette fixation induit un mouvement de la chaine latérale de la N44 pour lui permettre d'interagir avec l'hydroxyle 3 du mannose. Ce mouvement est permis par le basculement de la liaison peptidique entre la N44 et la S45. Mais ce mouvement va plus loin ; dans cette position, la N44 permet également de faire basculer la chaine latérale du résidu catalytique vers une position catalytiquement active, orientée vers le mannose. La structure contenant la *N*-acetylglucosamine seule (code PDB 4UDG, 1,6Å), logée dans le sous-site +1, ne montre pas un tel mouvement, et la fixation du résidu mannose dans le sous-site -1 apparait donc comme l'élément déclencheur de ce mouvement.

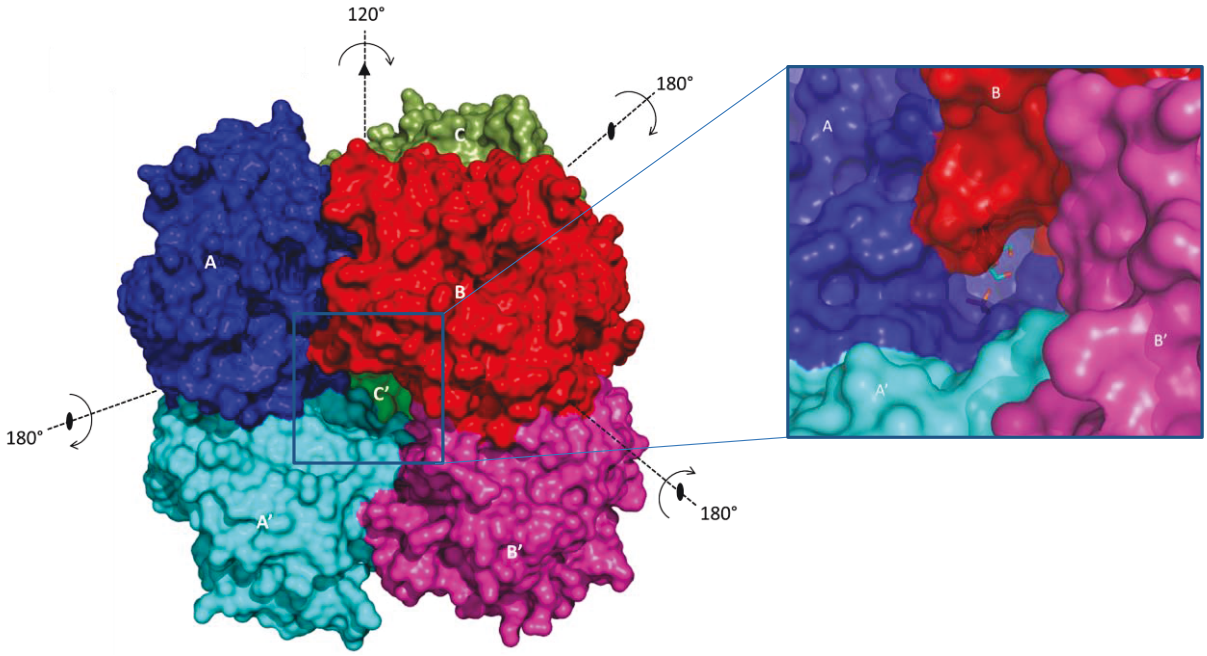


Figure 23: Structure 3D de Uhgb_MP.

La structure d'Uhgb_MP forme un hexamère, composée de trois dimères, notés A/A', B/B', C/C', et colorés respectivement en paires de bleu, rouge, vert, foncé et clair. Les axes de symétrie sont indiqués sur la figure, avec un axe d'ordre 2 pour chaque dimère, et un axe d'ordre 3 pour chaque trimère. Sur la droite est représenté un zoom de l'entrée du tunnel catalytique du monomère A, avec les molécules de mannose, *N*-acetylglucosamine et l'ion phosphate présent dans le site actif.

Les deux structures contenant la *N*-acetylglucosamine seule (code PDB 4UDG, 1,60Å), ou avec du mannose (code PDB 4UDK, 1,76Å), permettent d'expliquer la spécificité d'Uhgb_MP pour le motif cœur des N-glycanes eucaryotes. L'histidine 174 apparait comme un résidu déterminant car elle est directement impliquée dans une liaison hydrogène avec l'hydroxyle porté par le carbone 6 du sucre présent dans le sous-site +1. Le rôle critique joué par ce résidu est démontré par les résultats biochimiques car, de façon cohérente, le rhamnose (6-deoxy mannose) se comporte comme un inhibiteur d'Uhgb_MP. D'autres résidus sont impliqués dans la fixation du glycosyl présent dans le sous-site +1, interagissant avec l'hydroxyle porté par le carbone 3, comme Y103, localisée à la base de ce sous-site, pour servir de plate-forme aromatique permettant la fixation des sucres. De plus, des résidus apolaires tels la M67, permettent de former des interactions de Van der Walls avec le methyle du groupement *N*-acetyle. L'implication d'un tel résidu apolaire dans la sélectivité vis-à-vis des substrats *N*-acetylés a par ailleurs déjà été observée pour d'autres enzymes. Cependant, Uhgb_MP étant cristallisée en présence de P_i, il n'a pas été possible d'obtenir la structure d'Uhgb_MP en complexe avec son substrat préférentiel, le manno-chitobiose, bien que des expériences de co-cristallisation et de trempage aient permis d'enregistrer des jeux de données de diffraction. Ceux-ci ne montraient malheureusement que la *N*-acetylglucosamine présente dans le sous-site +1, la densité de celle présente dans le +2

étant trop faible pour permettre son positionnement dans le modèle 3D, ce qui suggère une forte mobilité de cette dernière ainsi que des résidus formant le sous-site +2.

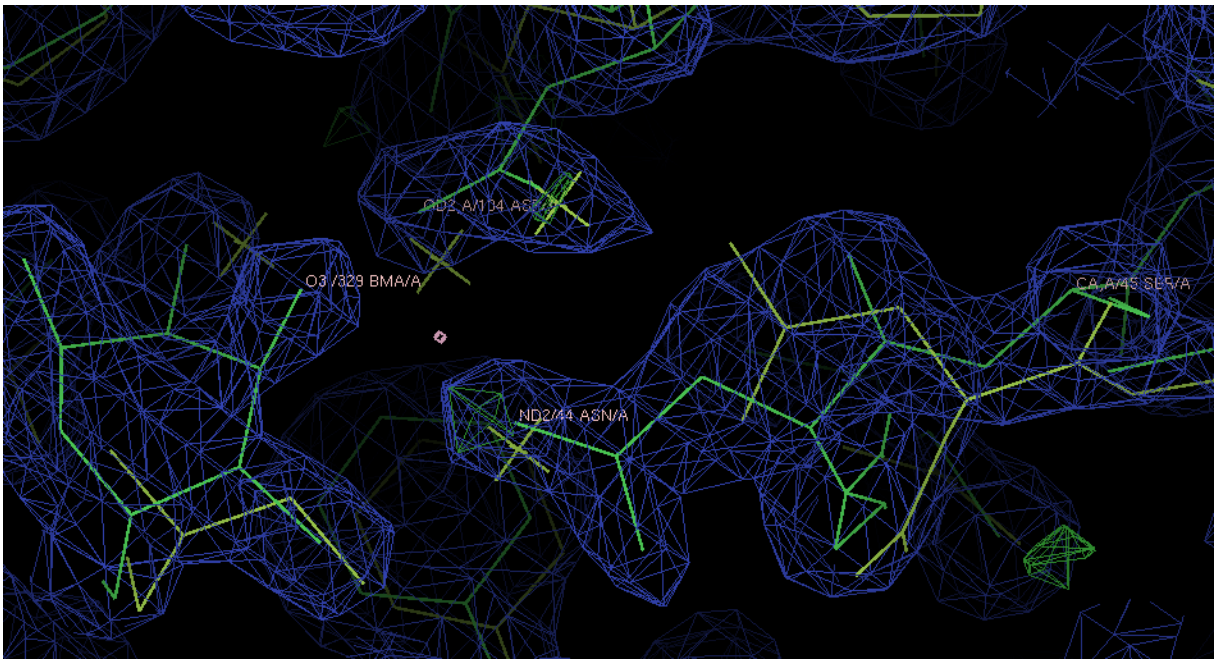


Figure 24: Superposition de structures d'Uhgb_MP sous formes *apo* et en complexe avec du mannose. Zoom au niveau du sous-site -1.

La forme *apo* est représentée en jaune, et la forme complexée au mannose en vert. La carte de densité affichée est celle du complexe avec le mannose. Dans la forme *apo*, une molécule de glycérol est présente dans le sous-site, et mime les interactions établies par la molécule de mannose. Elle interagit avec la N44, mais par l'intermédiaire d'une molécule d'eau (croix jaune), qui est éjectée dans la forme complexée avec le mannose, car s'est là que vient se positionner la N44. Le résidu mannose, sous sa forme bateau $B_{2,5}$, lorsque fixé dans le sous-site -1, déclenche le mouvement de la N44 dans sa direction, laquelle permet ensuite de stabiliser la chaine latérale du donneur de proton dans sa configuration active, tourné vers l'hydroxyle 3 du mannose.

La comparaison de cette nouvelle structure avec celles des autres membres de la famille pour lesquelles elles sont disponibles à également révélé des différences fondamentales expliquant leurs différences fonctionnelles. En effet, même si nous ne disposons que de 6 autres structures, au moins une est représentante de chaque cluster de séquences. Tout d'abord ces enzymes s'assemblent selon un état oligomérique différent, allant du monomère à l'hexamère en passant par le dimère, le tetramère et le pentamère. Cette observation est déjà hors du commun, car toutes les autres glycoside phosphorylases connues sont retrouvées actives en solution sous la forme de dimères. Par ailleurs, au niveau même de la structure de leurs protomères respectifs, des différences permettent d'expliquer leurs spécificités de substrats. En l'occurrence, certains éléments de structure secondaire, en particulier des boucles, différencient les sous-familles. Chez les GH130_1, une boucle qui leur est spécifique vient se loger à l'emplacement de l'extrémité du tunnel catalytique, ce qui explique qu'elles ne peuvent accommoder des substrats de degré de polymérisation supérieur à 2. Une autre boucle, celle-ci spécifique des GH130_NC, joue le même rôle de

blocage de l'extrémité du tunnel, ce qui est en accord avec les résultats biochimiques observés par Jaito *et al.* sur la GH130_NC Teth514_1788, incapable de synthétiser des oligosaccharides de DP>3, et sur ceux que nous avons obtenus sur la protéine BACOVA_03624, qui accommode très mal les oligosaccharides de DP>4.

L'ensemble de ces résultats structuraux a fait l'objet d'une publication parue dans la revue *Acta Crystallographica Section D* en 2015.

Afin de caractériser de façon plus fine la famille GH130, et plus particulièrement en ce qui concerne la métabolisation *in vivo* des mannanes et N-glycanes, nous avons par la suite entrepris de compiler et de compléter les informations disponibles sur certaines autres protéines de la famille. Ces différentes protéines sont des représentants de chacune des sous-familles, ainsi que de l'ensemble des séquences non classées, présentes chez l'organisme *Bacteroides ovatus ATCC* 8483. Le génome de cet organisme, un membre hautement prévalent du microbiote intestinal, code pour sept GH130, deux appartenant à la sous-famille 1, une à la sous-famille 2, et quatre regroupées dans l'ensemble non classifié. Afin de valider les hypothèses quant aux rôles physiologiques des différentes sous-familles GH130, des expériences de transcriptomique ont été entreprises en collaboration avec Pascale Mosoni et Jordane Despres, de l'INRA de Theix à Clermont-Ferrand. Les premiers résultats, encore préliminaires, confirment et complètent ceux déjà publiés (Martens et al, 2011) sur l'induction des PULs de *B. ovatus*, dont six mettant en jeu des GH130 dont l'expression est induite par les β-mannanes d'origine végétale. En particulier, ce travail a permis de valider expérimentalement l'implication de la GH130_NC BACOVA_04885 dans la métabolisation des β-mannanes. La seule GH130 de *B. ovatus* dont l'expression n'est pas induite par les mannanes de plantes est BACOVA_03624, répertoriée dans le cluster GH130_NC. La structure cristallographique de cette protéine est disponible, mais pas sa fonction. D'autre part, on retrouve chez cette enzyme un acide aminé acide à la position équivalente de la position 103 chez Uhgb_MP, ce qui suggère que cette enzyme agirait au moyen d'un mécanise hydrolytique. Afin de répondre à cette question, nous avons entrepris la caractérisation fonctionnelle de cette enzyme. Les résultats ont montré que BACOVA_03624 clive effectivement les liaisons glycosidiques par hydrolyse. Cependant, en excès de phosphate inorganique, elle est également capable, dans une moindre mesure, de catalyser des réactions de phosphorolyse de β-mannosides et de β-galactosides. Enfin, elle est capable de catalyser la synthèse de ces composés, par transglycosylation, qui reste une réaction mineure, et par phosphorolyse inverse. (fig. 25). BACOVA_03624 est ainsi la première hydrolase de la famille GH130. Ces travaux confirment notre hypothèse établie grâce à l'analyse des structures primaires des enzymes de cette famille. C'est aussi la seule GH130 capable d'agir à la fois sur des galactosides et des mannosides. Cette spécificité très laxe du sous-site -1 est très originale pour une enzyme capable de catalyser des réactions de phosphorylation et phosphorylation inverse. Compte tenu de sa spécificité de substrat, et de

son contexte génomique, nous proposons que BACOVA_03624 soit impliquée *in vivo* dans la métabolisation des N-glycanes eucaryotes, qu'ils proviennent de l'hôte ou des levures alimentaires.

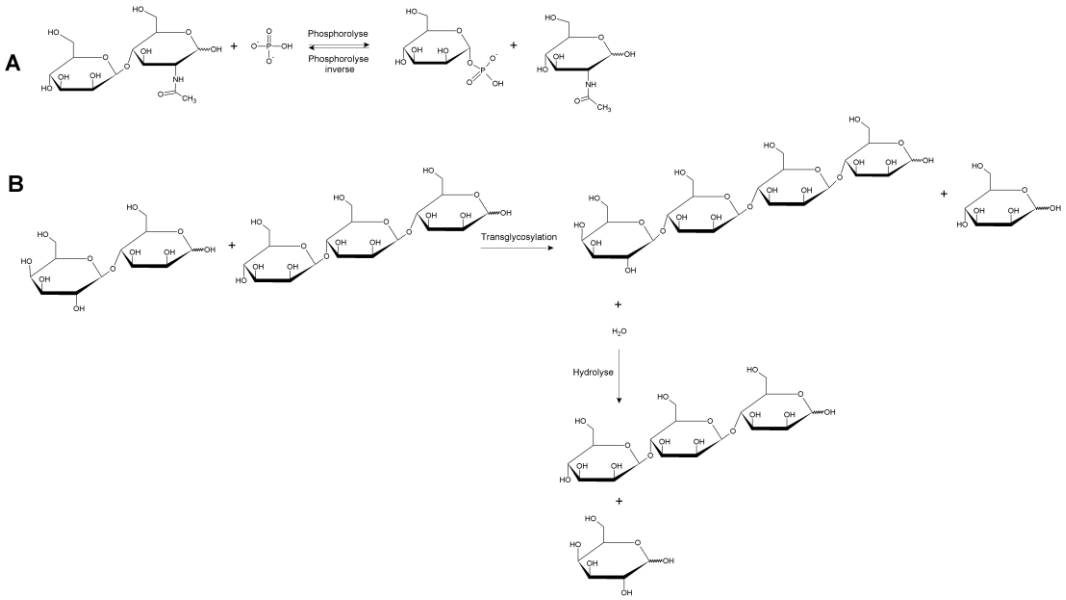


Figure 25: Réactions catalysées par BACOVA_03624.

Deux exemples de réactions catalysées par BACOVA_03624 sont représentés ici, illustrant à la fois les différents mécanismes catalytiques qu'elle utilise ainsi que sa capacité à tolérer des sucres différents dans ses sous-sites. A: réactions de phosphorolyse du β-D-Man*p*-1,4-D-Glc*p*NAc en α-D-Man*p*-1-phosphate et D-GlcNAc, et phosphorolyse inverse, en présence de phosphate inorganique. B: Réaction de transglycosylation,en absence de phosphate permettant la formation de β-D-Gal*p*-(1,4-β-D-Man*p*)$_3$ catalysées à partir de β-1,4-D-Mannotriose et β-d-Gal*p*-1,4-D-Man*p*, et d'hydrolyse permettant la dégradation du tétrasaccharide en β-1,4-D-Mannotriose et D-galactose.

L'ensemble de ces résultats, portant sur la caractérisation biochimique de BACOVA_03624, et les résultats préliminaires de transcriptomique sur *B. ovatus* ATCC 8483 feront l'objet d'une publication en cours de préparation (chapitre 7).

Comme l'ont démontré les résultats des analyses de prévalence et d'abondance des séquences GH130 au sein du métagénome intestinal humain, celles-ci sont potentiellement impliquées dans des mécanismes participant au déclenchement ou favorisant le maintien des maladies inflammatoires chroniques au niveau de l'épithélium intestinal humain. Par conséquent, l'emploi d'inhibiteurs de ces enzymes apparait comme stratégie séduisante afin d'en atténuer les effets. A ce titre, en collaboration avec l'équipe de Sébastien Gouin (CNRS Nantes), nous avons entrepris de mesurer l'effet d'un certain nombre de molécules inhibitrices synthétisées au sein de leur laboratoire sur l'activité d'Uhgb_MP. Ces molécules sont formées de dextranes linéaires ou ramifiés utilisés comme châssis pour la synthèse d'analogues de sucres ultra-valents de la famille des iminosucres (fig. 26). *In vitro*, certaines de ces molécules réduisent de façon dose-dépendante l'activité d'Uhgb_MP. Cependant, leur effet inhibiteur est limité, avec des IC_{50} de plusieurs centaines de µM. Par ailleurs, de

façon surprenante, l'une de ces molécules, un immino-mannose ultra-valent, présente un effet activateur pour des concentrations inférieures à 500µM. En effet, utilisé à cette concentration, le composé YB.2.135 permet d'augmenter d'un facteur 70 l'activité de Uhgb_MP sur *p*NP-β-D-Mannopyranose.

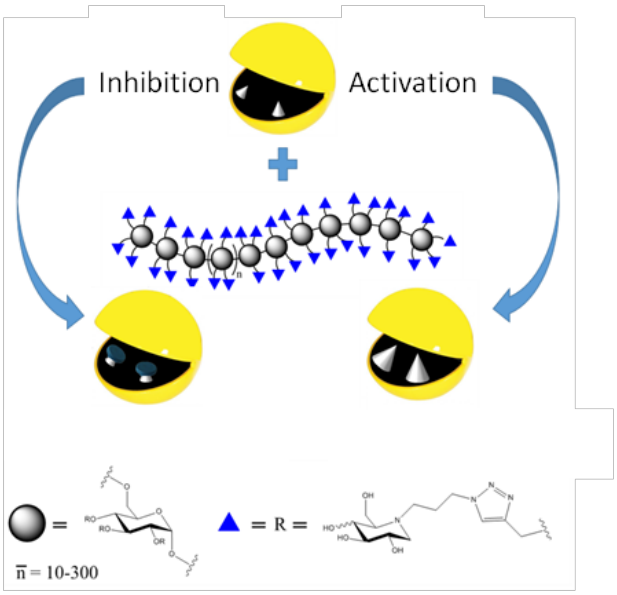


Figure 26: Representation schématique de l'effet de certains iminosucres multivalents sur Uhgb_MP.

Les molécules testées sont des molécules multivalentes formées d'un squelette de dextrane ([α-1,6-Glc]$_n$) linéaire ou ramifié avec des unités glucosyls branchés en α-1,2. Le taux de branchement et la longueur des chaines sont variables. Sur ce squelette sont ajoutés des iminosucres, connues pour leurs propriétés inhibitrices de certaines glycoside hydrolases.

Afin de mieux comprendre le mécanisme d'action de cette molécule, dont le rôle activateur a par ailleurs été observé pour plusieurs glycosidases d'intérêt médical, j'ai entrepris des expériences de biophysique. Les résultats de DLS démontrent clairement que cette molécule permet la formation d'agrégats de protéine, car le rayon hydrodynamique observé d'Uhgb_MP augmente de façon proportionnelle avec la concentration en molécule activatrice, jusqu'à atteindre une taille de 1990nm pour une concentration en activateur de 500µM, concentration au-delà de laquelle la protéine précipite sous forme inactive. Par conséquent, cette molécule entraine l'agrégation d'Uhgb_MP, probablement par pontage entre différentes protéines, ce qui augmente son activité. Ces résultats révèlent que les effets des imminosucres multivalents sont plus complexes que prévus, ces molécules pouvant désormais être considérées comme cofacteurs artificiel pour diverses applications thérapeutiques et biotechnologiques.

L'ensemble des résultats portant sur les effets inhibiteurs et activateurs des imminosucres multivalents, ont fait l'objet d'une publication, parue dans la revue *Bioconjugate Chemistry* en 2015.

Il apparait donc, au travers des différents résultats acquis au cours de cette thèse, que les enzymes de famille GH130 sont des acteurs importants au sein de l'écosystème intestinal humain. L'enzyme Uhgb_MP, principale cible de ce projet, est désormais complètement caractérisée, tant en ce qui concerne sa fonction que sa structure 3D. Cette enzyme et ses plus proches homologues, codés par des gènes très prévalents au sein de l'écosystème intestinal, apparaissent comme des éléments centraux dans la dégradation des N-glycanes de l'hôte par les bactéries du genre *Bacteroides*, qu'elles soient cultivées ou non cultivées. La large gamme de molécules acceptrices qu'elle tolère en fait aussi un outil biotechnologique très performant pour la synthèse d'oligosaccharides à très haute valeur ajoutée, à partir de biomasse végétale. L'élucidation de sa structure 3D a permis d'identifier les déterminants moléculaires de sa spécificité. Ces connaissances pourront être mises à profit dans le futur afin de cibler des modifications structurales à lui apporter en vue générer des mutants plus efficaces pour la production des molécules à haute valeur ajoutée telles que le β-D-Man*p*-1,4-D-Glc*p*NAc, le manno-chitobiose, ou d'autres molécules mannosylées d'intérêt industriel, et ce à partir de mannane, une matière première renouvelable et bon marché.

Ces travaux ont également permis d'accroitre les connaissances sur la famille GH130 dans son ensemble, de créer deux sous-familles, et de les associer à des fonctions physiologiques différentes. Cette famille, très originale, reste cependant mal connue. A ce jour, seuls 8 autres membres sont caractérisés sur le plan fonctionnel, et seules 6 autres structures sont connues. Il sera important dans l'avenir de caractériser d'autres membres de cette famille, afin de confirmer les hypothèses que nous avons émises concernant le rôle physiologique joué par les membres des deux sous-familles. Mais le chantier le plus important concerne certainement le groupe des GH130_NC, pour lesquelles seulement 3 membres ont été caractérisés, dont un dans le cadre de cette thèse. Ce groupe est hétérogène et est très probablement impliqué dans des rôles physiologiques divers, comme le prouve la spécificité unique des GH130 issues de *Thermoanaerobacter* sp. X514 caractérisées par Chiku *et al.*, et la polyvalence de la protéine BACOVA_03624 caractérisée ici. L'identification de nouvelles spécificités de fonction au sein des GH130_NC ainsi que l'accroissement de la taille de la famille dans son ensemble permettra sans doute par la suite de créer de nouvelles sous familles. En ce qui concerne BACOVA_03624, seule la structure 3D de cette enzyme sous sa forme *apo* est connue, ce qui limite notre compréhension de cette large spécificité de substrat ainsi que sa capacité à catalyser ces réactions à l'aide de différents mécanismes enzymatiques. Obtenir ces structures en complexe avec les différents substrats devrait être la prochaine étape à franchir dans la compréhension du fonctionnement de cette enzyme à l'échelle atomique. Ces nouvelles données pourraient également permettre d'entrevoir de futures modifications à apporter à l'enzyme afin d'améliorer ses propriétés de phosphorolyse inverse et de transglycosylation pour la synthèse d'oligosaccharides à haute valeur ajoutée. Egalement, des études de transcriptomique seront réalisées à court terme pour démontrer la réelle implication de

BACOVA_03624 dans la dégradation des N-glycans humains et/ou de levures. Les tout derniers résultats des tests de croissance dont nous disposons semblent aller dans ce sens, puisque *B. ovatus* ATCC 8483 est bien capable de croitre avec une glycoprotéine de levure fortement N-mannosylée comme seule source de carbone.

Nous avons par ailleurs commencé à caractériser la protéine *Tm*1225 issue de la bactérie marine *Thermotoga maritima* MSB8, appartenant à la sous-famille GH130_2 et partageant 61% d'identité de séquence avec Uhgb_MP. Nos premiers résultats montrent que *Tm*1225 est la β-mannoside polymerase la plus efficace connue à ce jour, puisqu'elle est capable de synthétiser par phosphorolyse inverse de chaines de β-manno-oligosaccharides d'un degré de polymérisation allant de 2 à 20, *a minima*. Les analyses RMN sont en cours et devraient bientôt permettre de déterminer la spécificité de liaison de cette enzyme. Elle fera l'objet de prochaines analyses permettant de vérifier ses capacité de polymérisation sur une large gamme de substrats glycosylés afin d'évaluer son potentiel dans la synthèse d'oligosaccharides à haute valeur ajoutée et de nouveaux copolymères biosourcés. Sa structure *apo*, qui a été résolue en 2004, s'avère être la plus semblable à celle de Uhgb_MP. En revanche, *Tm*1225 a cristallisé sous forme de dimère. Il serait alors très intéressant de confirmer cette organisation en solution, ainsi que d'identifier les bases moléculaires guidant la multimérisation au sein de cette famille de glycoside phosphorylases, la seule contenant des GPs non dimériques. Enfin, résoudre sa structure en complexe avec ses substrats permettrait d'identifier les bases moléculaires permettant la polymérisation de manno-oligosaccharides de haut degré de polymérisation.

Pour finir, la relation entre surabondance de séquences GH130 dans le microbiome intestinal et maladies inflammatoires reste à confirmer sur des cohortes de patients plus importantes, analysées à la fois par métagénomique et métatranscriptomique. Les effets des inhibiteurs glycosidiques et immino-glycosidiques identifiés dans le cadre de cette étude devront aussi être étudiés, par intégration de données « omiques » ciblant les souches cultivées portant des GH130, cultivées sur N-glycanes, en présence et absence de ces composés, voire sur des modèles animaux à flore humaine. Si les effets de ces composés sont confirmés à cette échelle, des études complémentaires devront être entreprises pour caractériser leurs interactions avec les enzymes cibles, notamment par diffraction des rayons X pour les composés de faible masse molaire. Le mécanisme d'interaction de l'activateur ultravalent pourra être quant à lui étudié par des expériences de SAXS, afin de connaitre l'évolution de la taille des particules en solution pour de faibles concentrations de ce composé, ou de composés similaires avec une valence et/ou un taux de branchement différents. Un modèle d'agrégation pourrait alors en découdre, ce qui pourrait guider la synthèse de nouvelles molécules afin de créer de nouveaux inhibiteurs ou activateurs, plus efficaces, valorisables respectivement dans les domaines de la santé et de l'industrie.

Appendix

GROSSET-FOURNIER & DEMACHY

DEMANDE DE BREVET INITIAL

EUROPEEN

DEPOSEE LE 31 JUILLET 2013

SOUS LE NUMERO 13 306 108.5

AUX NOMS DE

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE

INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE TOULOUSE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE D'AIX MARSEILLE

TITRE

«USE OF SPECIFIC GLYCOSIDE PHOSPHORYLASES FOR THE IMPLEMENTATION OF PHOSPHOROLYSIS OR REVERSE PHOSPHOROLYSIS REACTIONS»

INVENTEURS

POTOCKI de MONTALK-VERONESE, Gabrielle
LADEVEZE, Simon
TARQUIS, Laurence
LAVILLE, Elisabeth
HENRISSAT, Bernard
MONSAN, Pierre

Grosset-Fournier & Demachy
54, rue Saint-Lazare
75009 Paris
Téléphone : 01 42 81 09 58
Télécopie : 01 42 81 08 71

RÉSUMÉ:

Les relations entre bactéries intestinales, aliments et hôte jouent un rôle crucial dans le maintien de la santé humaine. La caractérisation fonctionnelle d'Uhgb_MP, une enzyme de la famille 130 des glycoside hydrolases découverte par métagénomique fonctionnelle, a révélé une nouvelle fonction de dégradation par phosphorolyse des polysaccharides de la paroi végétale et des glycanes de l'hôte tapissant l'épithélium intestinal. Les déterminants moléculaires de la spécificité d'Uhgb_MP vis-à-vis des mannosides ont été identifiés grâce à la résolution de sa structure cristallographique, sous forme *apo* et en complexe avec ses ligands. Un nouveau procédé de synthèse par phosphorolyse inverse d'oligosaccharides mannosylés à haute valeur ajoutée, a aussi été développé. Enfin, la caractérisation fonctionnelle de la protéine BACOVA_03624 issue de *Bacteroides ovatus* ATCC 8483, une bactérie intestinale hautement prévalente, a révélé que la famille GH130 comprend à la fois des glycoside-hydrolases et des glycoside-phosphorylases capables de dégrader les mannosides et les galactosides, et de les synthétiser par phosphorolyse inverse et/ou transglycosylation. L'ensemble de ces résultats, ainsi que l'indentification d'inhibiteurs des enzymes de la famille GH130, ouvrent de nouvelles perspectives pour l'étude et le contrôle des interactions microbiote-hôte.

SUMMARY:

The interplay between gut bacteria, food and host play a key role in human health. The functional characterization of Uhgb_MP, an enzyme belonging to the family 130 of glycoside hydrolases, discovered by functional metagenomics, revealed novel functions of plant cell wall polysaccharide and host glycan degradation by phosphorolysis. The molecular determinants of Uhgb_MP specificity towards mannosides were identified by solving its crystal structure, in *apo* form and in complex with its ligands. A new process of high added value mannosylated oligosaccharide synthesis by reverse-phosphorolysis was also developed. Finally, the functional characterization of the BACOVA_03624 protein from *Bacteroides ovatus* ATCC 8483, a highly prevalent gut bacterium, revealed that GH130 family both contains glycoside phosphorylases and glycoside hydrolases, which are able to degrade mannosides and galactosides, and to synthesize them by reverse-phosphorolysis and/or transglycosylation. All these results, together with the identification of GH130 enzyme inhibitors, open new perspectives for studying, and potentially also for controlling, interactions between host and gut microbes.